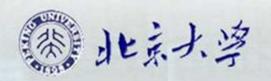


# 秩方差分析

主讲人: 刘宏志

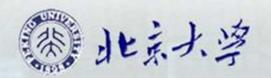
liuhz@ss.pku.edu.cn





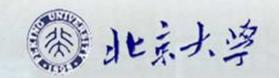
### Kruskal-Wallis秩和检验

- · 简称K-W检验, 也称H检验
- 由Kruskal和Wallis二人在1952年提出
- 一种单因素方差分析方法
- 将两个独立样本的秩和检验推广到3个或更 多组的检验
- 基本原理:
  - ▶若各组的处理效应相同,则混合编秩后,各组的平均秩应近似相等



# 单因素方差分析的数据结构

	因素(A)					
<b>-</b>	水平 $A_1$	水平 $A_2$	• • •	水平 $A_K$		
重 复	$x_{11}$	$x_{21}$	•••	$x_{k1}$		
观	$x_{12}$	$x_{22}$	•••	$x_{k2}$		
察量	:	:	:	:		
	:	:	:	:		
	$x_{1n_1}$	$x_{1n_2}$	•••	$x_{1n_k}$		

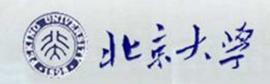




# 单因素方差分析的秩矩阵

	因素(A)					
重	水平A <sub>1</sub>	水平 $A_2$	• • •	水平 $A_K$		
复	$R_{11}$	$R_{21}$	•••	$R_{k1}$		
观 察	$R_{12}$	$R_{22}$	•••	$R_{k2}$		
量 的	:	<b>:</b>	:	<b>:</b>		
秩	:	:	:	•		
	$R_{1n_1}$	$R_{2n_2}$	•••	$R_{k n_k}$		
秩和	$R_1$ .	$R_2$ .	•••	$R_k$ .		

各组的平均秩:  $\bar{R}_{i\cdot} = R_{i\cdot}/n_i$  混合后的平均秩:  $\bar{R}_{\cdot\cdot} = (N+1)/2$ 



# 检验统计量: H值

$$H = \sum_{i=1}^{k} \frac{(R_{i.} - n_{i} \overline{R}_{..})^{2}}{n_{i} S^{2}}$$

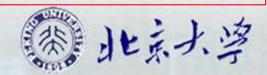
$$S^{2} = \frac{1}{N-1} \sum_{i=1}^{k} \sum_{j=1}^{n_{i}} (R_{ij} - \overline{R}_{..})^{2} = \frac{1}{N-1} \sum_{i=1}^{k} \sum_{j=1}^{n_{i}} R_{ij}^{2} - N \overline{R}_{..}^{2}$$

如果没有同秩(tie)现象,则有

$$S^{2} = \frac{1}{N-1} \left[ \frac{N(N+1)(2N+1)}{6} - \frac{N(N+1)^{2}}{4} \right] = \frac{N(N+1)}{12}$$

$$H = \frac{12}{N(N+1)} \sum \frac{R_i^2}{n_i} - 3(N+1)$$

当混合编秩过程中出现同秩现象时,需要对H统计量进行校正





# K-W检验:一般步骤

#### 1.建立假设、确定检验水准

 $H_0$ : 各样本分布位置相同

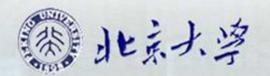
 $H_1$ : 各样本分布位置不全相同

 $\alpha = 0.05$ 

#### 2. 选择检验方法、计算统计量

- (1)混合编秩
- (2)求秩和 $(R_i)$ 和统计量H值:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{\infty} \frac{R_i^2}{n_i} - 3(N+1)$$



#### 3.确定P值,作出推断结论

#### (1)小样本情况:

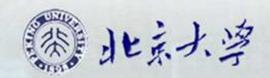
当组数 $k \leq 3$ ,且 $n_i < 5$ 时,查H界值表,确定P值。

如果 $H > H_{\alpha}$ , 则 $P < \alpha$ ; 反之, $P > \alpha$ 。

#### (2)大样本情况:

若k>3或 $n_i>5$ 时,理论上,H近似服从自由度为k-1的 $\chi^2$ 分布,可查 $\chi^2$ 界值表确定P值。

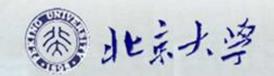
如果 $H > H_{\alpha}$ , 则 $P < \alpha$ ; 反之, $P > \alpha$ 。



# 示例: 杀灭钉螺的药物比较

### 三种药物杀灭钉螺的死亡率(%)比较

甲药		乙数		丙药	 丙药		
死亡率	 秩	死亡率	 秩	死亡率	<del></del>		
32.5	10	16.0	4	6.5	1		
35.5	11	20.5	6	9.0	2		
40.5	13	22.5	7	12.5	3		
46.0	14	29.0	9	18.0	5		
49.0	15	36.0	12	24.0	8		
$R_i$	63		38		19		
$n_i$	5	_	5		5		



### 1.建立假设、确定检验水准

Ho: 三种药物杀灭钉螺的死亡率总体分布位置相同

 $H_{\parallel}$ : 三种药物杀灭钉螺的死亡率总体分布位置不全相同

$$\alpha = 0.05$$

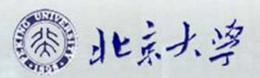
2. 求检验统计量H值:  $H = \frac{12}{N(N+1)} (\sum \frac{R_i^2}{n_i}) - 3(N+1)$ 

$$H = \frac{12}{15(15+1)} \left(\frac{63^2 + 38^2 + 19^2}{5}\right) - 3(15+1) = 9.74$$

### 3.确定P值,作出推断结论

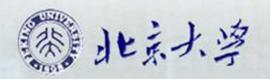
$$\chi^2_{0.05}(2)=5.99 < H$$

拒绝H<sub>0</sub>,接受H<sub>1</sub>,认为三种药物的效果不同



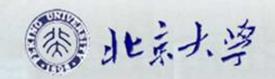
# 多个相关样本的比较: M检验

- Friedman M检验,用于推断随机区组设计的 多个相关样本所来自的多个总体分布的位 置是否有差别
- 检验假设与备择假设和多个独立样本比较的Kruskal-Wallis 检验相同



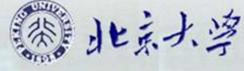
# 多个相关样本的比较: M检验

受试对象( <i>j</i> )-		因素(A)	) j	
	水平 $A_1$	水平 $A_2$	• • •	水平AK
1	$X_{11}$	$X_{21}$	• • •	$X_{k1}$
2	$X_{12}$	$X_{22}$	•••	$X_{k2}$
:	:	:	:	:
:	:	:	•	:
п	$X_{1n}$	$X_{2n}$	•••	$X_{kn}$



# 多个相关样本的比较: M检验

受试对象( <i>j</i> ) -		因素(A)	) i	
	水平 $A_1$	水平A2	•••	水平AK
1	$R_{11}$	$R_{21}$	• • •	$R_{k1}$
2	$R_{12}$	$R_{22}$	•••	$R_{k2}$
:	•	:	:	:
:	:	:	:	:
n	$R_{1n}$	$R_{2n}$	•••	$R_{k n}$
<b>秩和</b>	$R_1$ .	$R_2$ .	•••	$R_{k}$
对每个受试对象	在不同水平-	下的观测值{R	1i, R <sub>2i</sub> ,	, R <sub>i</sub> }独立编科



### 8名受试对象对4种不同频率声音刺激的反应率(%)

受试号·	频率 A		频率 B		频率 C		频率 D	
	反应率	秩	反应率	秩	反应率	秩	反应率	秩
1	8.4	1	9.6	2	9.8	3	11.7	4
2	11.6	1	12.7	4	11.8	2	12.0	3
3	9.4	2	9.1	1	10.4	4	9.8	3
4	9.8	2	8.7	1	9.9	3	12.0	4
5	8.3	2	8.0	1	8.6	3.5	8.6	3.5
6	8.6	1	9.8	3	9.6	2	10.6	4
7	8.9	1	9.0	2	10.6	3	11.4	4
8	7.8	1	8.2	2	8.5	3	10.8	4
$\overline{R_i}$		11		16		23.5		29.5

### M统计量

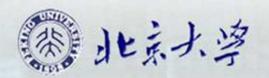
$$\overline{R}_{i\cdot} = \frac{1}{n} \sum_{j=1}^{n} R_{ij}$$

$$\overline{R}_{\cdot\cdot} = \frac{1}{nk} \sum_{j=1}^{n} \sum_{i=1}^{k} R_{ij}$$

$$SS_{t} = n \sum_{i=1}^{k} \left( \overline{R}_{i\cdot} - \overline{R}_{\cdot\cdot} \right)^{2}$$

$$M = \frac{12n}{k(k+1)} \sum_{i=1}^{k} \left( \overline{R}_{i\cdot} - \overline{R}_{\cdot\cdot} \right)^{2}$$

M服从什么分布?





# M检验:一般步骤

#### 1.建立假设

 $H_0$ : 各样本分布位置相同

 $H_1$ : 各样本分布位置不全相同

#### 2. 计算检验统计量

(1)每个对象独立编秩

(2)求统计量M值: 
$$M = \frac{12n}{k(k+1)} \sum_{i=1}^{k} \left(\overline{R}_{i\cdot} - \overline{R}_{\cdot\cdot}\right)^{2}$$

#### 3.确定P值,作出推断结论

- (1) 根据检验水平 $\alpha$ 和自由度k-1查 $\chi^2$ 分布表,确定临界值
- (2) 比较临界值和统计量值,做出结论

