

复习课

考试和答疑时间

- 考试时间：12月20日三四节课（暂定）
 - 答疑时间：12月15日 12:30-14:00
-

题型

- 1、选择
 - 2、填空
 - 3、简答
 - 4、综合计算
-

第 1 章

1、非参数统计和参数统计的区别和联系。

2、秩检验统计量的性质

P22 定理1.3 推论1.2 推论1.3的结论

3、有结数据的秩的处理P23

4、p值的概念 P6

5、秩的概念，秩方法的思想

哪些检验是秩方法检验？

第 1 章

设样本 x_1, x_2, \dots, x_n 来自连续型总体 X , 记 x_i 的秩为 $R_i (i = 1, 2, \dots, n)$,

则 R_i 的分布列为_____, $i \neq j$ 时,

R_i, R_j 的联合分布列为_____

且 $E(R_i) =$ _____, $D(R_i) =$ _____, $\text{Cov}(R_i, R_j) =$ _____

第 1 章

2、在观察值中有相等的值(即有结) 时, 用_____方法确定结里观察值的秩.

例 样本数据有 12 个数, 其值, 秩和结统计量 (用 τ_i 表示, 为第 i 个结中的观察值数量) 为:

表 2.1. 结的计算

观察值	2	2	4	7	7	7	8	9	9	9	9	10
秩												

其中有 6 个结, 每个结长分别为

第2章 单一样本的推断问题

- 1、符号检验 要求：会作检验 并能求出精确p值
P40 例2.2 ★
 - 2、Wilcoxon符号秩检验 P87 2.1 2.4（会给出
P324附表2部分信息）★
 - 3、符号检验和Wilcoxon符号秩检验 的区别和联系
P60
 - 4、Cox-Staut趋势存在性检验P47 例2.6 ★
 - 5、Kolmogorov-Smirnov正态性检验的原理，
它与卡方拟合优度检验的区别。P75-76
-

第 2 章 单一样本的推断问题

1、一般可用_____、_____、_____来表示数据的中心位置。在参数分析方法中，总体的中心位置常用总体的_____表示；在非参数分析方法中总体的中心位置常用总体的_____表示。符号检验方法是讨论对_____的检验问题。

第2章 单一样本的推断问题

2、对称中心的概念

X 的分布关于 θ 对称, 则对任意的 $a > 0$, 应满足
_____.

3、Wilcoxon符号秩和检验是对_____的
检验问题。

4、若 X 的分布关于点 θ_0 对称, 则 $Z = X - \theta_0$ 关于
_____对称。

第3章 两独立样本数据的位置和尺度推断

1、Brown-Mood中位数检验 P92例3.1 ★

要求：能利用P91的分布求 p 值

2、Wilcoxon-Mann-Whitney秩和检验 ★

例3.4（会给出检验的临界值表

P332附表4 部分信息）

P95定理3.1 $E(W_Y), Var(W_Y)$ 的推导

统计量 W_Y 的取值情况。

第4章 多组数据位置推断

1、Kruskal-Wallis单因素方差分析 P119 例4.3 ★

$E(H)$ 的推导

2、Friedman秩方差分析法 P144 4.4 ★

(注意处理和区组的区别)

$E(Q)$ 的推导

3、两者间的区别

假设某公司有三种方法供员工执行某生产任务使用。为检验这三种方法之间有没有差异，有以下两个方案：

方案 1：随随机地抽取 6 个员工。每一个员工先后用三种方法执行该生产任务。6 个员工完成该生产任务所用的时间（分钟）由下表给出。

员工	1	2	3	4	5	6
方法 1	6.0	5.0	7.0	6.2	6.1	6.4
方法 2	5.4	5.2	6.5	5.9	6.0	5.8
方法 3	6.4	5.4	6.7	6.3	6.8	7.2

方案 2：先随机地抽取 6 个员工，其中的每个员工都使用方法 1。然后随机地抽取另 6 个员工，其中的每个员工都使用方法 2。最后再随机地抽取另 6 个员工，其中的每个员工都使用方法 3。他们完成该生产任务所用的时间（分钟）由下表给出。

方法 1：7.2 6.8 7.3 6.0 6.6 7.0

方法 2：6.5 6.2 5.1 6.1 5.9 6.9

方法 3：7.6 7.1 7.4 6.3 7.5 6.4

①你认为在使用这两个方案的数据检验这三种方法之间有没有差异时，能用相同的方法吗？为什么？

②请根据方案 1 得到的数据，检验这三种方法之间有没有差异。

③请根据方案 2 得到的数据，检验这三种方法之间有没有差异。

第5章 分类数据的关联分析

★ 1、卡方检验 P178 5.2

2、Fisher精确性检验

★ 3、关联规则 Apriori算法（会找强关联规则）

概念：支持度 可信度

P154例 P179 5.8(2)

4、Cochran检验 P135 例4.10

关联规则挖

假设最小值支持度为50%，最小置信度为50%

交易ID	购买商品
2000	A,B,C
1000	A,C
4000	A,D
5000	B,E,F

频繁项集	支持度
{A}	75%
{B}	50%
{C}	50%
{A,C}	50%

▶ 对于规则 $A \Rightarrow C$:

▶ 支持度 = $\text{support}(\{A, C\}) = 50\%$

▶ 置信度 = sup 规则 $A \Rightarrow C$ 满足最小支持度和最小置信度，所以它是强关联规则
= 66.6%

Apriori算法举例

现有A、B、C、D、E五种商品的交易记录表，找出所有频繁项集，假设最小支持度 $\geq 50\%$ ，最小置信度 $\geq 50\%$

交易号	商品代码
T1	A、C、D
T2	B、C、E
T3	A、B、C、E
T4	B、E

Apriori算法举例_产生频繁项集

K=1

	项集	支持度
C1	{A}	50%
	{B}	75%
	{C}	75%
	支持度<50	
	{E}	75%



L1	{A}	50%
	{B}	75%
	{C}	75%
	{E}	75%

K=2



	项集	支持度
C2	支持度<50	
	{A,C}	50%
	支持度<50	
	{B,C}	50%
	{B,E}	75%
	{C,E}	50%

L2	{A,C}	50%
	{B,C}	50%
	{B,E}	75%
	{C,E}	50%

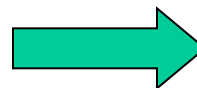


Apriori算法举例_产生频繁项集

L2	{ A,C }	50%
	{ B,C }	50%
	{ B,E }	75%
	{ C,E }	50%



从K2中求可用来计算的的三项集	
{ A,C } + { B,C }	{ A,B,C }
{ A,C } + { B,E }	超过三项
{ A,C } + { C,E }	{ A,C, E }
{ B,C } + { B,E }	{ B,C, E }
{ B,C } + { C,E }	{ B,C, E }
{ B,E } + { C,E }	{ B,C, E }



L3	{ B, C, E }	50%
----	-------------	-----



C3	支持度 < 50	
	支持度 < 50	
	{ B, C, E }	50%

Apriori算法举例_产生关联规则

- 对于频繁项集{B,C,E},它的非空子集有{B}、{C}、{E}、{B,C}、{B,E}、{C,E}。以下就是据此获得的关联规则及其置信度。

规则	置信度Confidence
$B \rightarrow CE$	66.7%
$C \rightarrow BE$	66.7%
$E \rightarrow BC$	66.7%
$CE \rightarrow B$	1
$BE \rightarrow C$	66.7%
$BC \rightarrow E$	1

置信度 $\geq 50\%$ (最小置信度),
都是强关联规则

第6章 秩相关和分位数回归

- 1、3个相关系数：Pearson相关系数，Spearman秩相关系数，Kendall τ 相关系数
- 2、Spearman相关检验
P182 例6.1
- 3、Kendall τ 相关检验 P187 例6.4
- 4、多变量Kendall协和系数（不出大题）



第6章 秩相关和分位数回归

例：零售商要了解每周的广告费和销售收入间的关系，记录了如下的数据：

广告费(10万元)	1	2	3	4	9	16	46
销售收入(100万元)	1	3	5	7	9	12	15

计算Pearson矩相关系数和Spearman秩相关系数，是否相等？为什么？

比较两个检验的方法

常用的比较方法:

(1)待比较的两个检验在相同的样本容量 n 的情况下, 给定相同的显著性水平 α , 比较犯第二类错误的大小.

犯第二类错误概率小的检验为好.

定义1: 功效 $\beta = 1 - \text{检验犯第二类错误的概率}$
 $= P(\text{拒}H_0 | H_1\text{真})$

即等价于功效大的检验为好.

(2)待比较的两检验取相同的显著性水平 α , 并取同样的功效 β , 比较要达到水平 α 和功效 β 所需要的样本容量.

所需样本容量较小的检验为好. (1),(2)两种方法等价

比较两个检验的方法

定义： $\frac{n_1}{n_2}$ 称为检验2对检验1的渐近相对效率

注：若渐近相对效率 $\frac{n_1}{n_2} > 1$, 说明检验2优于检验1;

若渐近相对效率 $\frac{n_1}{n_2} < 1$, 说明检验1优于检验2.