基础数理统计（研究生公共课）

肖柳青 博士 教授 主讲

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

第三章 假设检验
Hypothesis Testing

我们将讨论不同于参数估计的另一类重要的统计推断问题. 这就是根据样本的信息检验关于总体的某个假设是否正确. 这类问题称作假设检验问题.

假设检验
- 参数假设检验 —— 总体分布已知, 检验关于未知参数的某个假设
- 非参数假设检验 —— 总体分布未知时的假设检验问题

本章大纲

1. 假设检验的基本概念
2. 单个正态总体参数的假设经验
3. 两个正态总体参数比较的假设经验
4. 最佳检验的概念
5. 分布拟合的假设检验
6. 独立性检验

非参数检验方法

前面的参数假设检验方法是有前提条件的：总体的分布已知，而参数未知。但在实际问题中，常常不能与预知总体的分布函数。

所谓非参数假设检验主要包括：

1、总体分布的假设检验：根据样本来检验关于总体分布的各种假设；

2、样本独立性检验等等

§3.5 分布函数拟合检验（p. 78）

正态概率纸法：p.78

χ^2拟合优度检验:

检验一批分类数据来自总体的分布是否与某种理论分布一致，总体分为r 类，样本 分类数据 判断总体各类出现的概率与已知概率是否相等.

例如检验一颗骰子是否均匀，投掷若干次，记录每面出现的次数，检验各面出现的概率是否为1/6.

一、χ^2 拟合检验法 （p.80）

1. χ^2检验法的定义

这是在总体的分布未知的情况下,根据样本 $X_1, X_2, \cdots, X_n$ 来检验关于总体分布的假设

H_0: 总体 X 的分布函数为 $F(x)$,

H_1: 总体 X 的分布函数不是 $F(x)$,

的一种方法.

说明

(1)*在这里备择假设*H_1*可以不必写出.*

(2) 若总体 X 为离散型：*则上述假设相当于*

H_0：总体 X 的分布律为 $P\{X=t_i\}=p_i, i=1,2,\cdots.$

(3) 若总体 X 为连续型：*则上述假设相当于*

H_0：总体 X 的概率密度为 $f(x)$.

(4) 在使用 χ^2 检验法检验假设 H_0 时，若 $F(x)$ 的形式已知，但其参数值未知，需要先用最大似然估计法估计参数，然后作检验.

2. χ^2检验法的基本思想

将随机试验可能结果的全体 Ω 分为 k 个互不相容的事件 $A_1, A_2, \cdots, A_n$ ($\sum_{i=1}^{k} A_i = \Omega$, $A_i A_j = \varnothing, i \neq j$, $i, j = 1, 2, \cdots, k$). 于是在假设 H_0 下,我们可以计算 $p_i = P(A_i)$ (或 $\hat{p}_i = \hat{P}(A_i)$), $i = 1, 2, \cdots, k$. 在 n 次试验中,事件 A_i 出现的频率 $\dfrac{f_i}{n}$ 与 p_i (或 $\hat{p}_i$) 往往有差异,但一般来说,若 H_0 为真,且试验次数又多时,这种差异不应很大.

一、诸 p_i 均已知

如果H_0成立，则对每一类A_i，其频率n_i/n与概率p_i应较接近。即观测频数n_i与理论频数np_i应相差不大。据此，英国统计学家K.Pearson提出如下检验统计量：

$$\chi^2=\sum_{i=1}^{k}\frac{\left(n_i-np_i\right)^2}{np_i} \qquad (1)$$

并证明在H_0成立时对充分大的n, (1) 给出的检验统计量近似服从自由度为k-1的 χ^2 分布。

拒绝域为：

$$W=\left\{\chi^2\ge\chi^2_{1-\alpha}\left(k-1\right)\right\}$$

二、诸 p_i 不完全已知

若诸 $p_i,i=1,\cdots,k$ $(r<k)$ *个未知参数* $\theta_1,\ldots,\theta_r$ *确定，即* $p_i=p_i(\theta_1,\cdots,\theta_r),\quad i=1,\cdots,k.$

首先给出 $\theta_1,\cdots,\theta_r$ *的极大似然估计* $\hat\theta_1,\cdots,\hat\theta_r,$

然后给出诸 $p_i,i=1,\cdots,k$ *的极大似然估计*

$\hat p_i=p_i(\hat\theta_1,\cdots,\hat\theta_r).$ ***Fisher证明了***

$$\chi^2=\sum_{i=1}^{k}\frac{\left(n_i-n\hat p_i\right)^2}{n\hat p_i}$$

在H_0成立时近似服从自由度为k-r-1的 χ^2 分布，于是检验拒绝域为

$$\left\{\chi^2\ge\chi^2_{1-\alpha}\left(k-r-1\right)\right\}$$

3.皮尔逊定理

一、诸 p_i 均已知

设检验假设 H_0 的统计量为

$$\chi^2=\sum_{i=1}^{k}\frac{n}{p_i}\left(\frac{f_i}{n}-p_i\right)^2\left(\text{或}\chi^2=\sum_{i=1}^{k}\frac{f_i^{\,2}}{np_i}-n\right)$$

二、诸 p_i 不完全已知

定理3.5.2 p.82

若 n 充分大(≥ 50),则当 H_0 为真时(不论 H_0 中的分布属什么分布),上统计量总是近似地服从自由度为 $k-r-1$ 的 χ^2 分布,其中,r 是被估计的参数的个数.

于是，如果在假设 H_0 下，

$$\chi^2=\sum_{i=1}^{k}\frac{(f_i-np_i)^2}{np_i}\geq\chi^2(k-r-1),$$

则在显著性水平 α 下拒绝 H_0，否则就接受 H_0.

注意

在使用 χ^2 检验法时，n 要足够大，np_i 不太小.

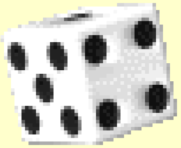
根据实践，一般 $n\geq 50$，每一个 $np_i\geq 5$.

引例1 把一颗骰子重复抛掷 300 次, 结果如下:

出现的点数	1	2	3	4	5	6
出现的频数	40	70	48	60	52	30

试检验这颗骰子的六个面是否匀称? (取 $\alpha = 0.05$)

解 根据题意需要检验假设



H_0: 这颗骰子的六个面是匀称的.

(或 $H_0 : P\{X = i\} = \dfrac{1}{6} \quad (i = 1,2,\cdots,6)$)

其中 X 表示抛掷这骰子一次所出现的点数 (可能值只有 6 个),

取 $\Omega_i = \{i\}, \quad (i = 1,2,\cdots,6)$

则事件 $A_i = \left\{X \in \Omega_i\right\} = \{X = i\} \quad (i = 1,2,\cdots,6)$ 为互不相容事件.

在 H_0 为真的前提下, $p_i = P(A_i) = \dfrac{1}{6}, \quad (i = 1,2,\cdots,6)$

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - np_i)^2}{np_i}$$

$$= \frac{(40 - 300 \times \frac{1}{6})^2}{300 \times \frac{1}{6}} + \frac{(70 - 300 \times \frac{1}{6})^2}{300 \times \frac{1}{6}} + \frac{(48 - 300 \times \frac{1}{6})^2}{300 \times \frac{1}{6}} +$$

$$\frac{(60-300\times\frac{1}{6})^2}{300\times\frac{1}{6}}+\frac{(52-300\times\frac{1}{6})^2}{300\times\frac{1}{6}}+\frac{(30-300\times\frac{1}{6})^2}{300\times\frac{1}{6}},$$

$\chi^2=20.16$, 自由度为$6-1=5$,

查$\chi^2(5)$表得$\chi^2_{0.05}=11.07$, $\chi^2=20.16>11.07$,

所以拒绝 H_0,

认为这颗骰子的六个面不是匀称的.

例2 在一试验中, 每隔一定时间观察一次由某种铀所放射的到达计数器上的 α 粒子数, 共观察了100次, 得结果如下表:

i	0	1	2	3	4	5	6	7	8	9	10	11	≥ 12
f_i	1	5	16	17	26	11	9	9	2	1	2	1	0
A_i	A_0	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	A_{12}

其中 f_i 是观察到有 i 个 α 粒子的次数. 从理论上考虑 X 应服从泊松分布 $P\{X=i\}=\dfrac{e^{-\lambda}\lambda^i}{i!}, i=0,1,2,\cdots,$

问 $P\{X=i\}=\dfrac{e^{-\lambda}\lambda^i}{i!}$ 是否符合实际? $(\alpha=0.05)$

解 所求问题为: 在水平 0.05 下检验假设

H_0: 总体 X 服从泊松分布

$$P\{X=i\}=\frac{e^{-\lambda}\lambda^{i}}{i!},\ i=0,1,2,\cdots,$$

由于在 H_0 中参数 λ 未具体给出, 故先估计 λ.

由最大似然估计法得 $\hat{\lambda}=\bar{x}=4.2,$

根据题目中已知表格, $P\{X=i\}$有估计

$$\hat{p}_i = \hat{P}\{X = i\} = \frac{e^{-4.2}4.2^i}{i!}, i = 0,1,2,\cdots,$$

如 $\hat{p}_0 = \hat{P}\{X = 0\} = e^{-4.2} = 0.015,$

$$\hat{p}_3 = \hat{P}\{X = 3\} = \frac{e^{-4.2}4.2^3}{3!} = 0.185,$$

$$\hat{p}_{12} = \hat{P}\{X \geq 12\} = 1 - \sum_{i=1}^{11}\hat{p}_i = 0.002,$$

具体计算结果见下页表,

例2的 χ^2 拟合检验计算表

A_i	f_i	$\hat{p}_i$	$n\hat{p}_i$	$f_i^2/n\hat{p}_i$
A_0	1 } ***6***	0.015 } ***0.078***	1.5	***4.615***
A_1	5	0.063	6.3	
A_2	16	0.132	13.2	19.394
A_3	17	0.185	18.5	15.622
A_4	26	0.194	19.4	34.845
A_5	11	0.163	16.3	7.423
A_6	9	0.114	11.4	7.105
A_7	9	0.069	6.9	11.739
A_8	2	0.036	3.6	
A_9	1	0.017	1.7	***5.538***
A_{10}	2 } ***6***	0.007 } ***0.065***	0.7	
A_{11}	1	0.003	0.3	$\sum$ ***=106.281***
A_{12}	0	0.002	0.2	

其中有些$n\hat{p}_i < 5$的组予以合并，使得每组均有$np_i \geq 5$，如表中第四列化括号所示.

并组后$k = 8$，故χ^2的自由度为$8-1-1=6$，

$$\chi_\alpha^2(k-r-1) = \chi_{0.05}^2(6) = 12.592 > 6.2815,$$

故接受H_0，认为样本来自泊松分布总体.

例3 自1965年1月1日至1971年2月9日共2231天中，全世界记录到里氏震级4级和4级以上地震共162次，统计如下：

$(\alpha = 0.05)$

(X 表示相继两次地震间隔天数，Y 表示出现的频数)

X	$0-4$	$5-9$	$10-14$	$15-19$	$20-24$	$25-29$	$30-34$	$35-39$	≥ 40
Y	50	31	26	17	10	8	6	6	8

试检验相继两次地震间隔天数 X 服从指数分布.

解 所求问题为：在水平 0.05下检验假设

H_0：X 的概率密度 $f(x)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{x}{\theta}}, & x>0,\\ 0, & x\le 0.\end{cases}$

由于在 H_0 中参数 θ 未具体给出，故先估计 θ.

由最大似然估计法得 $\hat{\theta}=\bar{x}=\dfrac{2231}{162}=13.77$,

X 为连续型随机变量,

将 X 可能取值区间 $[0,+\infty)$ 分为 $k=9$ 个互不重叠的子区间 $[a_i,a_{i+1}), i=1,2,\cdots,9.$ *(见下页表)*

例3的 χ^2 拟合检验计算表

A_i	f_i	$\hat{p}_i$	$n\hat{p}_i$	$f_i^2/n\hat{p}_i$
$A_1: 0 \le x \le 4.5$	50	0.2788	45.1656	55.3519
$A_2: 4.5 < x \le 9.5$	31	0.2196	35.5752	27.0132
$A_3: 9.5 < x \le 14.5$	26	0.1527	24.7374	27.3270
$A_4: 14.5 < x \le 19.5$	17	0.1062	17.2044	16.7980
$A_5: 19.5 < x \le 24.5$	10	0.0739	11.9718	8.3530
$A_6: 24.5 < x \le 29.5$	8	0.0514	8.3268	7.6860
$A_7: 29.5 \le x \le 34.5$	6	0.0358	5.7996	6.2073
$A_8: 34.5 < x \le 39.5$	6	0.0248	4.0176	14.8269
$A_9: 39.5 < x < \infty$	8	0.0568	9.2016	
			13.2192	$\sum$ ***=163.5633***

在 H_0 为真的前提下,

X 的分布函数的估计为 $\hat{F}(x)=\begin{cases}1-e^{-\frac{x}{13.77}}, & x>0,\\ 0, & x\le 0.\end{cases}$

概率 $p_i=P(A_i)$有估计

$$\hat{p}_i=\hat{P}(A_i)=\hat{P}\{a_i\le X<a_{i+1}\}=\hat{F}(a_{i+1})-\hat{F}(a_i),$$

如 $\hat{p}_2=\hat{P}(A_2)=\hat{P}\{4.5\le X<0.5\}$

$$=\hat{F}(9.5)-\hat{F}(4.5)=0.2196,$$

$$\hat{p}_9 = \hat{F}(A_9) = 1 - \sum_{i=1}^{8} \hat{F}(A_i) = 0.0568,$$

$$\chi^2 = 163.5633 - 162 = 1.5633, \qquad k = 8, r = 1,$$

$$\chi^2_\alpha(k - r - 1) = \chi^2_{0.05}(6) = 12.592 > 1.5633,$$

故在水平 0.05 下接受 H_0,

认为样本服从指数分布.

例4 下面列出了84个依特拉斯坎人男子的头颅的最大宽度(mm), 试验证这些数据是否来自正态总体? $(\alpha = 0.1)$

141	148	132	138	154	142	150	146	155	158
150	140	147	148	144	150	149	145	149	158
143	141	144	144	126	140	144	142	141	140
145	135	147	146	141	136	140	146	142	137
148	154	137	139	143	140	131	143	141	149
148	135	148	152	143	144	141	143	147	146
150	132	142	142	143	153	149	146	149	138
142	149	142	137	134	144	146	147	140	142
140	137	152	145						

解 所求问题为检验假设

H_0： X 的概率密度 $f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<\infty.$

由于在 H_0 中参数 μ,σ^2 未具体给出,故先估计μ,σ^2.

由最大似然估计法得 $\hat{\mu}=143.8,\quad \hat{\sigma}^2=6.0^2,$

将 X 可能取值区间 $(-\infty,\infty)$ 分为7个小区间,

(见下页表)

例4的 χ^2 拟合检验计算表

A_i	f_i	$\hat{p}_i$	$n\hat{p}_i$	$f_i^2 / n\hat{p}_i$
$A_1: x \le 129.5$	1	0.0087	0.73	
$A_2: 129.5 < x \le 134.5$	4	0.0519	4.36 (合并 **5.09**)	**4.91**
$A_3: 134.5 < x \le 139.5$	10	0.1752	14.72	6.79
$A_4: 139.5 < x \le 144.5$	33	0.3120	26.21	41.55
$A_5: 144.5 < x \le 149.5$	24	0.2811	23.61	24.40
$A_6: 149.5 < x \le 154.5$	9	0.1336	11.22	10.02
$A_7: 154.5 < x < \infty$	3	0.0375	3.15 (合并 **14.37**)	Σ =87.67

在 H_0 为真的前提下，X 的概率密度的估计为

$$\hat{f}(x)=\frac{1}{\sqrt{2\pi}\times 6}\mathrm{e}^{-\frac{(x-143.8)^2}{2\times 6^2}},-\infty<x<+\infty.$$

概率 $p_i=P(A_i)$有估计

如 $\hat{p}_2=\hat{P}(A_2)=\hat{P}\{129.5\le x<134.5\}$

$$=\Phi\left(\frac{134.5-143.8}{6}\right)-\Phi\left(\frac{129.5-143.8}{6}\right)$$

$$=\Phi(-1.55)-\Phi(-2.38)=0.0519\ .$$

$$\chi_{\alpha}^2(k-r-1)=\chi_{0.1}^2(5-2-1)=\chi_{0.1}^2(2)=4.605>3.67,$$

故在水平 0.1 下接受 H_0, 认为样本服从正态分布.

例5 一农场10年前在一鱼塘里按如下比例 20 : 15 : 40 : 25 投放了四种鱼: 鲑鱼、鲈鱼、竹夹鱼和鲇鱼的鱼苗. 现在在鱼塘里获得一样本如下:

序号	1	2	3	4	
种类	鲑鱼	鲈鱼	竹夹鱼	鲇鱼	
数量(条)	132	100	200	168	$\Sigma = 600$

检验各鱼类数量的比例较 10 年前是否有显著改变?

(取 $\alpha = 0.05$)

解 用 X 记鱼种类的序号,

根据题意需检验假设:

$H_0: X$ 的分布律为

X	1	2	3	4
p_i	0.20	0.15	0.40	0.25

所需计算列表如下 $(n=600)$

例5 的 χ^2 拟合检验计算表

A_i	f_i	p_i	np_i	$f_i^2/n\hat{p}_i$
A_1	132	0.20	120	145.20
A_2	100	0.15	90	111.11
A_3	200	0.40	240	166.67
A_4	168	0.25	150	188.16
				$\sum=611.14$

$$\chi^2 = 611.14 - 600 = 11.14, \quad k = 4, r = 0,$$

$\chi^2_{0.05}(k-r-1) = \chi^2_{0.05}(3) = 7.815 < 11.14$，故拒绝 H_0，

认为各鱼类数量之比较 10 年前有显著改 变 .

例6　为募集社会福利基金，某地方政府发行福利彩票，中彩者用摇大转盘的方法确定最后中奖金额，大转盘均分为20份，其中金额为5万、10万、20万、30万、50万、100万的分别占2份、4份、6份、4份、2份、2份。假定大转盘是均匀的，则每一点朝下是等可能的，于是摇出各个奖项的概率如下：

额度	*5万*	*10万*	*20万*	*30万*	*50万*	*100万*
概率	*0.1*	*0.2*	*0.3*	*0.2*	*0.1*	*0.1*

现20人参加摇奖，摇得5万、10万、20万、30万、50万、100万的人数分别为2、6、6、3、3、0，由于没有一个人摇到100万，于是有人怀疑大转盘是不均匀的，那么该怀疑是否成立？

解　这是一个典型的分布拟合优度检验，总体共有6类，其发生概率分别为0.1、0.2、0.3、0.2、0.1和0.1。

这里k＝6，检验拒绝域为{ $\chi^2 \ge \chi^2_{1-\alpha}(5)$}

若取 $\alpha = 0.05$，则查表知 $\chi^2_{0.95}(5) = 11.07$

由数据可以算出

$$\chi^2 = \frac{(2-2)^2}{2} + \frac{(6-4)^2}{4} + \frac{(6-6)^2}{6} + \frac{(3-4)^2}{4} + \frac{(3-2)^2}{2} + \frac{(0-2)^2}{2}$$

$$= 3.75$$

由于$\chi^2 = 0.75$ 未落入拒绝域，故接受原假设，没有理由认为转盘不均匀。

二. 正态性检验

1、正态概率纸(孙荣恒p. 138)

正态概率纸可用来做正态性检验，方法如下：利用样本数据在概率纸上的描点，用目测方法看这些点是否在一条直线附近，若是的话，可以认为该数据来自的总体是正态分布，若明显不在一条直线附近，则认为该数据来自非正态总体。

2、 Kolmogrov-Smirnov 检验*孙荣恒p.147

总体分布连续时，χ^2 拟合优度检验只是检验了 $\hat{p}_i = P(A_i) = P(a_{i-1} < X \le a_i) = \hat{F}(a_i) - \hat{F}(a_{i-1})$，

$i = 1, 2, \cdots, r$ 是否为真，并未真正检验总体分布是否为 $F(x)$。Kolmogrov-Smirnov 检验比 χ^2 拟合优度检验更为精确，既可检验经验分布是否佛服从某种理论分布（Kolmogrov 检验），又可检验两个样本是否来自同一总体（Smirnov 检验）。

1. Kolmogrov 检验　　检验经验分布是否佛服从某种理论分布

总体分布函数 $F(x)$ 连续，样本 $(X_1, X_2, \cdots, X_n)$，经验分布函数 $F_n(x)$。格里汶科定理

$$P(\lim_{n\to 0} \sup_x |F_n(x) - F(x)| = 0) = 1$$

随机变量 $D_n = \sup_x |F_n(x) - F(x)|$ 以概率 1 收敛到 0。

假设检验：$H_0 : F(x) = F_0(x) \leftrightarrow H_1 : F(x) \ne F_0(x)$, 至少存在某一点$x$使得不等号成立

统计量 $D_n = \sup_x |F_n(x) - F_0(x)|$

H_0 不真时，D_n 有偏大的趋势，拒绝域 $W = \{D_n \ge D_{n,\alpha}\}$，其中 $D_{n,\alpha}$ 是临界值

若 n 较大时（$n > 100$），$D_{n,\alpha} \approx \dfrac{\lambda_{1-\alpha}}{\sqrt{n}}$。

例 5 某矿区煤层厚度的 123 个数据的频数分布表如下所示，试用 Kolmogrov 法检验煤层厚度是否服从正态分布。

组号	厚度间隔 $(a_{i-1}, a_i]$	组中值 x_i	频数 n_i	组号	厚度间隔 $(a_{i-1}, a_i]$	组中值 x_i	频数 n_i
1	0.20–0.50	0.35	1	6	1.70–2.00	1.85	24
2	0.50–0.80	0.65	6	7	2.00–2.30	2.15	25
3	0.80–1.10	0.95	5	8	2.30–2.60	2.45	19
4	1.10–1.40	1.25	12	9	2.60–2.90	2.75	20
5	1.40–1.70	1.55	19	10	2.90–3.20	3.05	2

解：用 X 表示煤层厚度， $\hat{\mu} = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} n_i x_i = 1.884$， $\hat{\sigma}^2 = s_n^2 = \frac{1}{n}\sum_{i=1}^{n} n_i (x_i - \overline{x})^2 = 0.576^2$

假设检验 H_0： 总体 X 服从正态分布 $N(1.884, 0.576^2)$

$$\hat{p}_i = P(A_i) = P(a_{i-1} < X \le a_i) = \hat{F}(a_i) - \hat{F}(a_{i-1})$$

组号	厚度间隔	频率 $\frac{n_i}{n}$	组上限 x	$u=\frac{x-1.884}{0.576}$	经验分布函数值 $F_n(x)$	理论分布函数值 $F_0(x)$	$\lvert F_n(x_i)-F_0(x_i)\rvert$
1	0. 20−0. 50	0. 008	0. 50	−2. 40	0. 008	0. 0082	0. 0002
2	0. 50−0. 80	0. 049	0. 80	−1. 88	0. 057	0. 0301	0. 0269
3	0. 80−1. 10	0. 041	1. 10	−1. 36	0. 098	0. 0869	0. 0111
4	1. 10−1. 40	0. 098	1. 40	−0. 84	0. 196	0. 2005	0. 0045
5	1. 40−1. 70	0. 154	1. 70	−0. 32	0. 350	0. 3745	0. 0245
6	1. 70−2. 00	0. 195	2. 00	0. 20	0. 545	0. 5793	0. 0343
7	2. 00−2. 30	0. 203	2. 30	0. 72	0. 748	0. 7642	0. 0162
8	2. 30−2. 60	0. 154	2. 60	1. 24	0. 902	0. 8925	0. 0095
9	2. 60−2. 90	0. 081	2. 90	1. 76	0. 983	0. 9608	0. 0222
10	2. 90−3. 20	0. 017	3. 20	2. 28	1. 000	0. 9887	0. 0113

其中 $F(x)=P(X\le x)=P\left(\frac{X-1.884}{0.576}\le\frac{x-1.884}{0.576}\right)=\Phi\left(\frac{x-1.884}{0.576}\right)$，$\hat{D}_n=0.0343$

$n>100$ 时，$\alpha=0.05$，查表得 $\lambda_{1-\alpha}=1.36$，$D_{n,\alpha}\approx\frac{\lambda_{1-\alpha}}{\sqrt{n}}=\frac{1.36}{\sqrt{123}}=0.123$

$\hat{D}_n<D_{n,\alpha}$，接受原假设 H_0，即认为煤层厚度服从正态分布。

注：①分组列表统计得到的是$\hat{D}_n$的近似值，该值偏小。当$\hat{D}_n$与$D_{n,\alpha}$的值很接近时，接受H_0的结论需要慎重考虑；

②n充分大时，$P\left(D_n \le \frac{\lambda_{1-\alpha}}{\sqrt{n}}\right) \approx 1-\alpha$，$F(x)$的$1-\alpha$置信区间为：$F_n(x)-\frac{\lambda_{1-\alpha}}{\sqrt{n}} \le F_0(x) \le F_n(x)+\frac{\lambda_{1-\alpha}}{\sqrt{n}}$。

三. 两总体之间关系的假设检验

Smirnov检验 比较两个总体的真实分布是否相同

连续型总体 $X \sim F(x), Y \sim G(y)$，样本 $(X_1, \cdots, X_n)$，$(Y_1, \cdots, Y_m)$，两样本相互独立

假设检验 $H_0 : F(x) = G(x) \leftrightarrow H_1 : F(x) \neq G(x)$, 至少存在某一点$x$使得不等号成立

总体 X, Y 的经验分布函数分别为 $F_n(x)$ 和 $G_n(x)$，统计量 $D_{n,m} = \sup\limits_x |F_n(x) - G_m(x)|$

令 $n_1 = \dfrac{nm}{n+m}$，拒绝域 $W = \{D_{n,m} \geq D_{n_1,\alpha}\}$

若 n, m 较大时（$n, m > 100$），$D_{n_1,\alpha} \approx \dfrac{\lambda_{1-\alpha}}{\sqrt{n_1}}$。

例 6 自动车床加工某一种零件。工人刚接班时，抽取 $n=150$ 只零件作为一个样本，自动车床工作 2 小时后，再抽取 $m=100$ 只零件作为第二个样本，测定每个零件距离标准的偏差 X，列表如下：

偏差 X 的测量区间	样本 1 的频数	样本 2 的频数
$[-15,-10)$	10	0
$[-10,-5)$	27	7
$[-5,0)$	43	17
$[0,5)$	38	30
$[5,10)$	23	29
$[10,15)$	8	15
$[15,20)$	1	1
$[20,25)$	0	1
合计	150	100

试问此两样本是否来自同一总体？ $\alpha=0.05$

解：$H_0: F(x)=G(x) \leftrightarrow H_1: F(x)\neq G(x)$

组上限 x	样本 1 累积频数	样本 2 累积频数	$F_n(x)=\dfrac{n_i}{n}$	$G_m(x)=\dfrac{m_i}{m}$	$\lvert F_n(x)-G_m(x)\rvert$
−10	10	0	0.067	0.000	0.067
−5	37	7	0.247	0.070	0.177
0	80	24	0.533	0.240	0.293
5	118	54	0.787	0.540	0.247
10	141	83	0.940	0.830	0.110
15	149	98	0.993	0.980	0.013
20	150	99	1.00	0.990	0.010
25	150	100	1.00	1.000	0.000

$D_{n,m}=\sup\limits_{x}\lvert F_n(x)-G_m(x)\rvert=0.293$。

$n_1=\dfrac{nm}{n+m}=\dfrac{150\times100}{150+100}=60$，$\alpha=0.05$，$D_{n_1,\alpha}=0.17231$，$D_{n,m}>D_{n_1,\alpha}$，故拒绝 H_0，

即在自动机床上加工零件时，不能忽视时间延续的影响，最好能找出合适的时间间隔作定时调整。

§3.6 独立性检验 (p.87)

列联表的独立性检验

列联表是将观测数据按两个或更多属性（定性变量）分类时所列出的频数表。

抽取 n 个样本，样品按不同特性分类。两个特性：X_1，r 类；X_2，c 类

二维列联表 $(r\times c)$　频数表

$X_1 \setminus X_2$	B_1	B_2	$\cdots$	B_c	合计
A_1	n_{11}	n_{12}	$\cdots$	n_{1c}	$n_{1\cdot}$
A_2	n_{21}	n_{22}	$\cdots$	n_{2c}	$n_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
A_r	n_{r1}	n_{r2}	$\cdots$	n_{rc}	$n_{r\cdot}$
合计	$n_{\cdot 1}$	$n_{\cdot 2}$	$\cdots$	$n_{\cdot c}$	n

总体 $X=(X_1,X_2)$，　X_1 与 X_2 是否独立？

$$P(X\in A_i\cap B_j)=P(\{X_1\in A_i\}\cap\{X_2\in B_j\})=p_{ij}\text{，}\quad i=1,2,\cdots,r\text{，}\quad j=1,2,\cdots,c$$

$$p_{i\cdot}=P(X_1\in A_i)=\sum_{j=1}^{c}p_{ij}\text{，}\quad i=1,2,\cdots,r\text{；}\quad p_{\cdot j}=P(X_2\in B_j)=\sum_{i=1}^{r}p_{ij}\text{，}\quad j=1,2,\cdots,c$$

$$\sum_{i=1}^{r}p_{i\cdot}=\sum_{j=1}^{c}p_{\cdot j}=1$$

若 X_1 与 X_2 相互独立，则 $\forall i,j$， $p_{ij}=p_{i\cdot}\cdot p_{\cdot j}$

独立性检验： H_0： $p_{ij}=p_{i\cdot}\cdot p_{\cdot j}$， $\forall i,j$ $\quad H_1$： $p_{ij}\neq p_{i\cdot}\cdot p_{\cdot j}$， 至少有一对 (i,j)

$$\chi^2=\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij}-np_{ij})^2}{np_{ij}}=\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij}-np_{i\cdot}p_{\cdot j})^2}{np_{i\cdot}p_{\cdot j}}\text{，}\quad H_0\text{ 为真}$$

$p_{i\cdot},p_{\cdot j}$ 的 MLE： $\hat{p}_{i\cdot}=\dfrac{n_{i\cdot}}{n}$， $i=1,2,\cdots,r$； $\hat{p}_{\cdot j}=\dfrac{n_{\cdot j}}{n}$， $j=1,2,\cdots,c$

检验统计量： $\chi^2=\displaystyle\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij}-n\hat{p}_{i\cdot}\hat{p}_{\cdot j})^2}{n\hat{p}_{i\cdot}\hat{p}_{\cdot j}}$， $n\hat{p}_{i\cdot}\hat{p}_{\cdot j}=n\cdot\dfrac{n_{i\cdot}}{n}\cdot\dfrac{n_{\cdot j}}{n}=\dfrac{n_{i\cdot}n_{\cdot j}}{n}$

$r+c$ 个未知参数 $\hat{p}_{i\cdot},\hat{p}_{\cdot j}$， $\displaystyle\sum_{i=1}^{r}\hat{p}_{i\cdot}=1$， $\displaystyle\sum_{j=1}^{c}\hat{p}_{\cdot j}=1$， $r+c-2$ 个独立参数

rc 个自由参数 n_{ij}， 自由变化的 $\hat{p}_{i\cdot}, \hat{p}_{\cdot j}$ 共 $r+c-2$ 个，它们由 n_{ij} 表示，再加上约束

$\sum_{i=1}^{r}\sum_{j=1}^{c} n_{ij} = n$，则自由度：$rc-(r+c-2)-1 = rc-(r+c)+1 = (r-1)(c-1)$

H_0 为真，n 较大时，$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij}-n\hat{p}_{i\cdot}\hat{p}_{\cdot j})^2}{n\hat{p}_{i\cdot}\hat{p}_{\cdot j}} = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij}-n_{i\cdot}n_{\cdot j}/n)^2}{n_{i\cdot}n_{\cdot j}/n} \sim \chi^2\left((r-1)(c-1)\right)$，$n\hat{p}_{i\cdot}\hat{p}_{\cdot j} \geq 5$

给定显著性水平 α，拒绝域 $W = \left\{\chi^2 \geq \chi^2_{1-\alpha}\left((r-1)(c-1)\right)\right\}$

特例 若 $r=c=2$，可证明检验统计量 $\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(n_{ij}-n_{i\cdot}n_{\cdot j}/n)^2}{n_{i\cdot}n_{\cdot j}/n} = \frac{n(n_{11}n_{22}-n_{12}n_{21})^2}{n_{1\cdot}n_{2\cdot}n_{\cdot 1}n_{\cdot 2}}$，拒绝域

$W = \left\{\chi^2 \geq \chi^2_{1-\alpha}(1)\right\}$

例 7 调查 339 名 50 岁以上的吸烟者与慢性气管炎的关系，结果如下：

	患慢性气管炎者	未患慢性气管炎者	合计	患病率
吸烟	43	162	205	20. 1%
不吸烟	13	121	134	9. 7%
合计	56	283	339	16. 5

$\alpha = 0.01$ 水平上检验慢性气管炎的患病率与吸烟是否有关？

解： $r = 2$， $c = 2$， $\alpha = 0.01$， $\chi^2_{0.99}\left((2-1)(2-1)\right) = \chi^2_{0.99}(1) = 6.635$， $W = \left\{\chi^2 \geq 6.635\right\}$

$n\hat{p}_{i\cdot}\hat{p}_{\cdot j}$	患慢性气管炎者	未患慢性气管炎者	合计
吸烟	33. 86	171. 14	205
不吸烟	22. 14	111. 86	134
合计	56	283	339

$$\chi^2 = \frac{(43-33.86)^2}{33.86} + \frac{(162-171.14)^2}{171.14} + \frac{(13-22.14)^2}{22.14} + \frac{(121-111.86)^2}{111.86} = 7.48 > 6.635$$

（ $\chi^2 = \dfrac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1\cdot}n_{2\cdot}n_{\cdot 1}n_{\cdot 2}} = \dfrac{339\times(43\times121-162\times13)^2}{205\times134\times56\times283} = 7.47 > 6.635$ ）

所以拒绝原假设 H_0， 即在 $\alpha = 0.01$ 水平上认为慢性气管炎的患病率和吸烟有关。

例8 **为研究儿童智力发展与营养的关系，某研究机构调查了1436名儿童，得到下面的数据，试在显著性水平0.05下判断智力发展与营养有无关系。**

	智商				合计
	<80	80~89	90~99	>=100	
营养良好	367	342	266	329	1304
营养不良	56	40	20	16	132
合计	423	382	286	345	1436

解　用A表示营养状况，它有两个水平：A_1表示营养良好，A_2表示营养不良；B表示儿童智商，它有四个水平，B_1、B_2、B_3、B_4分别表示表中四种情况。沿用前面的记号，首先建立假设H_0：营养状况与智商无关联，即A与B是独立的。统计表示如下：

$$H_0: p_{ij} = p_{i\bullet} p_{\bullet j}, i=1,2, j=1,2,3,4$$

在原假设H_0成立下，可以计算诸参数的最大似然估计值，

$$\hat{p}_{1\bullet}=1304/1436=0.9081, \hat{p}_{2\bullet}=132/1436=0.0919$$

$$\hat{p}_{\bullet 1}=423/1436=0.2946, \hat{p}_{\bullet 2}=382/1436=0.2660$$

$$\hat{p}_{\bullet 3}=286/1436=0.1992, \hat{p}_{\bullet 4}=345/1436=0.2403$$

进而可给出诸$n\hat{p}_{ij}=n\hat{p}_{i\bullet}\hat{p}_{\bullet j}$，如

$$n\hat{p}_{11}=1436\times 0.9081\times 0.2496=384.1677$$

所以计算检验统计量的值

$$\chi^2=\frac{(367-384.1677)^2}{384.1677}+\frac{(342-346.8724)^2}{346.8724}+\cdots+\frac{(16-31.7120)^2}{31.7120}$$
$$=19.2785$$

此处**r=2**，**c=4**，**(r-1)(c-1)=3**，若取 $\alpha=0.05$，查表有 $\chi^2_{0.95}(3)=7.815$，由于**19.2785>7.815**，故拒绝原假设，认为营养状况对智商有影响。本例中检验的 **p** 值为**0.0002**。

§3.7 特殊差异性检验 *(孙荣恒p.153)

常用的方法：

1、符号检验法
2、符号秩和检验法

1、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

$$H_0: \text{甲乙两种酒一样好} \tag{1.2}$$

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 $b(n, p)$, $0 \leq p \leq 1$,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

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

$X = n_+$的具体值为x_0,记$x_0' = min\{x_0, n - x_0\}$,则检验的$p$值为

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

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:

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

设想请了13个人品尝甲、乙两种酒, 评分结果如下:

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

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$D = \{X = n_+ \geq 10,\ 或\ X \leq 2\}$.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.

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$D = \{n_+ \ge c$ 或 $n_+ \le d\}$.

因此, 在给定显著性水平α之后, c和d的值由

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

所确定.

在本例中$n=11$, $\alpha=0.10$, 查二项分布表知

$$\sum_{k=0}^{2}\binom{11}{k}\left(\frac{1}{2}\right)^{11}=0.0327,$$

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

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

$$\{n_+\leq 2 \text{ 或 } n_+\geq 9\}$$

作差值$z_i=x_i-y_i$,得

$$0.15,\quad -0.04,\quad -0.14,\quad 0.02,\quad -0.06,\quad -0.05,$$

$$-0.03,\quad -0.01,\quad -0.11,\quad -0.07,\quad -0.12,$$

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.

符号检验用于分位数（中位数）检验

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: $X\sim b(100,p)$, $0\le p\le 1$,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$D=\{X\ge c_2 \text{ 或 } X\le c_1\}$. 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

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

查表得$(c_1-50)/5=-1.96$, 解得$c_1=40.2$

类似地由

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

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

$$\{X : X \le 40.2 \text{ 或 } X \ge 59.8\}$$

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检验的关系（孙荣恒p.156）

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

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

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

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

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

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

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

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

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

2、符号秩和检验法

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

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

表 1.3

品酒人(i)	甲(x_i)	乙(y_i)	符号(z_i)	$\lvert Z_i\rvert = \lvert x_i - y_i\rvert$	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

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

$$H_0: \text{甲、乙两酒一样好}$$

成立时, W^+应当不大不小. 检验的否定域是

$$\{W^+ \leq d \text{ 或 } W^+ \geq c\}, \tag{1.8}$$

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:

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

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

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

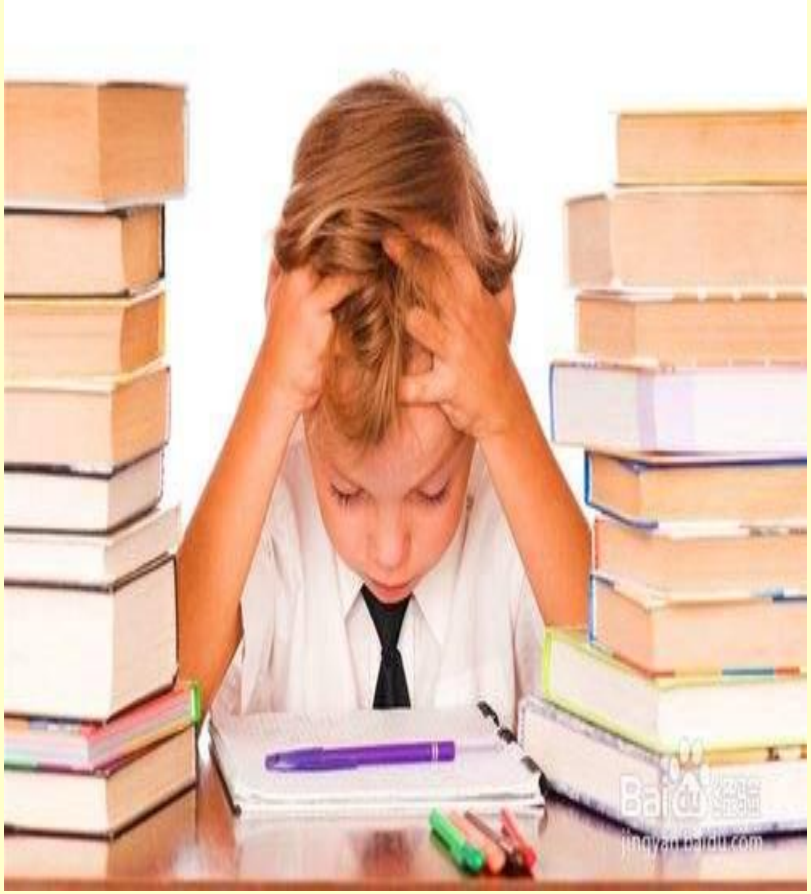
故例6.2.5的水平近似为α的双侧W^+检验的否定域为

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

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$H_0$.

第9次作业：

- 电子书上p.88 习题三
- 14、15、16、17、18、19、20.



谢 谢！