

# 医学统计学

Echo

## 摘要

此为问答整理，参考书本《医学统计学-陆守曾 (第 3 版)》及其他资料。

## 目录

<b>1 绪言</b>	<b>4</b>
1.1 简述变量的特点及其类型	4
1.2 小概率原理	4
<b>2 个体变异与变量分布</b>	<b>4</b>
2.1 频率分布表的绘制步骤	4
2.2 计量资料集中趋势描述指标、适用范围	4
2.3 计量资料离散趋势描述指标、适用范围	4
2.4 分类变量的描述指标 (相对数), 注意事项	5
2.5 率和构成比的鉴别	5
2.6 举例说明率的标准化法的基本思想	6
2.7 应用标准化法的注意事项	6
2.8 正态分布的特点及其应用	6
2.9 二项分布的特点及其应用条件	6
2.10 正态分布、二项分布、 <i>Possion</i> 分布之间有什么关系	7
2.11 参考值范围的确定	7
<b>3 抽样误差</b>	<b>8</b>
3.1 数理统计中心极限定理	8
3.2 均数的抽样误差、反映抽样误差大小的指标及其公式	8
3.3 均数的标准误与标准差的区别与联系	8
3.4 $t$ 分布特点, 和标准正态分布异同	8
<b>4 可信区间</b>	<b>9</b>
4.1 可信区间和假设检验的区别和联系	9
4.2 总体均数的可信区间与医学参考值范围的区别	9
4.3 可信区间的两个要素	10

<b>5 假设检验</b>	<b>10</b>
5.1 假设检验注意事项	10
5.2 第一类错误和第二类错误的区别与联系, 有何实际意义	10
<b>6 定量资料的分析</b>	<b>11</b>
6.1 配对 $t$ 检验 $Vs$ 成组 $t$ 检验	11
6.2 方差分析的基本思想、用途及应用条件	11
6.3 完全随机方差分析中, $MS(\text{总}) = MS(\text{组间}) + MS(\text{误差})$ 是否成立	12
6.4 比较完全随机设计与随机区组设计资料的方差分析的基本思想	12
6.5 重复测量方差分析和区组方差分析的区别和联系	13
6.6 单因素方差分析和双因素方差分析, 适用情况	13
6.7 多个均数的比较不能直接作两两比较的 $t$ 检验	13
6.8 方差分析中的 $F$ 检验为何是单侧检验	13
<b>7 定性资料的分析</b>	<b>13</b>
7.1 行列表资料, 分类, 分析步骤	13
7.2 $R \times C$ 表 $\chi^2$ 检验的注意事项	13
7.3 $\chi^2$ 检验的基本思想和适用范围	14
7.4 $\chi^2$ 检验各个公式适用条件	14
7.5 四格表资料、配对四格表资料的 $\chi^2$ 检验在设计方法、资料整理、假设检验等方面的差别	15
7.6 确切概率表的基本思想	15
<b>8 等级资料的分析</b>	<b>15</b>
8.1 简述非参数检验的适用范围	15
8.2 简述秩和检验, 优缺点、适用范围、常见类型	16
8.3 什么是参数检验/非参数检验, 区别	16
8.4 两样本比较秩和检验基本思想	17
8.5 配对设计秩和检验基本思想	17
8.6 两样本比较的秩和检验, 当 $n_1 > 10$ 、 $n_2 - n_1 > 10$ 时, 采用 $U$ 检验, 这时候检验是属于参数检验还是非参数检验	17
<b>9 资料分布的拟合优度检验</b>	<b>18</b>
<b>10 两指标间的直线相关</b>	<b>18</b>
10.1 相关分析的注意事项	18
10.2 $r$ 和 $r_s$ 的应用条件有何不同	18
<b>11 两指标间的直线回归</b>	<b>18</b>
11.1 直线回归应注意的问题	18
11.2 直线回归的适用条件和应用	19
11.3 直线回归的三个区间估计及其公式	19
11.4 如何说明回归方程有意义	19
11.5 相关与回归的区别和联系	19

<b>12 两指标间的曲线回归</b>	<b>20</b>
12.1 拟合曲线回归的步骤	20
<b>13 研究设计 (一)</b>	<b>20</b>
13.1 研究设计的作用	20
13.2 实验设计的基本要素	20
13.3 实验设计的基本原则	20
13.4 实验设计中什么是均衡性, 不满足均衡性有什么后果, 怎么处理?	21
13.5 对照组的条件	21
13.6 对照组种类	21
13.7 什么是随机化, 如何进行随机化?	21
13.8 随机分组	22
13.9 样本含量的决定因素及估算方法	22
13.10 实验研究样本量估算四要素	22
<b>14 研究设计 (二)</b>	<b>22</b>
14.1 完全随机设计/成组设计	22
14.2 配对设计	23
14.3 随机区组设计	23
14.4 拉丁方设计	24
14.5 析因设计	24
14.6 交叉设计	24
<b>15 研究设计 (三)</b>	<b>24</b>
15.1 寿命表分析	24
15.2 寿命表的几个指标分别代表什么与计算	24
15.3 Kaplan-Meier 法画生存曲线	24
15.4 随访资料的生存分析方法	24
15.5 去死因寿命表的意义, 用其研究各类死因对居民寿命影响的优点	24
<b>16 研究设计 (四)</b>	<b>24</b>
16.1 试验研究和调查研究的区别	24
16.2 常用随机抽样方法, 各自使用场合 (各自优缺点等见流行病 3.1)	24
<b>17 综合分析</b>	<b>25</b>

## 1 绪言

### 1.1 简述变量的特点及其类型

1. 定量资料：以定量值表达每个观察单位的某项观察指标。特点：(1) 观察指标在各观察单位间只有量的差别；(2) 数据间有连续性。
2. 定性资料：以定性方式表达每个观察单位的某项观察指标。特点：(1) 观察指标在各观察单位间或者相同，或者存在质的差别；(2) 有质的差别者之间无连续性。
3. 等级资料：以等级方式表达每个观察单位的某项观察指标。特点：(1) 观察指标在各观察单位间或者相同，或者存在质的差别；(2) 各等级间只有顺序，而无数值大小，故等级之间不可度量。

### 1.2 小概率原理

概率是随机事件发生可能性大小的一个度量，常用  $P$  来表示，取值范围为  $0 \leq P \leq 1$ 。当某事件发生的概率小于或等于 0.05 时，统计学通常称该事件为小概率事件，其涵义为该事件发生的可能性很小，进而认为其在一次抽样中不可能发生，此为小概率原理。

## 2 个体变异与变量分布

### 2.1 频率分布表的绘制步骤

1. 计算极差；
2. 确定组段数和组距：组距 =  $R/\text{组段数}$ ，组段数一般在 10~15 之间；
3. 确定各组段的上下限：第一组段包含最小值，最后一组段包含最大值；
4. 列表：纵标目为组段，横标目为频数、频率、累计频数、累计频率。

### 2.2 计量资料集中趋势描述指标、适用范围

平均数是用来描述定量资料集中位置的一组指标体系，包括算数均数、几何均数、中位数、众数、调和均数等。

1. 算术均数：简称均数。适用于对称分布，尤其是正态分布或近似正态分布的资料。
2. 几何均数：适用于观察值呈倍数变化或等比资料，对数分布资料等。如抗体滴度。
3. 中位数：适用于偏态分布资料、分布不明或分布的末端无确定值的资料。

### 2.3 计量资料离散趋势描述指标、适用范围

用于反映离散程度的变异指标包括全距、四分位数间距、方差、标准差和变异系数等。

- 1. 极差：适用于描述单峰分布，小样本资料的变异程度，初步了解资料的变异程度。虽然简便但稳定性较差。
- 2. 四分位数间距：适用于描述偏态分布资料，两端无确定值或分布不明的资料的离散程度，比极差稳定。
- 3. 方差：适用于描述对称分布，特别是正态分布或近似正态分布资料的离散程度。
- 4. 变异系数，主要用于量纲不同的指标间或均数相差较大的指标间变异程度的比较。

2.4 分类变量的描述指标 (相对数)，注意事项

相对数：定性的或等级的原始资料经过整理，获得若干”绝对数”，由绝对数可以计算出相对数。相对数表示相对关系，常用的相对数包括率、构成比、相对比等。

- 作用：
  - 1. 表示事物出现的频率；
  - 2. 便于比较；
- 常用相对数：
  - 1. 率：又称频率指标，用以说明某事物或某现象在其可能发生的范围内实际发生的频率或强度。
  - 2. 构成比：又称构成指标，表示某一事物内部各部分组成部分所占的比重或频率。
  - 3. 比：又称相对比，是 A、B 两个有关指标之比。
- 相对数应用时的注意事项
  - 1. 计算相对数时分母不能过小；
  - 2. 不要把构成比和率相混淆；
  - 3. 正确计算合并率 (分子相加/分母相加)；
  - 4. 注意资料的可比性；
  - 5. 注意使用率的标准化；
  - 6. 应用相对数进行比较时应进行参数估计和假设检验。

2.5 率和构成比的鉴别

	率	构成比
概念	发生的频率或强度	各组成部分所占的比重
强调点	随机发生事件	各部分的构成
资料获得	较难	容易
特点	不一定	合计为 100%

## 2.6 举例说明率的标准化法的基本思想

当比较组间内部构成不同,导致了组间总率的大小结果与各率比较的结果不一致。采用统一的标准(平台),以消除其内部构成上的差别。以该标准,分别对各比较组的率进行标准化,再行比较标准化率,反映比较组间的相对水平。

举例:甲疗法收治的病人以重症居多,乙疗法则以轻症为多,即甲组重症患者构成比相对较大。由于无论是甲疗法还是乙疗法,重症的治愈率均较低,因而使得甲疗法总的治愈率低于乙疗法。

## 2.7 应用标准化法的注意事项

1. 标准化法只适用于某因素两组内部构成不同,并可能影响两组总率比较的情况。对于其他条件不同产生的不具有可比性的问题,标准化法不能解决。
2. 由于选择的标准人口不同,算出的标准化率也不相同,因此当比较几个标准化率时,应采用同一个标准人口。
3. 标准化后的标准化率,已经不再反映当地的实际水平,它只表示相互比较的资料间的相对水平。
4. 两个样本标准化率是样本值,存在抽样误差。当比较两个样本量较小的标准化率时应进行假设检验。

## 2.8 正态分布的特点及其应用

### • 特点:

1. 钟形曲线,关于  $X = u$  对称,左右完全对称,两端与  $X$  轴永不相交。
2. 在  $X = u$  处取得该概率密度函数的最大值,在  $X = u \pm \delta$  处有拐点。
3. 曲线下面积为 1,曲线下对称于 0 的区间面积相等。
4.  $u$  决定曲线的位置, $u$  越大曲线右移, $u$  越小曲线左移。
5.  $\delta$  决定曲线的形状, $\delta$  越大曲线越矮胖, $\delta$  越小曲线越瘦高。

- 应用:估计频数分布,确定参考值范围,质量控制,统计分析方法的基础。

## 2.9 二项分布的特点及其应用条件

### • 特点:

1. 当  $p$  值较小且  $n$  不大时,分布时偏倚的。随  $n$  的增大分布趋于对称;
2. 当  $p$  值趋于 0.5 时,分布趋于对称;
3. 对于固定的  $n$  及  $p$ ,当  $X$  增加时,  $Pn(k)$  先随之增加并达到其极大值,以后又下降。

### • 应用条件:

1. 各观察单位只能有互相对立的两种结果之一;
2. 已知发生某一结果的概率为  $p$ ,其对立结果的概率则为  $1 - p = q$ ;
3.  $n$  次观察结果互相独立。

2.10 正态分布、二项分布、*Possion* 分布之间有什么关系

- **区别：**二项分布、*Possion* 分布是离散型概率分布，用概率函数描述其分布情况，而正态分布是连续型概率分布，用密度函数和分布函数描述其分布情况。
- **联系：**
  1. 当  $n$  足够大， $\pi$  或  $1 - \pi$  不太小时，二项分布近似于正态分布。
  2. 当  $n$  足够大， $\pi$  或  $1 - \pi$  很小时，二项分布近似于 *Possion* 分布。

2.11 参考值范围的确定

- **原则：**
  1. 选定同质的正常人作为研究对象；
  2. 控制检测误差：在检测过程中，通过人员培训、控制检测条件、重复测定等措施，严格控制检测误差；
  3. 判断是否分组：当变量值在不同年龄组之间、性别之间的差别有临床意义时应分组计算参考值范围；
  4. 单、双侧问题：根据专业知识判断参考值范围取单侧还是双侧界值；
  5. 选择百分界值：参考值范围的涵义是：绝大多数的正常人在该范围内。习惯上将绝大多数定义为正常人的 80%、90%、95%、99%。
  6. 确定可疑范围：若病人与正常人的数据有重叠，则确定任何界值都会产生假阳性或假阴性。重叠部分越大，产生假阳性或假阴性的错误比例就越大，故设定可疑范围。
- **确定方法：**
  1. 正态分布法：医学上许多指标服从或近似服从正态分布，有些呈偏态分布的资料若经某种转换成为正态分布，也可按正态分布规律处理。

双侧	$\bar{X} \pm u_{\alpha} s$
单侧	$\bar{X} + u_{\alpha} s$
	$\bar{X} - u_{\alpha} s$

2. 百分位数法：当资料不服从正态分布时，则可用百分位数法估计参考值范围。

双侧	$P_{2.5} \sim P_{97.5}$
单侧	低侧 $\geq P_5$
	高侧 $\leq P_{95}$

## 3 抽样误差

### 3.1 数理统计中心极限定理

1. 从均值为  $\mu$ 、标准差为  $\sigma$  的正态总体中独立、重复、随机抽取含量为  $n$  的样本，样本均数的分布仍为正态分布，其均数为  $\mu$ ，标准差为  $\sigma_{\bar{X}}$ 。
2. 即使从非正态总体 (均数为  $\mu$ 、标准差为  $\sigma$ ) 中独立、重复、随机抽取含量为  $n$  的样本，只要样本含量足够大 (如  $n \geq 50$ )，样本均数也近似服从于均数为  $\mu$ 、标准差为  $\sigma_{\bar{X}}$  的正态分布。

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

### 3.2 均数的抽样误差、反映抽样误差大小的指标及其公式

在统计理论上将样本统计量的标准差称为统计量的标准误 (SE)，用来衡量抽样误差的大小。据此，样本均数的标准差  $\sigma_{\bar{X}}$  称为均数的标准误，简称标准误， $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ 。 $\sigma$  往往是未知的，一般用样本标准差  $s$  代替  $\sigma$ ，求得  $\sigma_{\bar{X}}$  的估计值  $s_{\bar{X}}$ 。即： $s_{\bar{X}} = \frac{s}{\sqrt{n}}$

### 3.3 均数的标准误与标准差的区别与联系

- 区别：

1. 概念不同：标准差描述的是观察值 (个体值) 之间的变异程度；标准误描述的是样本均数的抽样误差；
2. 用途不同：标准差与均数结合估计参考值范围、计算变异系数、计算标准误等；标准误用于估计参数的可信区间、进行假设检验等。
3. 它们与样本含量的关系不同：当样本含量足够大时，标准差趋于稳定；而标准误随样本含量的增大而减小，甚至趋于 0。

- 联系：

1. 标准差、标准误均为变异指标；
2. 样本均数的标准差即为标准误；
3. 当样本含量不变时，标准误与标准差成正比。

### 3.4 $t$ 分布特点，和标准正态分布异同

- $t$  分布特点：

1. 单峰分布，以 0 为中心，左右对称；
2.  $v$  越小， $t$  值越分散，曲线峰值越矮，尾部越高；



3.  $v$  越大,  $t$  分布逐渐趋近标准正态分布,  $v \rightarrow \infty$  时,  $t$  分布趋向标准正态分布。标准正态分布是  $t$  分布的特例。
- $t$  分布与标准正态分布异同:
    1. 区别:
      - (1) 正态分布是与自由度无关的一条曲线,  $t$  分布是依自由度而变的一组曲线。
      - (2)  $t$  分布较正态分布顶部略低而尾部稍高
    2. 联系: 随着  $v$  增大,  $t$  分布逐渐趋近标准正态分布; 当  $n > 30$  时二者相差很小; 当  $v \rightarrow \infty$  时,  $t$  分布趋向标准正态分布。标准正态分布是  $t$  分布的特例。

4 可信区间

4.1 可信区间和假设检验的区别和联系

- 区别: 可信区间用于推断总体均数的范围, 假设检验用于推断两总体均数是否不同。
- 联系:
  1. 可信区间也可回答假设检验的问题, 可信区间比假设检验提供更多信息;
  2. 可信区间只能在已设定的概率水平  $\alpha$  的前提下进行计算, 假设检验能获得较为准确的  $p$  值, 故将两者结合才是完整的分析;
  3. 根据样本统计量来推断总体参数;
  4. 都要借助  $t$  分布、标准误来计算统计值;
  5. 以 95% 的可信区间得出上下界在一定范围内假设检验是否在假设范围内;
  6. 实际上可信区间与假设检验就是同一个问题, 从不同角度去阐述统计学问题;
  7. 都要考虑样本均数抽样误差。

4.2 总体均数的可信区间与医学参考值范围的区别

	均数可信区间	参考值范围
意义	按预先给定的规律（置信度），确定未知参数的可能范围。实际上一抽样算得的可信区间要么包含总体均数，要么不包含。但可以说：该区间有多大的可能性包含了总体均数。	“正常人”的解剖、生理、生化等指标的波动范围。
相关因素	与标准误大小有关，标准误越大，可信区间越宽。	与标准差大小有关，该范围越宽，分布也就越散。
用途	估计总体均数。	判断观察对象某项指标正常与否。

### 4.3 可信区间的两个要素

- 准确性：反映为可信度  $(1 - \alpha)$  的大小，显然可信度越接近 1 越好。
- 精确性：反映可信区间的范围，范围越小越好。
  1. 精确性与变量的变异度大小，样本量和  $1 - \alpha$  的取值有关；
  2. 在  $n$  确定的情况下，准确度  $\nearrow$ ，精确度  $\searrow$ ；
  3. 一般常用 95% 可信区间，能较好兼顾准确性和精确性；
  4. 在可信度确定的情况下，增加样本例数，可提高精确性。

## 5 假设检验

### 5.1 假设检验注意事项

1. 严密的研究设计：根据设计类型确定统计分析的步骤；
2. 不同资料选用不同的检验方法：若适合  $t$  检验的数据用秩和检验会降低检验效能；
3. 正确理解  $p$  值的含义；
4. 结论不能绝对化；
5. 正确理解统计学意义、临床意义和生物学意义；
6. 可信区间与假设检验的区别和联系：置信区间具有假设检验的主要功能；置信区间可提供假设检验没有提供的信息；假设检验比置信区间提供的信息多。

### 5.2 第一类错误和第二类错误的区别与联系，有何实际意义

- 区别：
  1.  $I$  类错误是拒绝了实际上成立的  $H_0$ ，也称为“弃真”错误，用  $\alpha$  表示。统计推断时，根据研究者的要求来确定。
  2.  $II$  型错误是不拒绝实际上不成立的  $H_0$ ，也称“存伪”错误，用  $\beta$  表示。它只能与特定的  $H_1$  结合起来才有意义，一般难以确切估计。
- 联系：
  1. 当抽样例数一定时， $\alpha$  越大， $\beta$  越小；反之  $\alpha$  越小， $\beta$  越大。
  2. 统计推断中， $I$ 、 $II$  型错误均有可能发生，若使两者都减小，可适当增加样本量。
  3. 根据研究者要求， $n$  一定时，可通过  $\alpha$  水平来控制  $\beta$  大小。
- 实际意义：
  1. 可用于样本量的估计；
  2. 可用来计算可信度  $(1 - \alpha)$ ，表明统计推断的可靠性大小；
  3. 可用来计算把握度  $(1 - \beta)$ ，来评价检验方法的效能等；
  4. 有助于研究者选择适当的检验水准；

5. 可以说明统计结论的概率保证。

## 6 定量资料的分析

### 6.1 配对 $t$ 检验 $V_s$ 成组 $t$ 检验

- 适用条件不同:

1. 配对  $t$  检验适用于配对设计两样本均数差异显著性检验, 两样本来自同一整体;
2. 成组  $t$  检验适用于非配对设计或成组设计两样本均数差异显著性检验, 两样本相互独立。

- 假设检验不同:

1. 配对设计资料假设检验可视为样本均数与总体均数  $\mu_d = 0$  的比较:

$$H_0: \mu_d = 0 \text{ (即差值的总体均数为 0)}$$

$$H_1: \mu_d \neq 0 \text{ (即差值的总体均数不为 0)}$$

2. 成组  $t$  检验无效假设  $H_0: \mu_1 = \mu_2$

$$\text{备择假设 } H_1: \mu_1 \neq \mu_2$$

- 计算公式不同:

1. 配对  $t$  检验计算  $t$  值公式:

$$t = \frac{\bar{d} - \mu_0}{s_d / \sqrt{n}}$$

2. 成组  $t$  检验计算  $t$  值公式:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

- 检验效率不同:

样本例数相同时, 计量资料的成组  $t$  检验比配对  $t$  检验检验效率低。因为采用配对方法, 把一些对实验结果有影响的因素进行匹配, 消除了这些因素带来的干扰, 降低了误差。

### 6.2 方差分析的基本思想、用途及应用条件

- 基本思想:

方差分析的基本思想是把全部观察值间的变异按设计和需要分解成两个或多个部分, 然后将各影响因素产生的变异和随机误差进行比较, 以判断各部分的变异与随机误差相比较是否具有统计学意义。具体说

来就是根据实验设计的类型，将全部测量值总的离均差平方和及其自由度分解为两个或多个部分，除随机误差的作用外，每个部分的变异可有某个因素的作用加以解释。如组间变异可由处理因素的作用加以解释，通过比较不同变异来源的均方，借于  $F$  分布做出推断，从而推论各种研究因素对实验结果的影响。

• 用途：

1. 检验两个或多个样本均数间的差异有无统计学意义；
2. 回归方程的线性假设检验；
3. 检验两个或多个因素间有无相互作用。

• 应用条件：

1. 各样本是相互独立的随机样本；
2. 均服从正态分布；
3. 各样本的总体方差相等，即是指满足方差齐性。

### 6.3 完全随机方差分析中， $MS(\text{总}) = MS(\text{组间}) + MS(\text{误差})$ 是否成立

不成立

$$\therefore \begin{cases} SS_{\text{总}} = SS_{\text{组间}} + SS_{\text{误差}} \\ V_{\text{总}} = V_{\text{组间}} + V_{\text{误差}} \end{cases}$$

$$\therefore MS_{\text{总}} = SS_{\text{总}} / V_{\text{总}} = (SS_{\text{组间}} + SS_{\text{误差}}) / (V_{\text{组间}} + V_{\text{误差}}) \quad ①$$

$$MS_{\text{组间}} = SS_{\text{组间}} / V_{\text{组间}} \quad ②$$

$$MS_{\text{误差}} = SS_{\text{误差}} / V_{\text{误差}} \quad ③$$

$$② + ③ \text{ 得 } MS_{\text{组间}} + MS_{\text{误差}} = \frac{(SS_{\text{组间}} + SS_{\text{误差}}) + \frac{V_{\text{误差}}}{V_{\text{组间}}} \cdot SS_{\text{组间}} + \frac{V_{\text{组间}}}{V_{\text{误差}}} \cdot SS_{\text{误差}}}{V_{\text{组间}} + V_{\text{误差}}} \quad ④$$

$$\therefore ④ > ①$$

$\therefore$  上式不成立

### 6.4 比较完全随机设计与随机区组设计资料的方差分析的基本思想

1. 完全随机设计：将同质的受试对象随机地分配到各处理组，再观察其实验效应，各组样本含量可以相等，也可以不等。是最常见的研究单因素水平或多水平的实验设计方向，属于单向方差分析。
2. 随机区组设计：又称为配伍组设计，通常是将受试对象按性质相同或相近者组成数个区组（配伍组），每个区组中的受试对象分别随机分配到  $k$  个处理组中去。其方差分析属于无重复数据的双向方差分析。

## 6.5 重复测量方差分析和区组方差分析的区别和联系

- 区别:

1. 重复测量设计中的“处理”是在区组 (受试者) 间随机分配, 区组内的各时间点是固定的, 不能随机分配。
2. 重复测量设计区组内单位彼此不独立, 资料间呈一定的相关性。

- 联系: 当单样本重复测量的资料满足“球对称”条件时, 可用随机完全区组设计方差分析进行假设检验。

## 6.6 单因素方差分析和双因素方差分析, 适用情况

1. 单因素方差分析: 即影响样本的因素只有一个, 用于完全随机设计的多个样本均数间的比较, 其统计判断是判断各样本所代表的总体均数是否相等。
2. 双因素方差分析: 影响样本的因素有两个, 双因素方差分析是对影响因素进行检验, 研究是一个因素在起作用还是两个因素都起作用, 或是两个因素的影响都不明显。

## 6.7 多个均数的比较不能直接作两两比较的 $t$ 检验

两均数比较的  $t$  检验进行多重比较, 将会加大  $I$  类错误的概率  $\alpha$ , 从而可能把本无差别的两个总体均数判为有差别。例如, 有 4 个均数, 两两组合数为  $C_4^2 = 6$ , 若用  $t$  检验做 6 次比较, 且每次比较的检验水准为  $\alpha = 0.05$ , 则每次比较不犯  $I$  类错误的概率为  $(1 - 0.05)$ , 6 次均不犯  $I$  类错误的概率为  $(1 - 0.05)^6$ , 这时候总的检验水准变为  $1 - (1 - 0.05)^6 = 0.26$ , 比 0.05 大得多, 因此不能用  $t$  检验进行多个均数的比较。

## 6.8 方差分析中的 $F$ 检验为何是单侧检验

方差分析中检验统计量  $F$  值的计算通常是用某部分的均方 (如处理因素、交互效应等) 除以误差的均方, 其中, 分母误差的均方仅含随机因素的作用, 分子误差的均方含随机因素和处理因素两部分, 因此算得的  $F$  值从理论上讲应大于或等于 1, 而不会小于 1, 所以方差分析中的  $F$  界值采用单侧检验的界值。

# 7 定性资料的分析

## 7.1 行列表资料, 分类, 分析步骤

## 7.2 $R \times C$ 表 $\chi^2$ 检验的注意事项

1. 单项有序行列表, 当等级数  $\geq 3$  时, 一般用秩和检验。
2.  $\chi^2$  检验结论有统计学意义时, 仅表示多组之间有差别, 即多组之间至少有两组的概率分布是不同的, 但并不是任意两组之间都有差别。此时进一步检验要使用卡方分割。

3.  $\chi^2$  检验要求理论频数不宜太小, 一般不宜有  $\frac{1}{5}$  以上的格子理论频数小于 5, 或不宜有 1 个理论频数小于 1, 否则可能产生偏差。
4. 若  $\chi^2$  检验中理论频数太小, 则做以下处理: 删除理论频数太小的行或列; 将理论频数太小的行或列与性质相近的邻行或邻列合并; 增大样本量; 使用 *fisher* 确切法。

### 7.3 $\chi^2$ 检验的基本思想和适用范围

- **基本思想:** 如果  $H_0$  假设成立, 那么实际频数与理论频数应该比较接近。如果实际频数与理论频数相差较大, 超出了抽样误差所能解释的范围, 则可以认为  $H_0$  假设不成立, 即两样本对应的总体率不等。

$$\chi^2 = \sum \frac{(A_i - T_i)^2}{T_i} = \frac{(ad - bc)^2 * n}{(a + b)(c + d)(a + c)(b + d)}$$

- **适用范围:**

1. 推断两个及两个以上总体率或构成比是否有差别, 两个分类变量间有无相关关系。
2. 频数分布的拟合优度检验。
3. 配对设计样本率比较。
4. 线性趋势检验。

### 7.4 $\chi^2$ 检验各个公式适用条件

- **完全随机设计两样本率的比较**

1. 当  $n \geq 40$ , 且  $T \geq 5$  时, 用非连续校正  $\chi^2$  检验, 若所得  $p \approx \alpha$ , 则改用四格表的 *fisher* 确切概率法;

$$\chi^2 = \sum \frac{(|A - T|)^2}{T} = \frac{(ad - bc)^2 * n}{(a + b)(c + d)(a + c)(b + d)}$$

2. 当  $n \geq 40$ , 且有  $1 \leq T < 5$  时, 用连续校正  $\chi^2$  检验;

$$\chi_c^2 = \sum \frac{(|A - T| - 0.5)^2}{T} = \frac{(|ad - bc| - n/2)^2 * n}{(a + b)(c + d)(a + c)(b + d)}$$

3.  $n < 40$ , 或有  $T < 1$  时, 不能用  $\chi^2$  检验, 应当用四格表的 *fisher* 确切概率法。

- **配对设计四格表**

1. 当  $b + c \geq 40$  时, 进行非连续性校正的 *McNemar's* 配对  $\chi^2$  检验,  $\chi^2 = \frac{(|b - c|)^2}{b + c}$ ;
2. 当  $20 \leq b + c < 40$  时, 进行连续校正的 *McNemar's* 配对  $\chi^2$  检验,  $\chi_c^2 = \frac{(|b - c| - 1)^2}{b + c}$ ;
3.  $b + c < 20$  时直接计算 *fisher* 确切概率法。

- **行列表资料**

1.  $R \times C$  表格中要求理论频数不宜太小, 一般不宜有  $\frac{1}{5}$  以上的格子理论频数小于 5, 或不宜有 1 个理论频数小于 1, 否则可能产生偏差。
2. 若  $R \times C$  表格中理论频数太小, 则做以下处理: 删除理论频数太小的行或列; 将理论频数太小的行或列与性质相近的邻行或邻列合并; 增大样本量; 使用 *fisher* 确切法。

## 7.5 四格表资料、配对四格表资料的 $\chi^2$ 检验在设计方法、资料整理、假设检验等方面的差别

### • 设计方法:

1. 普通四格表资料为成组设计: 将同质的实验对象随机地分配到研究因素各水平所对应的处理组, 进行处理并观察, 或从不同特征的总体中随机抽样, 并对样本进行对比观察。
2. 配对四格表为配对设计: 将实验单位按一定条件配成对子, 再将每对中的两个实验单位随机分配到两个处理组中。配对的因素一般是可能影响研究结果的非处理组因素。

### • 资料整理:

1. 普通四格表为独立二分类样本表格形式, 横标题和纵标题为不一致的分类。
2. 配对四格表为相关样本表格形式, 横标题和纵标题为相关或相同的分类。

### • 假设检验:

1. 普通四格表资料进行 *Person* 卡方检验, 视具体情况进行 *Yate's* 连续性校正或 *fisher* 确切概率法。
2. 配对四格表资料进行 *McNemar's* 配对卡方检验, 视具体情况进行连续性校正或 *fisher* 确切概率法。

## 7.6 确切概率表的基本思想

保持周边合计数不变, 计算交叉表中各个实际频数变动的所有可能组合所对应的概率, 再将获得现有样本的概率以及比它更极端的所有概率求和, 直接求出单侧或双侧的累计概率进行推断。

# 8 等级资料的分析

## 8.1 简述非参数检验的适用范围

非参数检验适用于以下三种情况:

1. 顺序类型的资料, 这类数据的分布形态一般是未知的;
2. 虽然是连续数据, 但总体分布形态未知或者非正态, 这和卡方检验一样, 称自由分布检验;
3. 总体分布虽然正态, 数据也是连续类型, 但样本容量极小, 如 10 以下 (虽然 *T* 检验被称为小样本统计方法, 但样本容量太小时, 代表性毕竟极差, 最好不要用要求较严格的参数检验法)。

## 8.2 简述秩和检验，优缺点、适用范围、常见类型

- **秩和检验：**基于秩转换的非参数检验称为秩和检验，是用秩和检验作为统计量进行假设检验的方法。其中“秩”又称为等级、即按数据大小排定的次序号，次序号的和称为“秩和”。秩和检验是对从两个非正态总体中所得到的两个样本之间的比较，其零假设为两个样本从同一总体中抽取的。
- **优点：**
  1. 不受总体分布限制，适用面广；
  2. 适用于等级资料及两端无确定值的资料；
  3. 易于理解，易于计算。
- **缺点：**符合参数检验的资料，用秩和检验，则不能充分利用信息，检验功效低。
- **适用范围：**
  1. 等级资料；
  2. 定量资料，但数据的某一端或两端无确定值；
  3. 定量资料，但数值分布是极度偏态的；
  4. 定量资料，但各组离散程度相差悬殊，即使经变量变换，也难以达到方差齐性；
  5. 定量资料，但分布型尚未可知；
  6. 兼有等级和定量性质的资料。
- **常见类型：**
  1. 成组：*Wilcoxon* 秩和检验
  2. 配对：符号秩和检验
  3. 多样本：*Kruskal - Wallis* 法
  4. 区组设计的等级资料：*M* 检验 (*Friedman* 秩和检验)

## 8.3 什么是参数检验/非参数检验，区别

- **定义：**
  1. 参数检验：当总体分布已知，根据样本数据对总体分布的统计参数进行推断。
  2. 非参数检验：当总体分布未知，或知道甚少的情況下，利用样本数据对总体分布形态等进行推断的方法。
- **区别：**



参数检验		非参数检验
适用条件	特定的已知总体分布的统计问题。资料具有正态性、独立性、方差齐性。	总体分布未知的统计问题。等级资料、开口资料、极度偏态分布、各组明显方差不齐、不满足参数检验的结果
检验内容	参数	分布
检验效能	高	低
优点	符合应用条件时，检验效能高	应用范围广，对资料要求不高，计算简便，容易掌握
缺点	对资料要求严格，不能用于等级、一端或两端有不明确数据，另外也要求资料分布类型已知和总体方差齐性。	若对符合参数检验的资料，采用了非参数检验，则会降低检验效能，如需检验出同样大小的差异，非参数检验要求增大样本量。

8.4 两样本比较秩和检验基本思想

如果待比较的两个样本 (样本含量分别为  $n_1$  及  $n_2$ ) 来自同一个总体或分布相同的两个总体 (即  $H_0$  成立), 则含样本量为  $n_1$  的样本之实际秩和与其理论之和秩和  $n_1(N+1)/2$  之差  $[T - n_1(N+1)/2]$  纯系抽样误差所致, 故此差值一般不会很大, 而差值越大的概率越小, 若从现有样本中计算得的  $T$  与其理论秩和相差很大, 则说明从  $H_0$  规定的总体中随机抽得现有样本及更极端样本的概率  $P$  很小, 如小于等于检验水准  $\alpha$ , 则可能拒绝  $H_0$ 。

8.5 配对设计秩和检验基本思想

假定两种处理办法的效应相同, 则样本非 0 之差值之产生系抽样误差所致, 其总体分布应以 0 为中位数, 且越接近 0, 频数分布越密集; 同理, 在自身前后对比的研究中, 若处理无效, 则每一受试对象处理前后所得结果之差值的总体亦应以 0 为中位数。若从样本求得一个偏离  $n(n+1)/4$  很远的  $T$  值, 且其相应的  $P$  小于检验水准  $\alpha$  时, 根据小概率原理, 我们就有理由拒绝  $H_0$ , 接受  $H_1$ , 反之若  $P$  不是太小, 则没有理由拒绝  $H_0$ 。

8.6 两样本比较的秩和检验, 当  $n_1 > 10$ 、 $n_2 - n_1 > 10$  时, 采用  $U$  检验, 这时候检验是属于参数检验还是非参数检验

非参数检验, 当  $n_1 > 10$ 、 $n_2 - n_1 > 10$  时,  $T$ (秩和) 分布接近均数为  $n_1(N+1)/2$ , 方差为  $n_1n_2(N+1)/12$  的正态分布。此时不必要再编制比  $n$  大的  $T$  检验界值表, 而利用秩和分布随  $n$  增大渐进正态分布的性质, 进行  $u$  检验, 检验内容是分布而不是参数, 故仍属于非参数检验。

## 9 资料分布的拟合优度检验

## 10 两指标间的直线相关

### 10.1 相关分析的注意事项

1. 在进行相关分析前先绘制散点图，当散点图有线性趋势时才进行相关分析；
2. 出现离群值时要慎用相关；
3. 线性相关分析要求两个变量都是随机变量，而且仅适用于二元正态分布资料；
4. 相关关系不一定是因果联系。即两变量相关并不代表两变量之间一定存在内在联系；
5. 分层资料盲目合并易出现假象。

### 10.2 $r$ 和 $r_s$ 的应用条件有何不同

- 直线相关系数 $r$

1. 要求两随机变量要符合双变量正态分布，而且不分主次，处于同等地位；
2. 散点有线性趋势；
3. 两随机变量相关分析有实际意义。

- 等级 (秩) 相关系数 $r_s$

1. 不服从双变量正态分布的资料；
2. 总体分布类型未知的资料；
3. 一端或两端有不确定数值的资料；
4. 等级资料。

## 11 两指标间的直线回归

### 11.1 直线回归应注意的问题

1. 作回归分析要有实际意义，不能把毫无关联的两种现象随意进行回归分析，即要求能从专业理论上作出合理解释；
2. 在进行直线回归分析前，应绘制散点图，当观察点的分布呈直线趋势时，才适宜做直线回归分析；
3. 直线回归分析的实用范围一般以自变量的取值范围为限。若无充分理由证明超过自变量取值范围外还是直线，应避免外延；
4. 双变量的小样本经  $t$  检验只能推断总体两变量间有无直线关系，而不能推断相关的紧密程度，要推断相关的紧密程度，样本含量必须很大；
5. 两变量的直线关系不一定是因果关系，也可能是伴随关系，有相关关系不一定表明变量间确有内在联系。

## 11.2 直线回归的适用条件和应用

### • 适用条件:

1. 因变量  $Y$  和自变量  $X$  呈线性关系; (Linear)
2. 每个个体观察值之间相互独立; (Independence)
3. 在一定范围内, 任意给定的  $X$  对应的随机变量  $Y$  都服从正态分布; (Normal distribution)
4. 在一定范围内, 不同的  $X$  值所对应的随机变量  $Y$  方差相等。 (Equal variance)

### • 应用:

1. 描述两变量间的依存关系;
2. 利用回归方程进行预测;
3. 利用回归方程进行统计控制。

## 11.3 直线回归的三个区间估计及其公式

1. 回归系数  $\beta$  的可信区间的估计: 如果经过假设检验, 认为  $X$  与  $Y$  之间的回归关系成立, 那么进一步估计总体回归系数  $\beta$  的  $100(1-\alpha)\%$  的可信限为  $b \pm t_{\alpha, n-2} S_b$ 。
2.  $\mu_{\hat{Y}}$  的可信区间估计:  $\mu_{\hat{Y}}$  是当  $X$  固定时  $Y$  的总体中的条件均数,  $\hat{Y}$  是其估计值, 但  $\hat{Y}$  是有抽样误差的, 其标准误  $s_{\hat{Y}}$  按下式计算:  $s_{\hat{Y}} = s_{Y \cdot X} \sqrt{\frac{1}{n} + \frac{(X - \bar{X})^2}{\sum (X - \bar{X})^2}}$ , 当  $X = \bar{X}$  时,  $s_{\hat{Y}}$  最小, 为  $s_{Y \cdot X} / \sqrt{n}$ 。  $X$  与  $\bar{X}$  相差越大,  $s_{\hat{Y}}$  越大。  $\mu_{\hat{Y}}$  的  $100(1-\alpha)\%$  可信区间公式:  $\hat{Y} \pm t_{\alpha, n-2} s_{\hat{Y}}$ 。
3. 个体  $Y$  值的容许空间估计: 容许空间就是总体中当  $X$  固定时, 个体  $Y$  值的波动范围, 其标准差为  $s_Y = s_{Y \cdot X} \sqrt{1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{\sum (X - \bar{X})^2}}$ , 显见, 当  $X = \bar{X}$  时,  $s_Y$  最小;  $X$  与  $\bar{X}$  相差越大,  $s_Y$  越大。个体值  $\mu_Y$  的  $100(1-\alpha)\%$  可信区间公式:  $\hat{Y} \pm t_{\alpha, n-2} s_Y$ 。

## 11.4 如何说明回归方程有意义

## 11.5 相关与回归的区别和联系

### • 区别:

1. 资料要求不同: 相关要求  $X$ 、 $Y$  服从双变量的正态分布; 回归要求  $Y$  在给定  $X$  值时服从正态分布。
2. 应用情况不同: 相关反映两变量间的相互关系; 回归反映两变量间的依存关系。
3. 相关系数和回归系数的区别:
  - (1) 意义不同: 相关系数  $r$  说明具有线性关系的两变量之间的密切程度与相关方向; 回归系数  $b$  表示  $X$  每变化一个单位所导致  $Y$  的变化量。
  - (2) 取值不同:  $-1 \leq r \leq 1$ ;  $-\infty \leq b \leq +\infty$ 。
  - (3) 单位不同:  $r$  没有单位, 不受  $X$ 、 $Y$  单位的影响;  $b$  有单位

• 联系:

1. 正负符号: 对于既可以做相关又可以做回归分析的同一组数据,  $r$  和  $b$  的方向一致。  $r > 0, b > 0$ ;  $r < 0, b < 0$ 。
2. 假设检验: 在同一资料,  $r$  与  $b$  的假设检验统计量  $t$  值相等:  $t_r = t_b$ , 自由度亦相等, 故  $r$  和  $b$  的假设检验等价。
3. 换算关系: 相关系数和回归系数间可以相互换算:  $b = r\sqrt{l_{yy}/l_{xx}}$
4. 在相关回归中的贡献:  $r^2$  称为决定系数,  $r^2 = \frac{l_{xy}^2}{l_{xx}l_{yy}} = \frac{l_{xy}^2/l_{xx}}{l_{yy}} = \frac{SS}{SS}$ 。  $r^2$  越大 (越接近 1) 则回归解释的部分越大, 反映了实际意义越大。

## 12 两指标间的曲线回归

### 12.1 拟合曲线回归的步骤

1. 定曲线型;
2. 直线化或对称化;
3. 求方程;
4. 作图。

## 13 研究设计 (一)

### 13.1 研究设计的作用

1. 合理安排试验因素, 提高研究质量。
2. 控制误差, 使研究结果保持较好的稳定性。
3. 通过较少的观察例数, 获取尽可能丰富的信息。

### 13.2 实验设计的基本要素

1. 受试对象
2. 处理因素
3. 实验效应

### 13.3 实验设计的基本原则

1. 对照原则: 空白对照、安慰剂对照、标准对照、自身对照、历史对照
2. 随机原则;

3. 重复原则;
4. 均衡原则。

### 13.4 实验设计中什么是均衡性, 不满足均衡性有什么后果, 怎么处理?

- 均衡性: 实验组和对照组或各实验组之间, 除了处理因素以外, 其他一切非处理因素方面应尽可能达到均衡一致。
- 不满足后果: 导致结果产生偏性, 使得实验组和对照组不可比。均衡性使试验因素不同水平组中的受试对象受到非处理因素影响的完全平衡, 确保处理因素各水平组间不受其他处理因素或重要的非处理因素的不平衡的干扰和影响, 以便使所考察的处理因素取不同水平条件下对观测结果的影响能真实地显露出来。不满足均衡性则可能降低假设检验效率; 增加了混杂因素的干扰作用;
- 处理: 通过随机化处理, 包括抽样随机、分组的随机以及实验顺序的随机, 均衡干扰因素的影响, 使实验组和对照组具有可比性, 从而控制实验误差。

### 13.5 对照组的条件

1. 对等: 除研究因素外, 对照组具备与研究组对等的一切因素。
2. 同步: 对照组与研究组设立后, 在整个研究进程中始终处于同一空间和同一时间。
3. 专设: 任何一个对照组都是为相应的研究组专门设立的。不得借用文献上的记载或以往研究的、其他研究的资料作为本研究之对照组。

### 13.6 对照组种类

1. 空白对照: 观察对象在没有研究因素时的自然状态或“本底状态”。如: 注射生理盐水、服用安慰剂等。
2. 标准对照: 以标准品或标准效应作为对照者。如药物标准品之效价、正常值、常规疗法等。
3. 试验对照: 将某个 (或某组) 混杂因素设入对照组。
4. 相互对照: 设几个研究组互作对照, 而并不确定哪一个为对照组。

### 13.7 什么是随机化, 如何进行随机化?

为了使各对比组间在大量不可控制的非研究因素的分布方面尽量保持均衡一致, 统计学采取的重要措施就是随机化。具体体现在三个方面:

1. 抽样随机: 每一个符合条件的研究对象被收入研究范畴的机会相同;
2. 分组随机: 每个研究对象被分配到某处理组的机会相同;
3. 实验顺序随机: 每个研究对象先后接受处理的机会相同。

### 13.8 随机分组

例：将 20 个实验动物随机分配到 A、B 两组：

首先将动物 1 至 20 编号，在计算机上以一定的随机数种子产生 20 个随机数，前 10 个对应于 A 组，后 10 个对应于 B 组。按随机数的大小排序，其对应的秩次就是实验动物的编号。

### 13.9 样本含量的决定因素及估算方法

1. 数据种类；
2. 指标间差值；
3. 变异度；
4. 相关程度；
5. 研究设计质量；
6. 设计方法；
7. 各组例数分配；
8. 所定义  $\alpha$ 、 $\beta$  水准。

$$N = \left( \frac{1}{Q_1} + \frac{1}{Q_2} \right) \frac{(u_\alpha + u_\beta)^2 \sigma^2}{\delta^2}$$

### 13.10 实验研究样本量估算四要素

1. 欲比较的两总体参数的差值。差值越大，所需样本含量越大。
2. 有关总体变异性的信息。该值越大，所需样本含量越多。
3. 第 I 类错误概率的大小。 $\alpha$  越小，所需样本量越大。
4. 第 II 类错误概率  $\beta$  或检验功率  $(1 - \beta)$  的大小。 $\beta$  越小，所需样本量越大。

## 14 研究设计 (二)

### 14.1 完全随机设计/成组设计

1. 使用最为广泛的方法之一。对其它因素不加以控制，故操作较为简单、灵活。可用于两组或两组以上间进行比较的情形。
2. 总样本量一定时，样本量相等的设计效率高于各组样本量不等时，故建议样本量相等。
3. 个体同质性较差时，容易造成组间不均衡性，尤其是在样本量较小时。此时可采用带有区组控制的设计方法，如配对设计或随机区组设计等。
4. 严格遵循随机化原则可以在一定程度上提高组间均衡性。若组间不均衡，则需借助于统计学方法控制混杂因素的干扰。如分层分析、协方差分析，或其他多元统计方法。
5. 当处理数大于 2 时，组间的比较需考虑多重比较的问题。

## 14.2 配对设计

- 异体配对
  - 自身配对
1. 当个体间同质性相对较差时, 正确使用配对设计可减少抽样误差, 比完全随机设计得到更高的检验效能。
  2. 配对设计的成败取决于对子中两个个体的同质性和相关性。
  3. 采用自身前后配对设计要求实验前后观察时间不能间隔过长, 且实验前后各种条件要尽可能相同, 如实验条件、实验对象自身的一些状况, 如心理变化、疾病的自然进展等。
  4. 应尽量减少缺失现象的发生。
  5. 采用异体配对时, 配对条件不能过多、过严, 否则将难于找到足够多可以配对的样本, 但宽松的配对条件又可能造成配对效果欠佳。
  6. 有时配对设计结合回归分析可得到更丰富的结论。

## 14.3 随机区组设计

1. 可以提高各处理组间的可比性和均衡性, 减少误差, 提高检验效能。
2. 与配对设计要求类似, 同一区组内的个体应达到同质。
3. 结合回归分析可以提取更丰富的信息。
4. 尽可能使观察值不缺失。
5. 在同质实验单位数少于处理组时, 导致区组内个体数不足以放置所有水平时, 可考虑采用平衡不完全区组设计。

- 14.4 拉丁方设计
- 14.5 析因设计
- 14.6 交叉设计

15 研究设计 (三)

- 15.1 寿命表分析
- 15.2 寿命表的几个指标分别代表什么与计算
- 15.3 Kaplan-Meier 法画生存曲线
- 15.4 随访资料的生存分析方法
- 15.5 去死因寿命表的意义，用其研究各类死因对居民寿命影响的优点

16 研究设计 (四)

16.1 试验研究和调查研究的区别

试验研究		调查研究
研究条件	由研究者严格设定	由设计者在现实状态下依据需要进行选择
观察对象	观察对象集中于所控制的时空区间内进行	除设计中特定项目外，一般只做观察而不干预观察对象原来的生活及社会生活
例数	通常较少	较多

16.2 常用随机抽样方法，各自使用场合（各自优缺点等见流行病 3.1）

1. 简单随机抽样：适用于总体中个体数少，抽取的样本容量也较小的抽样类型。
2. 系统抽样：总体中的个体数较多，但在总体发生周期性变化的场合不适宜用这种办法。
3. 分层抽样：(1) 总体有差异明显的几部分组成；(2) 产品的质量验收。
4. 整群抽样：(1) 样本单元的分布相较集中的大规模抽样调查；(2) 连续性生产的质量控制。



## 17 综合分析