第二节　秩和检验
Sum of Rank

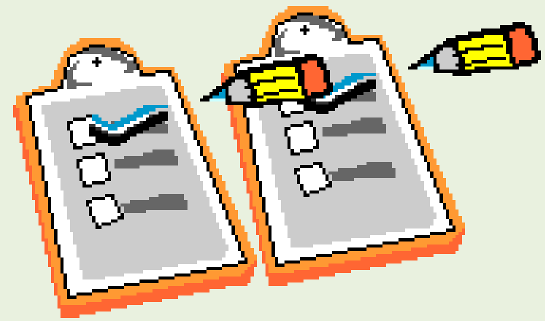
一、基本概念

二、例题选讲

三、小结

一、基本概念

1.假设中的等价问题



设有两个连续型总体,

它们的概率密度函数分别为

$f_1(x), f_2(x),$ (均为未知

已知 $f_1(x) = f_2(x - a)$, a为未知常数

要检验的各假设为

$$\boxed{\begin{aligned} &H_0: a = 0, \quad H_1: a < 0. \\ &H_0: a = 0, \quad H_1: a > 0. \\ &H_0: a = 0, \quad H_1: a \neq 0. \end{aligned}}$$

设两个总体的均值存在,分别记为μ_1, μ_2,

由于f_1, f_2最多只差一平移,则有 $\mu_2 = \mu_1 - a$.

$$\boxed{\begin{array}{ll} H_0: a = 0, & H_1: a < 0. \\ H_0: a = 0, & H_1: a > 0. \\ H_0: a = 0, & H_1: a \neq 0. \end{array}}$$

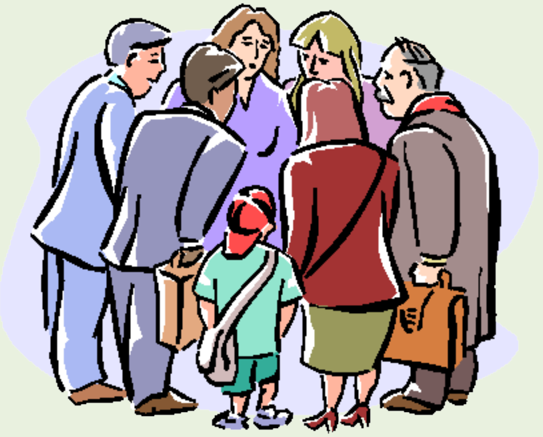
此时,上述各假设分别等价于

$$\boxed{\begin{array}{ll} H_0: \mu_1 = \mu_2, & H_1: \mu_1 < \mu_2. \\ H_0: \mu_1 = \mu_2, & H_1: \mu_1 > \mu_2. \\ H_0: \mu_1 = \mu_2, & H_1: \mu_1 \neq \mu_2. \end{array}}$$

2.秩的定义

设 X 为一总体,将容量为 n 的样本观察值按自小到大的次序编号排列成 $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$,称 $x_{(i)}$ 的足标 i 为 $x_{(i)}$ 的秩, $i=1,2,\cdots,n$.

例如: 某旅行团人员的行李重量数据如表



重量(kg)	34	39	41	28	33

写出重量 33 的秩.

因为 $28<33<34<39<41$, 故 33 的秩为 2 .

特殊情况:

如果在排列大小时出现了相同大小的观察值, 则其秩的定义为**足标的算术平均值**.

例如：抽得的样本观察值按次序排成0,1,1,1,2,3,3,

则3个1的秩均为$\frac{2+3+4}{3}=3$,

两个3的秩均为$\frac{6+7}{2}=6.5$.

3.秩和的定义

现设1, 2 两总体分别抽取容量为n_1, n_2的样本, 且设两样本独立这里总假定$n_1 \le n_2$.

我们将这$n_1 + n_2$个观察值放在一起按自小到大的次序排列求出每个观察值的秩,然后将属于第1个总体的样本观察值的秩相加,其和记为R_1, 称为第1样本的秩和, 其余观察值的秩的总和记作R_2, 称为第2样本的秩和.

我们将这n_1+n_2个观察值放在一起,按自小到大的次序排列,求出每个观察值的秩,

显然, R_1和R_2是离散型随机变量,

且由等差数列的求和公式,我们有

$$R_1+R_2=\frac{1}{2}(n_1+n_2)(n_1+n_2+1).$$

4.秩和检验法的定义

秩和检验法是一种非参数检验法, 它是一种用样本秩来代替样本值的检验法.

用秩和检验法可以检验两个总体的分布函数是否相等的问题.

5. 双边检验 $H_0: a=0, H_1: a\neq 0$ 问题

分析: 当 H_0 为真时, $f_1(x)=f_2(x)$,

此时, 两个独立样本实际上来自同一个总体.

故而第1样本中诸元素的秩应该随机地、分散地在自然数 $1\sim n_1+n_2$ 中取值, 一般来说不应该过分集中取较小的或较大的值.

因而，当 R_1 的值 r_1 过分大或过分小时，我们都拒绝 H_0.

在给定显著性水平 α 下,

H_0 的拒绝域为 $R_1 \le C_U\left(\frac{\alpha}{2}\right)$ 或 $R_1 \ge C_L\left(\frac{\alpha}{2}\right)$,

其中临界点 $C_U\left(\frac{\alpha}{2}\right)$ 是满足 $P_{a=0}\left\{R_1 \le C_U\left(\frac{\alpha}{2}\right)\right\} \le \frac{\alpha}{2}$ 的最大整数,

临界点$C_L\left(\frac{\alpha}{2}\right)$是满足$P_{a=0}\left\{R_1 \geq C_L\left(\frac{\alpha}{2}\right)\right\} \leq \frac{\alpha}{2}$的最小整数

犯第I类弃真错误的概率是

$$P_{a=0}\left\{R_1 \leq C_U\left(\frac{\alpha}{2}\right)\right\}+P_{a=0}\left\{R_1 \geq C_L\left(\frac{\alpha}{2}\right)\right\} \leq \frac{\alpha}{2}+\frac{\alpha}{2}=\alpha.$$

如果知道R_1的分布,则$C_U\left(\frac{\alpha}{2}\right), C_L\left(\frac{\alpha}{2}\right)$不难求得

6.求临界点的方法

以$n_1=3$，$n_2=4$为例

当$n_1=3$，$n_2=4$时,第1个样本中各观察值的秩的不同取法共有$\begin{pmatrix}3+4\\3\end{pmatrix}=35$种,

见下表:

秩	R_1	秩	R_1	秩	R_1	秩	R_1	秩	R_1
123	6	136	10	167	14	247	13	356	14
124	7	137	11	234	9	256	13	357	15
125	8	145	10	235	10	257	14	367	16
126	9	146	11	236	11	267	15	456	15
127	10	147	12	237	12	345	12	457	16
134	8	156	12	245	11	346	13	467	17
135	9	157	13	246	12	347	14	567	18

由于这35种情况的出现是等可能的, 由上表可求得R_1的分布律和分布函数如下表:

R_1的分布律和分布函数如下表

R_1	6	7	8	9	10	11	12
$P\{R_1 = r_1\}$ $P\{R_1 \le r_1\}$	1/35 1/35	1/35 2/35	2/35 4/35	3/35 7/35	4/35 11/35	4/35 15/35	5/35 20/35

R_1	13	14	15	16	17	18	
$P\{R_1 = r_1\}$ $P\{R_1 \le r_1\}$	4/35 24/35	4/35 28/35	3/35 31/35	2/35 33/35	1/35 34/35	1/35 1	

对于不同的α值，

参照上表可以写出双边检验的临界值和拒绝域.

例如，给定$\alpha = 0.2$，　由上表可知　上 表

$$P_{a=0}\{R_1 \le 7\} = \frac{2}{35} = 0.057 < 0.1 = \frac{\alpha}{2},$$

$$P_{a=0}\{R_1 \ge 17\} = \frac{2}{35} = 0.057 < 0.1 = \frac{\alpha}{2},$$

则$C_U(0.1) = 7, C_L(0.1) = 17$. 故当$n_1 = 3$，$n_2 = 4$时，

双边检验 $H_0 : a = 0$，　$H_1 : a \ne 0$的拒绝域：

$$R_1 \le 7 \text{ 或 } R_1 \ge 17.$$

犯第I类错误的概率是

$$P_{a=0}\left\{R_1 \le C_U\left(\frac{\alpha}{2}\right)\right\}+P_{a=0}\left\{R_1 \ge C_L\left(\frac{\alpha}{2}\right)\right\}$$

$$=P_{a=0}\{R_1 \le 7\}+P_{a=0}\{R_1 \ge 17\}$$

$$=0.057+0.057=0.114.$$

7.左边检验和右边检验

左边检验 $H_0: a=0$，$H_1: a<0$ 的拒绝域为

$$r_1 \le C_U(\alpha)，（显著性水平为\alpha）$$

临界点 $C_U(\alpha)$ 是满足 $P_{a=0}\{R_1 \le C_U(\alpha)\} \le \alpha$ 的最大整数.

右边检验 $H_0: a=0$，$H_1: a>0$ 的拒绝域为

$$r_1 \ge C_L(\alpha)，（显著性水平为\alpha）$$

临界点 $C_L(\alpha)$ 是满足 $P_{a=0}\{R_1 \ge C_L(\alpha)\} \le \alpha$ 的最小整数.