Lec12: 非参数统计方法

张伟平

May 4, 2011

§1 一样本问题中的非参数假设检验

在上一章我们讨论了当总体分布族是正态情形, 关于均值得一样本t检验方法. 但是, 当我们无把握认为总体分布族为正态模型时, 则必须用其它方法来检验. 下面介绍几种常用的非参数方法, 即符号检验法、符号秩和检验法和Fisher置换检验法。

一、符号检验法

例1 为比较甲乙两种酒的优劣, 找了N个人去品尝. 同一个人品尝两种酒后, 请他们分别给两种酒评分. 这里, 每一个品酒人对甲、乙两种酒的评分结果构成一个对子, 正好是一个成对比较的模型.

以X_i记第i个品酒人对甲酒的评分, Y_i记第i个品酒人对乙酒的评分. 记$Z_i = X_i - Y_i$, $i = 1, \cdots, N$.如果假定$Z_i \sim N(\mu, \sigma^2)$,则甲、乙两酒是否有优劣的问题将转化为原假设$H_0:\ \mu = 0$的检验问题, 这就是我们在§5.2讨论过的一样本t检验问题. 可是在一些情况下, 我们不见得有根据去假定Z_i服从正态分布. 这时上述方法就失效了. 下面是一个替代方法: 每一个评就人的评分给出一个符号

$$S_i = \begin{cases} + & 若Z_i > 0 \\ - & 若Z_i < 0 \\ 0 & 若Z_i = 0 \end{cases} \tag{1.1}$$

即品就人给以“+”号表示他认为“甲酒优于乙酒”, 另两个符号的意义类推. 如此, 我们得到n个符号$S_1, \cdots, S_n$.原假设

$$H_0:\ 甲乙两种酒一样好 \tag{1.2}$$

的检验就建立在试验结果的这n个符号的基础上, 故称为符号检验(Sign Test).下面将会看到: 从统计模型而言, 符号检验不过是二项分布参数检验的一个特例. 符号检验的具体方法如下:

记N个试验结果$S_1, \cdots, S_n$中“+”号的次数有n_+次, 出现“$-$”号的有n_-次, 其余为0. 记$n = n_+ + n_-$.如果H_0成立, 即甲乙两种酒一样好, 则在n个非0结果中出现“+”或“$-$”的机会相同. 即每个非0试验结果中出现“+”号的概率$p = 1/2$;若甲、乙两酒确有优劣之分, 则每个非0结果中出现“+”的概率$p \neq 1/2$. 若记$X = n_+$,

放在这个情况下, n_+的分布服从$b(n, 1/2)$,若甲乙两种酒确有优劣之分, 则每个结果出现“+”号的概率$p \neq 1/2$. 则所提问题转化为检验问题: X二项分布 $b(n,p)$, $0 \le p \le 1$,要检验

$$H_0: p = \frac{1}{2} \longleftrightarrow H_1: p \neq \frac{1}{2}. \tag{1.3}$$

一个合适的检验为

$$\text{当 } |X-n/2|>c \text{ 时否定 } H_0.$$

临界值c要根据给定的检验水平α, 由二项分布来决定(见附表10). 为使α为真实水平, 必要时用随机化检验. 一个更确当的方法是计算检验的p值(见§5.3,四). 在此, 令由样本$S_1,\cdots,S_n$算得的$X=n_+$的具体值为x_0,记$x_0'=min\{x_0,n-x_0\}$,则检验的p值为

$$p=\sum_{i=0}^{x_0'}\binom{n}{i}\left(\frac{1}{2}\right)^n+\sum_{i=n-x_0}^{n}\binom{n}{i}\left(\frac{1}{2}\right)^n \tag{1.4}$$

若n为偶数, 而$x_0=n/2$,则取p值为$p=1$. p值越接近1, 则H_0越可信. 如给定检验水平α,则当$p<\alpha$时否定H_0.

在例1中,给定检验水平α,则检验问题(1.2)的否定域为

$$\{X=n_+\geq c,\ \text{或}\ X\leq d\},$$

其中c和d的值由下式确定:

$$\sum_{i=c}^{n}\binom{n}{i}\Big(\frac{1}{2}\Big)^n\leq\frac{\alpha}{2},\quad d=n-c.$$

在例1中,令$N=13,\ S_1,\cdots,S_{13}$中$+$号和$-$号的个数分别是$n_+=2,\ n_-=10$,因此$n=n_++n_-=12$.取检验水平$\alpha=0.05$,查附表10"符号检验临界值表"得$c=10$,故$d=n-c=2$.故检验的否定域$D=\{X=n_+\geq 10,\ \text{或}\ X\leq 2\}$.检验统计量$X=n_+=2$, 因此否定原假设. 即认为甲、乙两酒不一样.

对这一检验问题, 也可通过计算检验的p值来解决. 此处, $n=12,\ x_0=n_+=2$,按(1.4), $x_0'=min(2,12-2)=2$,查二项分布表得

$$p=\sum_{i=0}^{2}\binom{12}{i}\left(\frac{1}{2}\right)^n+\sum_{i=10}^{12}\binom{12}{i}\left(\frac{1}{2}\right)^n=0.0384<0.05$$

故在0.05显著性水平下应否定H_0.

例2 工厂的两个化验室, 每天同时从工厂的冷却水总取样, 测量水中的含氯量一次. 下面是$n=11$天的记录:

i	1	2	3	4	5	6	7	8	9	10	11
x_i	1.15	1.86	0.76	1.82	1.14	1.65	1.92	1.01	1.12	0.90	1.40
y_i	1.00	1.90	0.90	1.80	1.20	1.70	1.95	1.02	1.23	0.97	1.52

其中x_i表示化验室A的测量记录, y_i表示化验室B的测量记录. 问两个化验室测定的结果之间有无显著差异? 取$\alpha=0.10$.

解 分别记化验室A和B的测量误差为ξ和η.设ξ和η为连续型随机变量, 其分布函数分别为$F(x)$和$G(x)$.检验问题是

$$H_0:F(x)=G(x)\longleftrightarrow H_1:F(x)\neq G(x). \tag{1.5}$$

显然含氯量的测定值, 除了与化验室的不同有关外, 还与当日水中含氯量的多少有关. 我们可以认为X_i和Y_i具有数据结构:

$$X_i = \mu_i + \xi_i, \qquad Y_i = \mu_i + \eta_i, \quad i = 1, 2, \cdots, n.$$

其中μ_i为第i天水中的含氯量, ξ_i和η_i分别表示第i天化验室A、B的测量误差. 显然$\xi_1, \cdots, \xi_n$和$\eta_1, \cdots, \eta_n$ 都是不可观察的独立同分布的随机变量. 前者与$\xi \sim F(x)$同分布, 后者与$\eta \sim G(x)$同分布.

不同日的两个数据X_i与Y_i显然不一定是同分布的, 而且X_i与X_j, 以及Y_i与Y_j也不一定是同分布的. 它们之间的差异不但与测量误差有关, 而且也与μ_i和μ_j的差异有关. 因此虽然$X_1, \cdots, X_n$相互独立, 但不能假定它们同分布, $Y_1, \cdots, Y_n$也是如此. 所以两样本的统计比较方法, 如两正态样本的t检验方法以及后面要介绍的两样本非参数检验方法都不能用于这类数据的检验工作. 我们在§5.2中也提到过成对数据的上述特点.

处理成对数据检验问题, 很自然地想到如何把μ_i的影响消除掉. 由于对每个i,X_i与Y_i之间可比, 若将同一天的两个数据相减, 从而把μ_i的影响消除掉. 令

$$Z_i = X_i - Y_i = \xi_i - \eta_i, \quad i = 1, 2, \cdots, n. \tag{1.6}$$

显然Z_i仅与化验室A、B在第i日的测量误差之差有关. 记$Z = \xi - \eta$, 则$Z_1, \cdots, Z_n$可看成来自总体Z的随机样本, 即$Z_1, \cdots, Z_n$ 是独立同分布的样本. 由于Z是两个测量误差之差, 因此Z的均值为0, 且可证明它是关于原点对称的.

令n_+为$Z_1, \cdots, Z_n$中取正值的个数, n_-为$Z_1, \cdots, Z_n$中取负值的个数, 它们都是$r.v.$.由于假定了ξ和η是连续型随机变量, 故$Z_1, \cdots, Z_n$中取值为0的个数以概率为1取0. 因此可记$n = n_+ + n_-$.当H_0,即(1.5)成立时, 则在n个试验单元中Z_i取"+"和取"−"的可能性皆为$\frac{1}{2}$. 因此检验问题转化为: $n_+ \sim b(n, p),\ 0 \le p \le 1$,检验

$$H_0' : p = \frac{1}{2} \longleftrightarrow H_1' : p \neq \frac{1}{2}$$

否定域$D = \{n_+ \ge c \text{ 或 } n_+ \le d\}$.

因此, 在给定显著性水平α之后, c和d的值由

$$\sum_{k=c}^{n} \binom{n}{k} \Big(\frac{1}{2}\Big)^n \le \frac{\alpha}{2}, \quad d = n - c$$

所确定.

在本例中$n = 11,\ \alpha = 0.10$, 查二项分布表知

$$\sum_{k=0}^{2} \binom{11}{k} \Big(\frac{1}{2}\Big)^{11} = 0.0327,$$
$$\sum_{k=0}^{3} \binom{11}{k} \Big(\frac{1}{2}\Big)^{11} = 0.113,$$

所以$d = 2,\ c = 11 - 2 = 9$ (也可查附表10得$c = 9,\ d = n - c = 2$). 故水平$\alpha = 0.10$ 的符号检验的否定域为

$$\{n_+ \le 2 \text{ 或 } n_+ \ge 9\}$$

作差值$z_i = x_i - y_i$,得

$$\begin{array}{cccccc} 0.15, & -0.04, & -0.14, & 0.02, & -0.06, & -0.05, \\ -0.03, & -0.01, & -0.11, & -0.07, & -0.12, & \end{array}$$

其中取正数的个数为$n_+ = 2$, 因此在水平$\alpha = 0.10$下否定H_0,即认为化验室A、B测定结果之间有显著差异.

符号检验的另一个重要应用是分位数(特别是中位数)检验. 请看下例.

例3　检验某种维尼纶的纤度, 测得100个数据如下表所示　试问该维尼纶纤度的中位

表 1.1

编号	1	2	3	4	5	6	7	8	9	10
纤度	1.26	1.29	1.32	1.35	1.38	1.41	1.44	1.47	1.50	1.53
频数	1	4	7	22	23	25	10	6	1	1

数m_e是否为1.40? ($\alpha = 0.05$)

解　本题在显著水平$\alpha = 0.05$下, 检验假设

$$H_0 : m_e = 1.40 \longleftrightarrow H_1 : m_e \neq 1.40$$

若令表中所列100个数据的纤度值为$X_i,\ i = 1, \cdots, 100$, 令$Y_i = X_i - 1.40,\ i = 1, \cdots, 100$. 计算$Y_i$取正值得个数$n_+$和取负值的个数$n_-$, 取值为0的个数为0, 因此$n_+ +\ n_- = 100$.在$H_0$成立的前提下, 则每个$Y_i$为正或负的可能性皆为$1/2$, 故100个数据中$n_+$和$n_-$应差别不大, 若记$X = n_+$,易见$X \sim b(100, 1/2)$,因此检验问题转化为: $X \sim b(100, p),\ 0 \leq p \leq 1$,要检验

$$H_0 : p = \frac{1}{2} \longleftrightarrow H_1 : p \neq \frac{1}{2}, \quad \alpha = 0.05$$

否定域为$D = \{X \geq c_2 \text{ 或 } X \leq c_1\}$. 利用中心极限定理可知: 当$H_0$成立, 且$n \to \infty$时有

$$\frac{X - n/2}{\sqrt{n/4}} = \frac{2X - n}{\sqrt{n}} \xrightarrow{\mathscr{L}} N(0, 1)$$

本题中$n = 100$, 令

$$\sum_{i=0}^{c_1} \binom{100}{i} \Big(\frac{1}{2}\Big)^{100} \approx \Phi\Big(\frac{c_1 - 50}{5}\Big) = \frac{\alpha}{2} = 0.025,$$

查表得$(c_1 - 50)/5 = -1.96$, 解得$c_1 = 40.2$

类似地由

$$\sum_{i=c_2}^{100} \binom{100}{i} \Big(\frac{1}{2}\Big)^{n} \approx 1 - \Phi\Big(\frac{c_2 - 50}{5}\Big) = 0.025$$

查表得$(c_2 - 50)/5 = 1.96$,解得$c_2 = 59.8$,故否定域为

$$\{X : X \leq 40.2 \text{ 或 } X \geq 59.8\}$$

由表1.1算得$X = n_+ = 43$,它介于(40.2, 59.8)之间, 故不足以否定H_0, 故认为该维尼纶的纤维度的中位数是1.40.

符号检验与二项分布参数检验的关系:

假设我们感兴趣一个实值连续型随机变量U, 记其p_0分位数为m_q,即

$$p_0 = P(U \le m_q)$$

实际中我们往往不知道m_q的值, 即便是指定p_0的值,这是由于我们不知道U的分布. 对某个特定的m_0, 记

$$p = P(U \le m_0)$$

此时由于U的分布未知, 故而p未知. 由于U为连续型随机变量, 故而

$$\begin{aligned} m_q = m_0 &\quad \text{当且仅当} \quad p = p_0 \\ m_q \le m_0 &\quad \text{当且仅当} \quad p \ge p_0 \\ m_q \ge m_0 &\quad \text{当且仅当} \quad p \le p_0 \end{aligned}$$

于是关于m_q的假设等价于关于p的假设. 记U的一组样本为$U_1, \cdots, U_n$, 从而符号检验统计量为

$$T = \sum I(U_i \le m_0)$$

显然$T \sim B(n, p)$. 于是由二项分布的检验容易得到此时关于U的分位数的假设检验法则.

二、符号秩和检验

让我们再回顾一下符号检验, 仍就例1中品酒的问题来说明. 在计算$Z_i = X_i - Y_i$后, 我们放弃Z_i的具体数值而取其符号S_i时, 丢失了一些信息. 这种信息的丢失, 使符号检验的效率有所降低. 为此提出了符号秩和检验, 它是符号检验的改进.

例4 仍看例1, 设想请了13个人品尝甲、乙两种酒, 评分结果如下:

表 1.2

品酒人	1	2	3	4	5	6	7	8	9	10	11	12	13
甲 (x_i)	55	32	41	50.5	60	48	39	45	48	46	52.2	45	44
乙 (y_i)	35	37	43.1	55	34	50.3	43	46.1	51	47.3	55	46.5	44
符号(z_i)	+	−	−	−	+	−	−	−	−	−	−	−	0

此处$z_i = x_i - y_i$.试问甲乙两种酒是否一样好? 一共12个非0符号中, 有两个"+"号, 显示多数品酒人认为乙酒好. 在符号检验中我们就只能根据"+"、"−"号的数目去下结论. 但细看一下结果, 我们发现, 在认为"乙酒比甲酒优"的10人中, 乙酒的得分比甲酒高得不多, 而在认为"甲酒优于乙酒"的2人中, 甲的得分远远高于乙. 这个事实给2 : 10这个表面结果, 打了一个折扣, 它启示我们: 除了考虑符号外, 还应当把这一点考虑进来. 符号秩的概念提供了一种作法.

定义6.2.1 设$X_1, \cdots, X_n$为两两不相等的一组样本, 将其大小排列为$X_{(1)} < \cdots < X_{(n)}$, 若$X_i = X_{(R_i)}$, 则称$X_i$在样本$(X_1, \cdots, X_n)$中的秩为$R_i$.

显然, 若$X_1, \cdots, X_n$为来自连续型分布$F(x)$的样本, 则以概率为1保证$X_1, \cdots, X_n$ 中两两互不相等.

定义6.2.2 设$X_1, \cdots, X_n$为来自单个连续型总体的样本, 或来自多个连续型总体的合样本. 则$R = (R_1, R_2, \cdots, R_n)$称为$(X_1, \cdots, X_n)$的*秩统计量*, 其中$R_i$为$X_i$的秩. 由$R$导出的统计量也称为*秩统计量*. 基于秩统计量的检验方法称为*秩检验*.

现在仍回到例4, 把表1.2扩充成下表. 我们把符号为"+"的那两个秩(即11和12) 括起

表 1.3

品酒人(i)	甲(x_i)	乙(y_i)	符号(z_i)	$\|Z_i\| = \|x_i - y_i\|$	秩
1	55	35	+	20	[11]
2	32	37	−	5	10
3	41	43.1	−	2.1	4
4	50.5	55	−	4.5	9
5	60	34	+	26	[12]
6	48	50.3	−	2.3	5
7	39	43	−	4	8
8	45	46.1	−	1.1	1
9	48	51	−	3	7
10	46	47.3	−	1.3	2
11	52.2	55	−	2.8	6
12	45	46.5	−	1.5	3
13	44	44	0	0	不定秩

来, 它们的和是$W^+ = 11 + 12 = 23$, 叫做"符号秩和". 一般它可以用下列方式来定义: 记$Z_i = X_i - Y_i$, 令

$$\bar{V}_i = \begin{cases} 1 & 若Z_i > 0; \\ 0 & 其它. \end{cases}$$

R_i为$|Z_i|$在$(|Z_1|, \cdots, |Z_n|)$中的秩, 则Wilcoxon *符号秩和* (the sum of Wilcoxon signed rank)检验统计量定义为

$$W^+ = \sum_{i=1}^{n} \bar{V}_i R_i. \tag{1.7}$$

容易理解: 在例6.2.5中, 若甲优于乙, 则不仅"+"号会多, 且"+"号观察相应的秩, 一般也偏大, 故总的效果是W^+应偏大. 反之, 若乙优于甲, 则W^+将偏小. 因此检验问题(1.2), 即

$$H_0: 甲、乙两酒一样好$$

成立时, W^+应当不大不小. 检验的否定域是

$$\{W^+ \le d \text{ 或 } W^+ \ge c\}, \tag{1.8}$$

此处d和c取决于n (本例中$n = 12$), 及指定的检验水平α. 即, 当给定α时, c, d分别由下列两式决定:

$$P(W^+ \le d \,|H_0) \le \alpha/2, \qquad P(W^+ \ge c \,|H_0) \le \alpha/2.$$

H_0为真时W^+的分布见参考文献[4] P_{246}. 对某些特定的α及不大的n, c和d可以查表求得, 见书末附表11. 表中仅可查到c,而$d = n(n+1)/2 - c$.

由表1.3可知本题中$n=12$, $W^+=23$.取$\alpha=0.05$,查表中$\alpha/2$那一栏, 在$n=12$处得$c=65$, 算得$d=13$,按(1.8)得否定域为

$$\{W^+ \leq 13 \text{ 或 } W^+ \geq 65\}.$$

而$13<W^+=23<65$,故应接受H_0, 即所得观察结果不构成甲、乙有优劣之分的充分证据.

这个检验称为Wilcoxon双侧符号秩和检验 (以下简称双侧W^+检验) , 之所以取$\alpha/2$, 也是由于这个“双侧”而来.

可以证明:

$$E(W^+)=\frac{n(n+1)}{4},\quad D(W^+)=\frac{1}{24}n(n+1)(2n+1)$$

与下节的秩和统计量W类似, 当$n\to\infty$时, W^+的标准化随机变量

$$W_*^+=\frac{W^+-n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}\xrightarrow{\mathscr{L}}N(0,1) \tag{1.9}$$

故例6.2.5的水平近似为α的双侧W^+检验的否定域为

$$\left\{|W_*^+|\geq u_{\alpha/2}\right\}$$

取$\alpha=0.05$,算得$|W_*^+|=1.26<1.96=u_{0.025}$,故接受$H_0$, 根据现有观察值不足以否定$H_0$.

我们可以看到例1和例6.2.5中的同一个检验问题用符号检验和符号秩和检验得到两种不同的结论. 按符号检验否定H_0,即认为甲、乙两酒有优劣之分, 且乙优于甲. 按符号秩和检验的小样本和大样本方法, 都接受H_0, 即表明无充分证据否定“甲、乙两酒一样好”. 这里我们看到: 同一个问题, 同一批数据, 用不同方法, 检验结果不同, 这不足为怪. 正如用同一批数据去估计正态总体的数学期望值, 用样本均值估计与用中位数估计, 两者结果不同. 这就产生了一个问题: 这两种检验法哪一种好? 这个问题不能一概而论, 有兴趣的读者可查看参考文献[9] P_{156}中表9.1所列的结果. 可以指出的是: 符号检验全然不看数值而只看符号; 基于正态假定的t检验则要看数值, W^+检验介于二者之间: 它既不忽视数值, 也不全看数值(数值只用于决定秩, 而不用其本身值) .

三、 Fisher的置换检验*

例5 为比较A、B两种施肥方法何种为优, 选择15块一样大的地, 把每块分成形状大小一样的两小块, 随机地将其中的一块分给A,另一小块给B. 收获时得到各小块的产量如下:

块号	1	2	3	4	5	6	7	8
A	188	96	168	176	153	172	177	163
B	139	163	160	160	147	149	149	122
$A-B$	49	-67	8	16	6	23	28	41
块号	9	10	11	12	13	14	15	
A	146	173	186	168	177	184	96	
B	132	144	130	144	102	124	144	
$A-B$	14	29	56	24	75	60	-48	

算出$\sum(A-B)=314$,现在要检验假设

$$H_0: A、B \text{ 的效果一样.} \tag{1.10}$$

若(1.10)成立, 每块内$A-B$值(即$49, -67, \cdots$等)不一样, 并非由于A、B效果不同, 而是由于其两小块的差别. 但随机化的结果, 每一小块有同等可能分给A或B. 因此, 如在第一块, 依随机化的结果不同, $A-B$可以是49, 也可以是-49,要看较好的那块派给A还是B. 这样一来, 这个试验的全部可能的$\sum(A-B)$值有2^{15}个:

$$\pm(49), \pm(-67), \pm(8), \cdots, \pm(60), \pm(-48),$$

实际得出的$\sum(A-B)=314$是2^{15}中的一个. 当A、B效果有较大差别时$|\sum(A-B)|$应取大值. 对2^{15}个可能结果中的每一个算出$\sum(A-B)$,用x_i记之, $i=1,2,\cdots,2^{15}$. 将它们按照它们的绝对值从大到小的顺序排列, 不妨记为

$$x_1,\ x_2,\ \cdots,\ x_{2^{15}} \tag{1.11}$$

即满足

$$|x_1|>|x_2|>\cdots>|x_{2^{15}}|$$

(1.11)中的2^{15}个值中, 在H_0成立前提下, 为等可能发生, 即每个出现的概率都是$1/2^{15}$.检验问题(1.10)的否定域为

$$\{|\sum(A-B)|>c\}$$

观测到得$|\sum(A-B)|=314$, 从而检验的P值为

$$P(|\sum(A-B)|>314|H_0)=\frac{m}{2^{15}}$$

其中m为排序(1.11)中满足$x_m=314$.

具体计算可知$p_{314}<0.0001$ 因此有理由否定H_0.

置换检验的缺点是: 在具体实施时计算量大, 使用起来不方便. 但现在有了高速计算机, 利用计算机来实施也不算难事了.

Fisher自己和其它许多学者, 都研究过这样的问题: 当n很大时, 可否找到一种近似的方法去实施置换检验, 以大大简化计算? 研究结果证明了: 在很一般的条件下, 这种简化的方法不仅存在, 且就是通常的t检验! 这是一个很有意思的结果. 因为一开始, t检验是局限在正态模型中导出的. 通过这个途径发现, 即使在更为广泛的模型下, 只要试验次数足够大, t检验仍是适用的, 因此可以说, 置换检验的理论从一个侧面加强了t检验的地位.

§2　两样本问题中的非参数假设检验

在两样本的比较问题中, 当样本的随机误差不服从正态分布时, 就需要提出更一般得假设, 并使用相应的非参数检验方法. 这方面的理论和方法较多, 但大都很专门, 这里只对Wilcoxon秩和检验和置换检验作一简略介绍.

一、引言及定义

我们首先来看一看这一检验的实际背景. 两样本检验问题的一般提法如下: 设$X_1,\cdots,X_m$和$Y_1,\cdots,Y_n$分别是从具有分布为F_1和F_2的一维总体中抽取的简单样本, 且假定合样本$X_1,\cdots,X_m$, $Y_1,\cdots,Y_n$全体相互独立. 要检验下列假设

$$H_0: F_1=F_2 \longleftrightarrow H_1: F_1\neq F_2. \tag{2.1}$$

在数理统计学中, 习惯上称这个检验问题为"两样本问题". 我们来分别考虑下列几种请况:

1. 设根据问题的实际背景, 如果我们有理由假定F_1和F_2为具有相同方差的正态分布, 即假定

$$F_1 \sim N(a, \sigma^2), \quad F_2 \sim N(b, \sigma^2)$$

其中a、b和σ^2皆未知, $-\infty < a,\ b < +\infty,\ \sigma^2 > 0$,这时检验问题转化为

$$H_0' : a = b \longleftrightarrow H_1' : a \neq b. \tag{2.2}$$

在这个假定下, 总体分布F_1和F_2只依赖于三个未知参数a、b和σ^2, 检验问题(2.1)归结为检验这些未知参数是否满足(2.2). 按§5.1所述这属于"参数型假设检验问题". 这就是§5.2中讨论的两样本t检验.

2. 如果我们对问题的实际背景所知甚少, 我们只好认为对F_1和F_2完全未知. 在这样宽广的假定下, 我们再不能使用通常的两样本t检验. 处理这个问题的一种方法是"斯米尔洛夫"(Smirnov)检验, 这将在本章第五节中讨论.

在这一情形下, 总体分布F_1和F_2不能用有限个实参数去刻画, 因此称为非参数检验问题.

3. 现在我们讨论一种中间情况. 设X是一种产品在一定生产工艺下的质量指标, 而Y是该产品在另一生产工艺下的质量指标. 有理由认为, 改变生产工艺不影响产品质量指标的概率分布, 而只能使此分布发生一些平移. 也就是说, 若以F记X的分布, 则Y分布为$F(x-\theta)$, 这里θ是一个未知的位置参数. 在这个假定下, "X、Y同分布"的假设相当"$\theta = 0$", 而对立假设为"$\theta \neq 0$". 因此检验(2.1)归结为检验

$$H_0 : \theta = 0 \longleftrightarrow H_1 : \theta \neq 0. \tag{2.3}$$

(2.3)是一个很重要的假设检验问题. 在这一模型中, 我们假定F未知, 因而比正态模型为广. 另外这一模型又比"*斯米尔洛夫检验*"中的模型窄一些, 因为对后者而言, 两分布F_1和F_2毫无关系, 而在此F_1和F_2之间有$F_2(x) = F_1(x-\theta)$.

虽然表面上看(2.3)象一个参数检验问题: 假设中只涉及θ,而它是一个实参数. 其实不然, 因为总体的分布与F和θ都有关, 而F的分布未知, 因此按非参数统计问题的定义, (2.3)应视为非参数检验问题.

一般地, 两样本问题(2.1)还有一些具有实际背景的中间情况. 例如$F_2(x) = F_1(x/\sigma)$, 此$\sigma > 0$为未知的刻度参数, 分布F也未知. 检验问题(2.1)在此情况下转化为

$$H_0^* : \sigma = 1 \longleftrightarrow H_1^* : \sigma \neq 1. \tag{2.4}$$

Wilcoxon两样本秩和检验就是考虑(2.3)的假设检验问题. 下面首先给出Wilcoxon 两样本秩和统计量的定义.

定义6.3.1 设$X_1, \cdots, X_m, Y_1, \cdots, Y_n$这$n+m$个值两两不相同, 把它们按大小排列, 结果为

$$Z_1 < Z_2 < \cdots < Z_N, \quad N = m + n, \tag{2.5}$$

显然, 每个Y_i必为(2.5)中的某一个. 若$Y_i = Z_{R_i}$,则Y_i在合样本$X_1, \cdots, X_m$, $Y_1, \cdots, Y_n$中的秩为R_i. 而$Y_1, \cdots, Y_n$的秩和为

$$W = R_1 + \cdots + R_n, \tag{2.6}$$

它称为*Wilcoxon两样本秩和统计量*. 这是Wilcoxon在1945年的一项工作中引进的.

二、 Wilcoxon两样本秩和检验—小样本方法

Wilcoxon两样本秩和检验就是考虑(2.3)的假设检验问题, 即设$X_1, \cdots, X_m$ i.i.d. $\sim F(x)$, $Y_1, \cdots, Y_n$ i.i.d. $\sim F(x-\theta)$,且合样本独立. 要检验(2.3), 即

$$H_0 : \theta = 0 \longleftrightarrow H_1 : \theta \neq 0.$$

设样本$Y_1, \cdots, Y_n$的秩和W由(2.6)给出. 现在这样推理: 每个R_i都可取$1, 2, \cdots,$ N之一为值. 若原假设H_0成立, 则全部样本来自同一总体, 每个都不占特殊位置, 不会取较小或较大的值, W所取之值应集中在平均数$n(N+1)/2$附近. 故得到下列检验:

$$\text{当 } W \leq d \text{ 或 } W \geq c \text{ 否定 } H_0 \tag{2.7}$$

如何确定c和d? 它们的确在原则上可以解决: 当H_0成立时, 合样本独立同分布, 由此根据对称性的考虑, 易知$(R_1, \cdots, R_n)$的联合分布为

$$P(R_1 = r_1, \cdots, R_n = r_n) = \begin{cases} \frac{1}{N(N-1)\cdots(N-n+1)} & \text{当} r_1, \cdots, r_n \leq N \text{为} \\ & \text{互不相同的自然数,} \\ 0 & \text{其它.} \end{cases}$$

由此不难形式地写出W的分布. 从而由

$$\alpha = P(W \leq d \text{ 或 } W \geq c | H_0)$$

定出c和d. 对较小的m、n已制成表.

如果假设检验是单边的, 即要检验

$$H_0^{'}: \ \theta \leq 0 \longleftrightarrow H_1^{'} : \theta > 0, \tag{2.8}$$

由于$W = \sum\limits_{i=1}^{n} R_i$是$Y_1, \cdots, Y_n$在合样本中的秩和. 若$\theta > 0$则因每个$Y_i$的分布与$X_i + \theta$的分布相同, Y_i与X_i相比较倾向于取更大的值. 即Y_i取值大于X_i的“机会”更多, 而小于它的机会则少. 这样一来$R_1, \cdots, R_n$当$\theta > 0$时倾向于取集合$\{1, 2, \cdots, N\}$ 中较大的值(此处$N = m+n$), 同样$\theta < 0$, 则$R_1, \cdots, R_n$倾向于取集合$\{1, 2, \cdots, N\}$中较小的值. 因W在$\theta > 0$时倾向于取较大的值, 在$\theta < 0$时倾向于取较小的值. 故检验问题(2.8)的检验为:

$$\text{当 } W \geq c \text{ 时, 否定 } H_0'.$$

同样检验问题

$$H_0^{''}: \ \theta \geq 0 \longleftrightarrow H_1^{''} : \theta < 0, \tag{2.9}$$

的检验为:

$$\text{当 } W \leq d \text{ 时, 否定 } H_0''.$$

将上述三类检验列成下表:

表6.3.1 Wilcoxon两样本秩和检验(小样本情形)

H_0	H_1	否定域
$\theta = 0$	$\theta \neq 0$	$W \leq d$ 或 $W \geq c$
$\theta \leq 0$	$\theta > 0$	$W \geq c$
$\theta \geq 0$	$\theta < 0$	$W \leq d$

如何确定临界值c和d? 对较小的m、n已经制成表, 见书末附表12. 关于此表, 作以下两点说明:

(1) 分别记$(X_1,\cdots,X_m)$和$(Y_1,\cdots,Y_n)$在合样本中的秩和为W_1和W_2,则

$$W_1+W_2=1+2+\cdots+(m+n)=\frac{(m+n)(m+n+1)}{2}$$

是一个常数. 因此, 在使用Wilcoxon秩和检验法, 比较两总体分布时, 用W_1作为检验统计量与用W_2作为检验统计量是一回事.

(2) 在附表12中只给出了秩和检验的临界值: $P(W\geq c)\leq\alpha$中c的值, 对$P(W\leq d)\leq\alpha$的临界值d如何利用此表求出? 可以证明 $P(W\leq\alpha)=P(W\geq n(m+n+1)-d)$ 即若记$c=n(m+n+1)-d$, 先对给定的 α 、m、n 求出 $P(W\geq c)\leq\alpha$ 的临界值 c, 然后 d 由公式 $d=n(m+n+1)-c$ 算出.

例6 某种羊毛在进行某种工艺处理之前和处理之后, 各随机抽取一个样本, 测得其含脂率如下

处理前: 0.20, 0.24, 0.66, 0.42, 0.12;

处理后: 0.13, 0.07, 0.21, 0.08, 0.19.
问处理后含脂率是否下降? $(\alpha=0.05)$

解 设X和Y分别表示处理前、后羊毛的含脂率, 它们的分布函数分别为$F(x)$和$F(x-\theta)$. 则检验问题

$$H_0:\ \theta\geq 0\longleftrightarrow H_1:\theta<0.$$

即将"处理后羊毛含脂率没有下降"作为原假设.

由表6.3.1可知: 当$W\leq d$时否定原假设. 由前面关于附表12的使用说明(2), 先从附表中由$P(W\geq c)\leq\alpha$查出c, 则$d=n(m+n+1)-c$.

本题中$m=n=5,\ \alpha=0.05$,由附表12, 查出$c=36$,故有

$$d=n(m+n+1)-c=5\times 11-36=19.$$

这表明

$$P(W\leq 19)=P(W\geq 36)\leq 0.05.$$

因此上述检验问题的否定域:

$$D=\{(\mathbf{X},\mathbf{Y}):\ W\leq 19\}.$$

现将两组样本观察值按从小到大排成一列成下表

0.07	0.08	0.12	0.13	0.19	0.20	0.21	0.24	0.42	0.66
$\underline{1}$	$\underline{2}$	3	$\underline{4}$	$\underline{5}$	6	$\underline{7}$	8	9	10

下划线的数是处理后羊毛含脂率(Y)的观察值的秩. 故Y的观察值的秩和为

$$W=1+2+4+5+7=19$$

将其与否定域中临界值比较$19\leq d\ (d=19)$,因此否定H_0,即认为处理后羊毛含脂率下降了.

一般化

假设:

> 给定两组独立的随机样本$X_1, X_2, \cdots, X_n$ 和$Y_1, Y_2, \cdots, Y_m$, 要求
>
> 1. 至少为顺序尺度
> 2. 感兴趣的变量是连续型的
> 3. $F(x)$和$G(x)$分别表示了X和Y的分布函数

则对假设

(A) $H_0: F(x) = G(x)$ 对所有$x \Longleftrightarrow H_1: F(x) \neq G(x)$ 对某些x

(B) $H_0: F(x) \leq G(x)$ 对所有$x \Longleftrightarrow H_1: F(x) > G(x)$ 对某些x

(C) $H_0: F(x) \geq G(x)$ 对所有$x \Longleftrightarrow H_1: F(x) < G(x)$ 对某些x

假设(B)和(C)中的H_1分别表示了"X倾向于比Y小" 和"X倾向于比Y大". 于是记W表示X样本在合样本中的秩和, 则检验的法则分别为

(A) 对双边检验$H_1: F(x) \neq G(x)$: 若$W < c_1$ 或$W > c_2$, 拒绝H_0; 否则不足以拒绝H_0.

(B) 对单边假设$H_1: F(x) > G(x)$: 若$W < c$,拒绝H_0 ; 否则不足以拒绝H_0.

(C) 对单边假设$H_1: F(x) < G(x)$: 若$W > c$, 拒绝H_0; 否则不足以拒绝H_0.

三、 Wilcoxon两样本秩和检验—大样本方法

前面讨论了当m、n较小时Wilcoxon两样本秩和检验的小样本方法. 当m、n较大时检验统计量W的分布的计算很复杂, 对给定的α没有现成的表可以查到否定域的临界值. 因此得依靠极限定理. 方法如下: 容易求得

$$E(W) = n(m+n+1)/2 = n(N+1)/2,$$

$$D(W) = mn(n+m+1)/12 = mn(N+1)/12.$$

此处W由(2.6)式给出, $N = m+n$.记

$$W^* = \frac{W - E(W)}{\sqrt{D(W)}} = \frac{W - n(N+1)/2}{\sqrt{mn(N+1)/12}}$$

可以证明在原假设$H_0: \theta = 0$成立之下, 当$m,\ n \to \infty$时有

$$W^* = \frac{W - n(N+1)/2}{\sqrt{mn(N+1)/12}} \xrightarrow{\mathscr{L}} N(0,1)$$

因此可得检验问题(2.3)得水平近似为α的否定域:

$$D = \{(\mathbf{X}, \mathbf{Y}):\ |W^*| \geq u_{\alpha/2}\}$$

类似可得检验问题(2.8)和(2.9)的水平近似为α的检验的否定域, 详见下表:

表6.3.2　Wilcoxon两样本秩和检验(大样本情形)

H_0	H_1	否定域
$\theta=0$	$\theta\neq 0$	$\lvert W^*\rvert > u_{\frac{\alpha}{2}}$
$\theta\leq 0$	$\theta>0$	$W^*>u_\alpha$
$\theta\geq 0$	$\theta<0$	$W^*<-u_\alpha$

例7　有甲乙两台机床加工同样的产品. 从这两台机床加工的产品中随机地抽取若干产品, 测得产品直径(单位:mm)为

甲: 18.1, 17.7, 17.2, 19.1, 17.0, 17.5, 17.8, 18.7

乙: 18.3, 19.0, 18.9, 17.3, 16.9, 18.4, 17.6, 18.6, 18.0

设两台机床的精度相同, 试问甲、乙两台机床加工产品的直径有无显著差异? ($\alpha=0.10$)

解　设X和Y分别表示甲乙两台车床加工产品的直径, 它们的分布函数分别为$F(x)$ 和$F(x-\theta)$.则检验问题为:

$$H_0:\theta=0\longleftrightarrow H_1:\theta\neq 0$$

此例中$m=8,\ n=9$.将两组数据案按从小到大排成一列, 算得第二组数据的秩依次为: 1, 4, 6, 9, 11, 12, 13, 15, 16, 其秩和是

$$W=1+4+6+9+11+12+13+15+16=87.$$

对检验水平$\alpha=0.10$,由表6.3.2可知大样本否定域为

$$D=\{(\mathbf{X},\mathbf{Y}):\ |W^*|\geq u_{0.05}=1.96\}$$

其中

$$|W^*|=\Big|\frac{W-n(n+m+1)/2}{\sqrt{mn(m+n+1)/12}}\Big|=\Big|\frac{87-81}{\sqrt{108}}\Big|=0.58<1.96$$

故接受H_0,即根据已有数据不足以否定这辆台机床加工的直径无显著差异的假定.

有人可能会觉得, 这种秩和检验的效率不会高, 因为它只利用了样本中的大小关系而完全忽略了其具体数值, 其实不然. 近代关于秩检验的大样本理论证明了, 一般秩检验至少在样本大小较大时, 与传统的参数检验相比并不逊色. 拿Wilcoxon两样本秩和检验与两样本t检验相比较, 即使在随机误差分布为正态情形, Wilcoxon检验的效率相对于t检验也达到$3/\pi\approx 0.95$ (样本容量较大时) . 对误差分布非正态时, 这个相对效率可任意接近1, 且总不会低于0.864.

以上我们假定了合样本$X_1,\cdots,X_m,Y_1,\cdots,Y_n$彼此不同, 因而$Y_i$的秩$R_i$可以唯一确定. 若假定$F$处处连续, 这一事实将以概率为1成立, 这时不存在问题. 当F不连续时, 合样本中可能出现相同的, 即所谓“结”的问题. 例如说

$$x_2<x_1=x_4<x_5<x_3,$$

习惯上把相同的n个变量称为一个“结”. 结外的x_i的秩唯一确定. 如此处x_2, x_5和x_3的秩分别为1, 4 和5. 结内的x_i的秩就不清楚了. 此时取这些相继秩数的平均值作为结内各个x_i的秩. 如此处x_1和x_4占有秩2和3, 取2.5分别作为x_1和x_4的秩. 对所有的“结”作了处理之后, 按前面所述的方法讨论两样本秩和检验问题.

四、 两样本置换检验法*

在数理统计学中, “处理”一词的含意极广. 它可以表示一种工艺流程, 一个种子品种, 一种治疗方法等等. 两样本置换检验的思想与例6.2.6 相似. 设两处理的全部试验结果为$X_1, \cdots, X_m, Y_1, \cdots, Y_n$. 将其排列在一起, 改为$Z_1, \cdots, Z_{n+m}$. 如果甲、乙两处理无差别, 则$Z_1, \cdots, Z_{n+m}$之间的差别不是由于处理不同而来, 而是由于这$n+m$个试验单元的分配方法而来. 在$n+m$ 个单元中分配m个给处理甲, 不同的方法有$\binom{n+m}{m}$种. 若在每种分配方法之下都计算以下的量:

$$g = \frac{1}{m}(\text{甲处理试验值之和}) - \frac{1}{n}(\text{乙处理试验值之和}).$$

它等于$Z_1, \cdots, Z_{n+m}$中的m个平均数减去剩下的n个的平均数. 那m个则要看试验单元是如何分配的. 这样, 一共能得到$N = \binom{n+m}{m}$ 个值: $g_1, g_2, \cdots, g_N$,将其绝对值按大小排列, 无妨令其为:

$$|g_1| \le |g_2| \le \cdots \le |g_N|, \tag{2.10}$$

检验问题为:

$$H_0 : \text{甲乙两处理效果一样} \tag{2.11}$$

当H_0成立时(2.10)中那N各值中的每一个, 有同等出现的机会$1/N$, 就实际样本算出g之值, 记为$g^* = \bar{X} - \bar{Y}$. 如果H_0不成立, $|g^*|$倾向于取较大之值. 因此, 给定检验水平α后, 找N_α, 使得

$$\alpha = P(|g_i| > |g_{[N_\alpha]}|,\ i \in \{1, \cdots, N\} \,|H_0) \tag{2.12}$$

由于$g_1, g_2, \cdots, g_N$中的每一个发生是等可能$1/N$, 因此(2.12)式等价于找N_α,使得

$$\frac{N - [N_\alpha]}{N} = \alpha$$

因此否定域为

$$D = \{\, |g_{[N_\alpha]+1}|,\ |g_{[N_\alpha]+2}|, \cdots, |g_N| \,\},$$

因此给定样本$\mathbf{X} = (X_1, \cdots, X_m)$和$\mathbf{Y} = (Y_1, \cdots, Y_n)$,算出$g^* = \bar{X} - \bar{Y}$, 若$|g^*| = |\bar{X} - \bar{Y}| > |g_{[N_\alpha]}|$, 则$|g^*|$必在否定域$D$中, 故$D$也可表为

$$D = \{\, (\mathbf{X}, \mathbf{Y}) :\ |g^*| = |\bar{X} - \bar{Y}| > |g_{[N_\alpha]}| \,\}.$$

此处$[N_\alpha]$表示N_α的整数部分.

若改原假设为

$$H_0' : \text{处理甲不优于乙},$$

则将$g_1, g_2, \cdots, g_N$按大小排列为

$$h_1 \le h_2 \le \cdots \le h_N$$

对给定的α,找N_α,使得

$$\alpha = P(h_i > h_{[N_\alpha]},\ i \in \{1, 2, \cdots ., N\} \,|H_0').$$

因此, 一旦有了样本$\mathbf{X} = (X_1, \cdots, X_m)$和$\mathbf{Y} = (Y_1, \cdots, Y_n)$ 算出$g^* = \bar{X} - \bar{Y}$,否定域可表为

$$D = \{(\mathbf{X}, \mathbf{Y}) : g^* = \bar{X} - \bar{Y} > h_{[N_\alpha]}\}.$$

当m、n都很大时, 上述置换检验接近于两样本t检验. 如同在一样本情形, 这个性质从一个侧面加强了两样本t检验的地位.