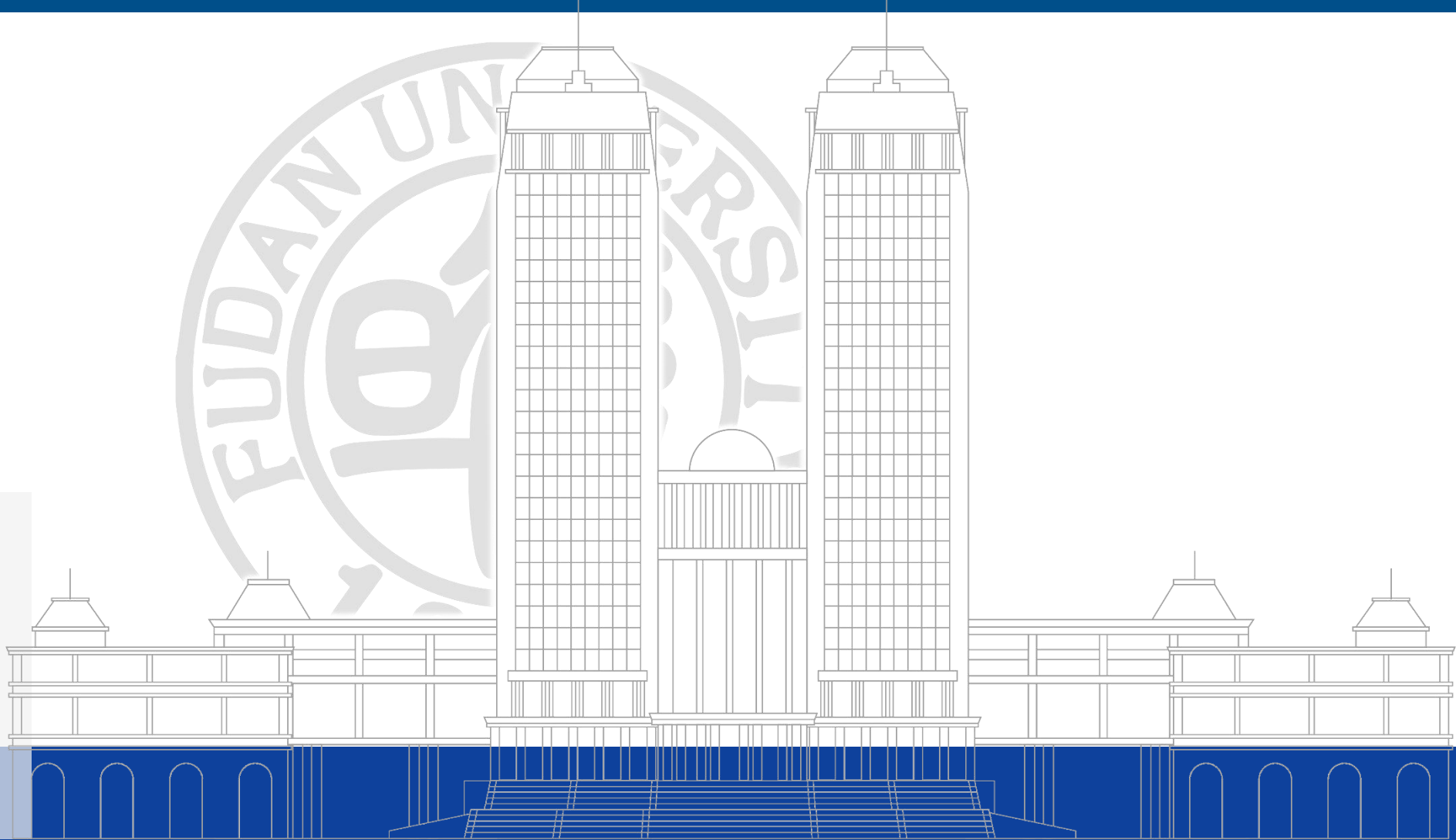




常用医学统计分析方法的SPSS实现

01



统计学基础知识



为什么要做统计分析?

医学统计 & SPSS

- 某医生收治20名患某疾病患者，随机分成2组，每组10人。
 - A组用试验药，B组用对照药；
 - 经过一个疗程治疗；
 - A药组治愈了8人，B药组治愈了7人。
- 请根据现有结果评价下列说法是否正确：
 - A药组的疗效高于B药组； **样本疗效**
 - A药的疗效高于B药。 **人群疗效**



为什么要做统计分析?

医学统计 & SPSS

- 仅仅基于样本均数或样本有效率的结果所下结论是没有意义的。
- 针对研究人群所下的结论才有意义。
- 基于样本分析的结果，经过统计学分析和推断，当结果具有统计学意义时（双侧P值 <0.05 或单侧P值 <0.025 ），可以得到针对研究人群的结论。



如何学统计?

医学统计 & SPSS

- 基础知识（资料的特点，统计学设计的特点，统计学的基本知识等）
- 不同统计分析方法的适用条件
- 至少掌握一种统计分析软件
- 工具：无他，唯手熟尔

知识+软件+实际操作

实际问题是最好的老师



基础知识-变量类型

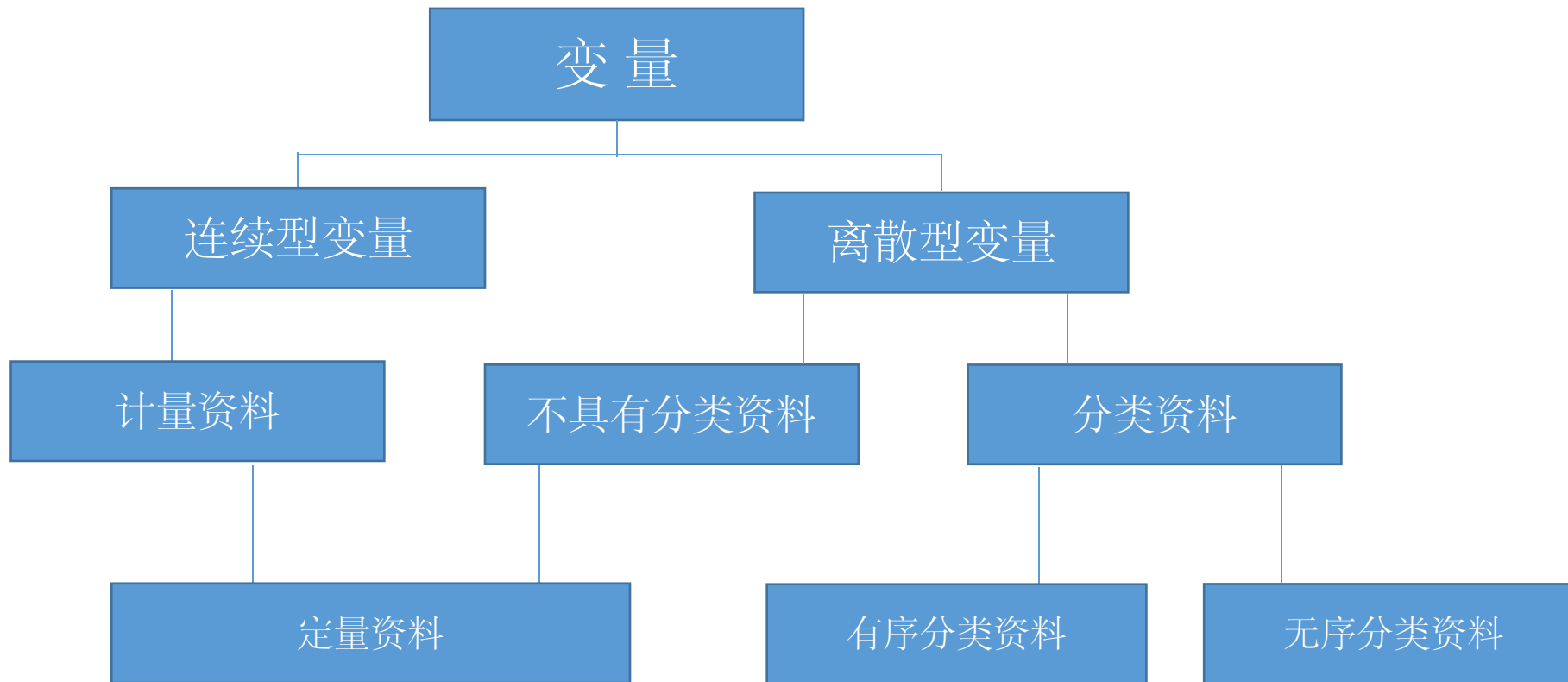
医学统计 & SPSS

- **连续型变量**---可能取值范围是一个区间，连续取值。即：在某一区间内的任意一个值都是可能被取到的，如血压，HbA1c;
- **离散型变量**---取值范围是有限个值或一个数列构成。离散型变量中取值表示分类情况的离散型变量又称为分类变量
 - ✓ 无序变量：两分类和无序多分类，如血型。也可用数字进行编码，但没有大小关系。
 - ✓ 有序多分类变量/等级变量：在研究背景下有等级顺序，如疗效（无效、有效、显效）
- **变量的转化**：连续型---有序---分类（信息损失）如：年龄进行分组，疗效归为有效或无效等。



基础知识-变量类型

医学统计 & SPSS





基础知识-变量类型

医学统计 & SPSS

例：调查某地某年1岁儿童的生长发育情况

人群：某地某年1岁的儿童

变量：性别、身高、体重、出牙、营养状况

变量类型：??

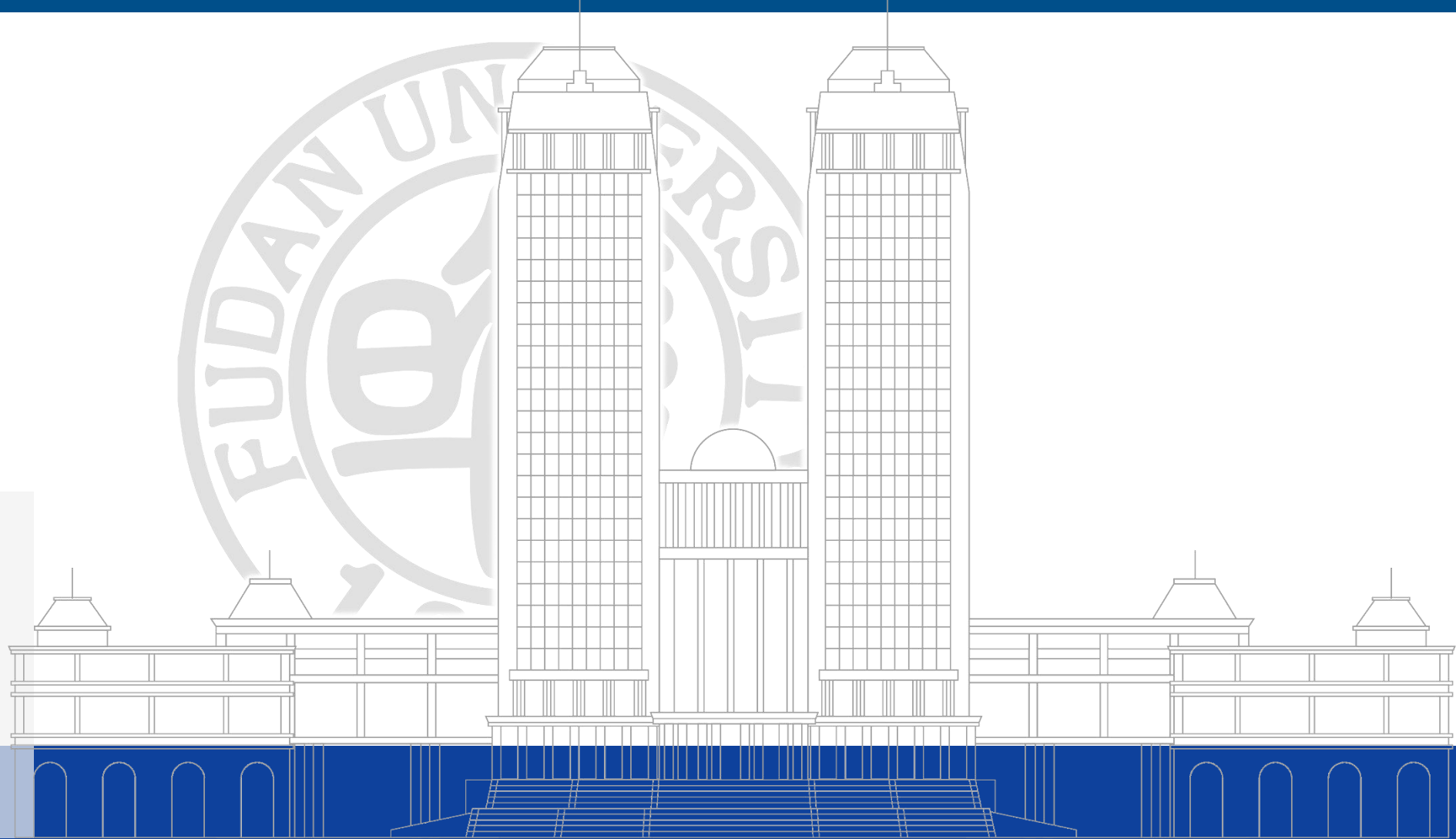


基础知识-统计学设计

医学统计 & SPSS

- **完全随机设计**：又称成组设计，主要有以下两种形式：一是采用完全随机分组，将全部同质受试对象随机分配到各个处理组，各组分别接受不同的处理；二是分别从不同的总体中进行随机抽样。
- **配对设计**：包括异体配对和自体配对。
- **随机区组设计**：将受试对象按性质相同或相近分成若干区组(或称配伍组)，将每个区组的各个受试对象随机分配到不同的处理组接受不同的处理。
- **重复测量设计**：给予一种或多种处理后，在多个时间点从同一个受试对象重复获得某指标观察值的设计称为重复测量设计。

02



SPSS软件基础知识



SPSS概述

医学统计 & SPSS

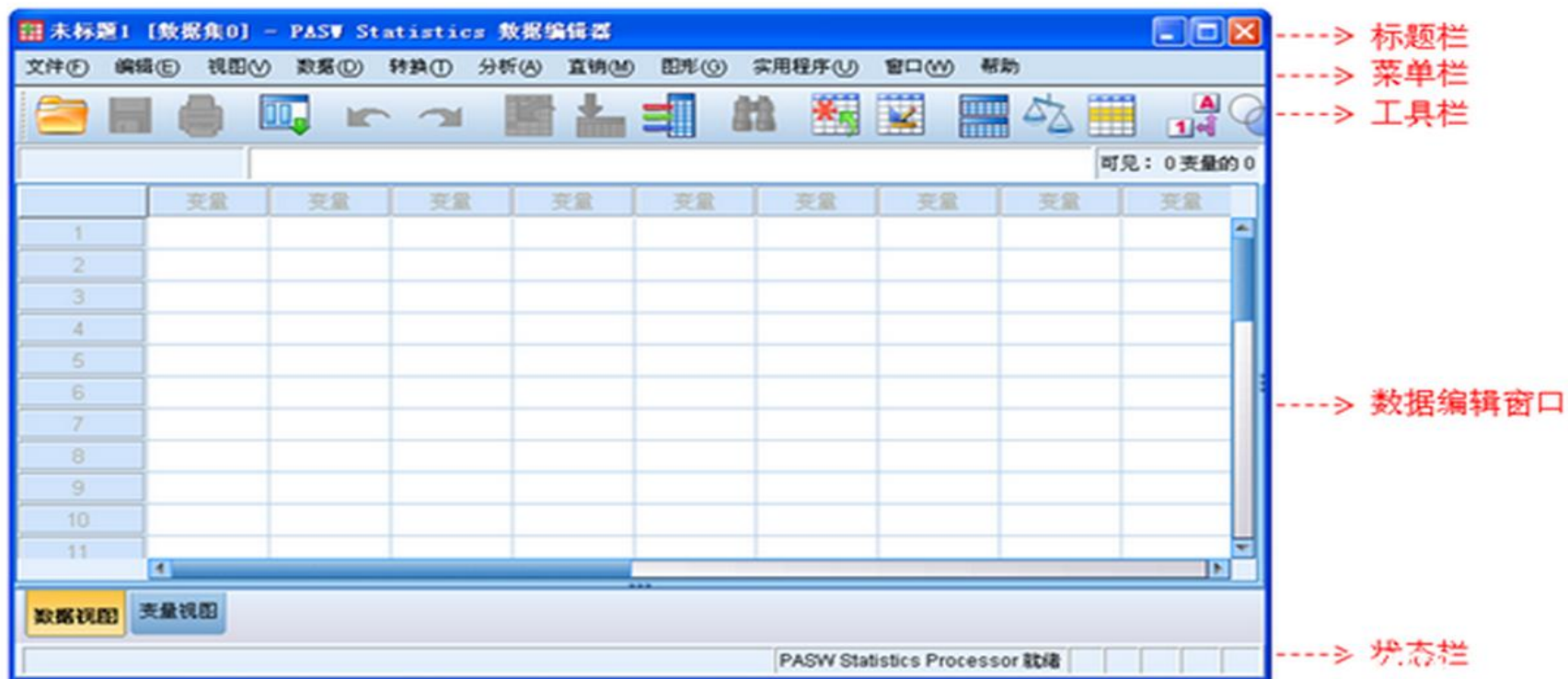


- SPSS是世界上最早的统计分析软件，由美国斯坦福大学的三位研究生Norman H. Nie、C. Hadlai (Tex) Hull 和 Dale H. Bent 于1968年研究开发成功，同时成立了SPSS公司，并于1975年成立法人组织、在芝加哥组建了SPSS总部。
- 2009年7月28日，IBM公司宣布将用12亿美元现金收购统计分析软件提供商SPSS公司。如今SPSS已出至版本22.0，而且更名为IBM SPSS。迄今，SPSS公司已有40余年的成长历史。



SPSS界面

医学统计 & SPSS



SPSS主界面主要包括两个：数据编辑窗口和结果输出窗口。

启动SPSS后，出现如图所示的空白数据编辑窗口，是一个类似于MS EXCEL界面的空文件。



SPSS界面

医学统计 & SPSS

The screenshot shows the SPSS Variable View window for the file '1991 U.S. General Social Survey.sav'. The window title is '*1991 U.S. General Social Survey.sav [数据集1] - PASW Statistics 数据编辑器'. The menu bar includes '文件(F)', '编辑(E)', '视图(V)', '数据(D)', '转换(T)', '分析(A)', '直销(M)', '图形(G)', '实用程序(U)', '窗口(W)', and '帮助'. The toolbar contains icons for file operations, data management, and analysis. The main area displays a table of variables with columns for '名称', '类型', '宽度', '小数', '标签', and '值'. The '值' column shows missing values for several variables. At the bottom, there are tabs for '数据视图' and '变量视图', with '变量视图' selected. The status bar at the bottom right indicates 'PASW Statistics Processor 就绪'.

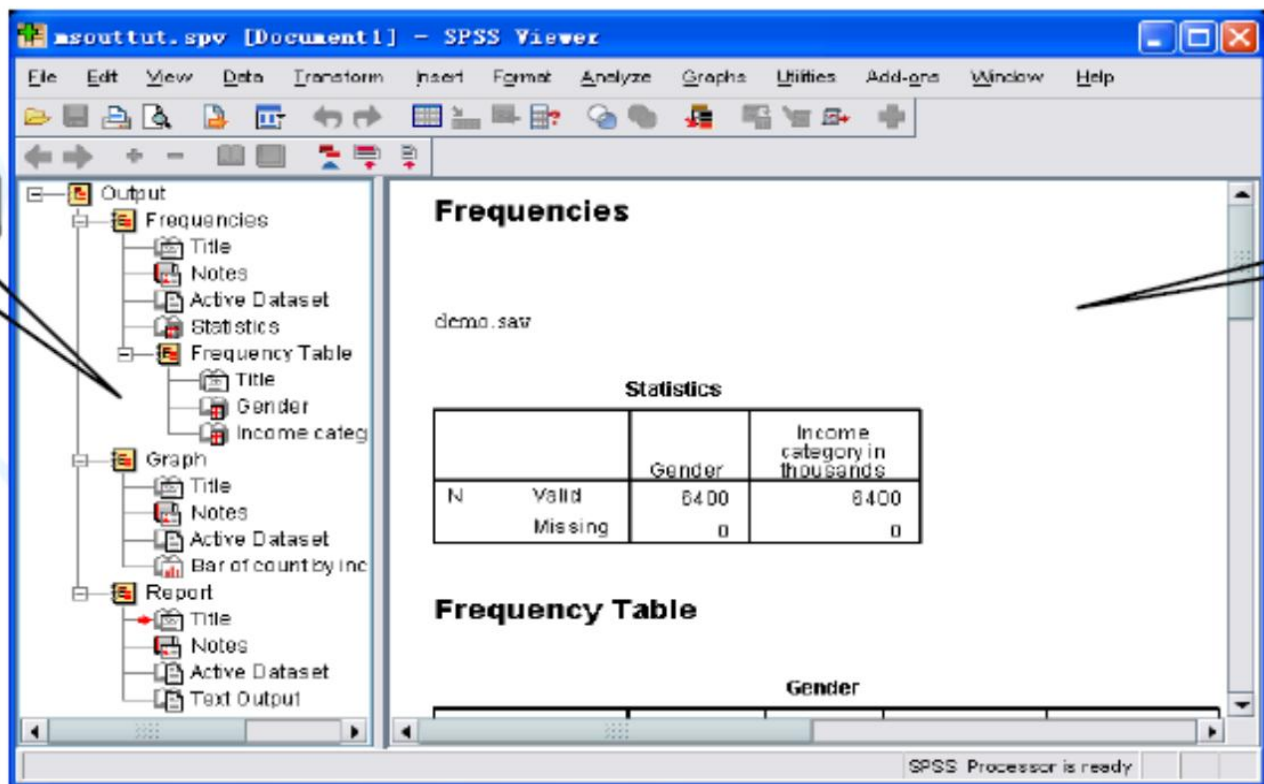
| | 名称 | 类型 | 宽度 | 小数 | 标签 | 值 |
|----|--------|-------|----|----|---------------------------------------|-----------|
| 1 | sex | 数值(N) | 1 | 0 | Respondent's Sex | {1, Male} |
| 2 | race | 数值(N) | 1 | 0 | Race of Respondent | {1, White |
| 3 | region | 数值(N) | 8 | 2 | Region of the United States | {1.00, No |
| 4 | happy | 数值(N) | 1 | 0 | General Happiness | {0, NAP} |
| 5 | life | 数值(N) | 1 | 0 | Is Life Exciting or Dull | {0, NAP} |
| 6 | sibs | 数值(N) | 2 | 0 | Number of Brothers and Sisters | {98, DK} |
| 7 | childs | 数值(N) | 1 | 0 | Number of Children | {0, 0} |
| 8 | age | 数值(N) | 2 | 0 | Age of Respondent | {98, DK} |
| 9 | educ | 数值(N) | 2 | 0 | Highest Year of School Completed | {97, NAP |
| 10 | paeduc | 数值(N) | 2 | 0 | Highest Year School Completed, Father | {97, NAP |
| 11 | maeduc | 数值(N) | 2 | 0 | Highest Year School Completed, Mother | {97, NAP |
| 12 | speduc | 数值(N) | 2 | 0 | Highest Year School Completed, Spouse | {97, NAP |

- 变量视图窗口是SPSS数据变量格式设置窗口，包含了名称、类型、宽度、小数、标签、值、缺失、列、度量标准、角色等11项，其中常用的是名称、类型、宽度、标签、值、度量标准等。



SPSS界面

医学统计 & SPSS



- SPSS结果输出窗口名为 Viewer，它可以显示和管理 SPSS 统计分析结果、报表及图形。该窗口中的内容能够以结果文件*.spv 的形式保存。



SPSS数据导入

医学统计 & SPSS

- 方法一：在数据编辑窗口录入，需同时录入数据，并且定义变量的属性。
 - 方法二：外部文件导入SPSS，但也需要定义变量的属性。
- ✓ 定义变量属性需要定义变量名，变量类型，变量长度（小数位数），变量标签和变量的格式等。



SPSS数据导入-读取EXCEL文件

医学统计 & SPSS





数据格式

医学统计 & SPSS

01

第一

Wide型

不同观测对象的数据不能在同一记录中出现，即同一观测数据应当独占一行。

02

第二

Long型

每一个测量指标或影响因素只能占据一列的位置，即同一指标的观测值都应当录入到同一个变量中去。



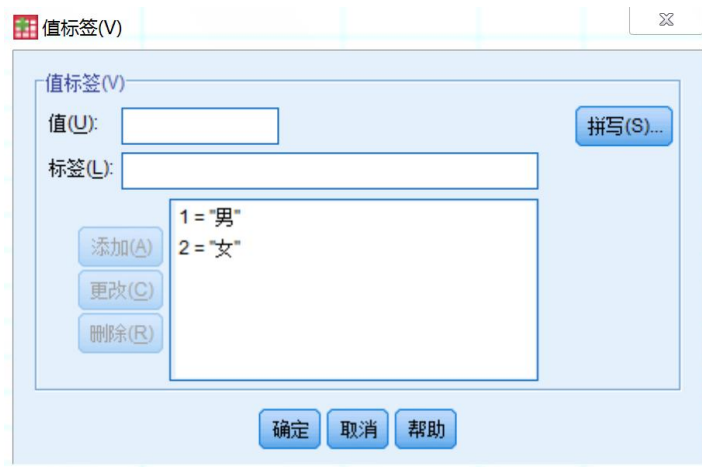
SPSS变量属性设置

医学统计 & SPSS

SPSS变量类型

- 标准数值型 (NUMERIC) : 12, 0.25...
- 逗号数值型 (COMMA) : 即整数部分每3位数加一逗号, 如: 100,000;
- 加点数值型 (DOT)
- 科学计数型 (SCIENTIFIC NOTATION)
- 货币数值型 (DOLLAR)
- 自定义型 (CUSTOM CURRENCY)
- 字符型 (STRING) : 张三, ENGLISH...
- 日期型 (DATE) : 出生日期...

变量值标签

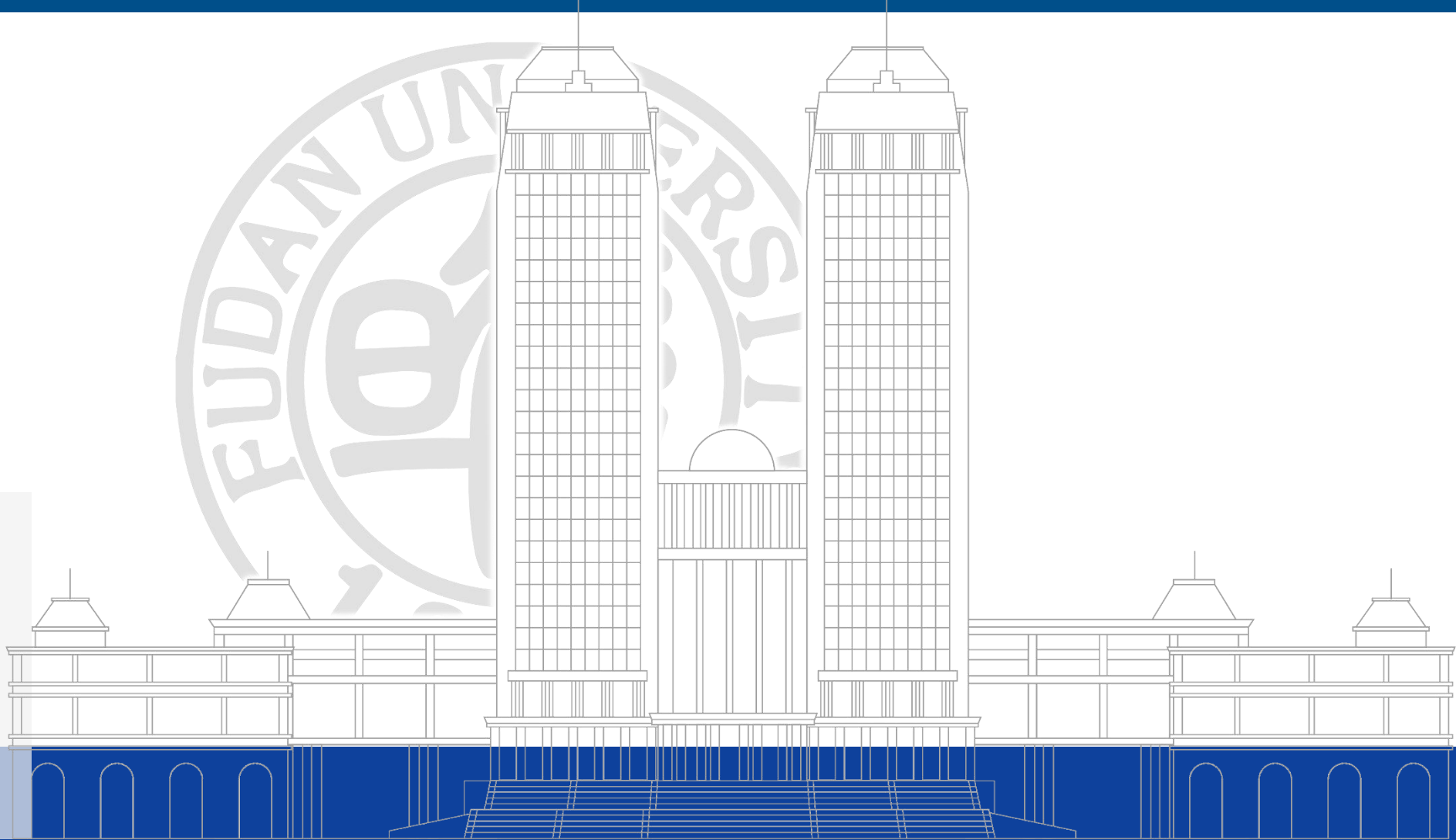


度量标准



- 度量: 连续性变量
- 序号: 有序或者等级变量
- 名义: 无序变量

03



统计描述及SPSS实现



统计描述指标

医学统计 & SPSS

定量资料

• 集中趋势:

➢ 算数均数(mean)

➢ 几何均数 $G = \sqrt[n]{X_1 X_2 \cdots X_n}$

➢ 中位数(median)

• 离散趋势:

➢ 极差(range)

➢ 四分位间距(P75-P25)或者四分位数范围(P25~P75)

➢ 方差(variance)和标准差 $s^2 = \frac{\sum (X - \bar{X})^2}{n-1}$ $s = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}}$

➢ 变异系数(CV): 标准差/均数*100%

□ 如果资料近似呈对称分布, 则 $\bar{x} \pm s$

□ 如果资料取对数后才呈对称分布, 则

$$\bar{y}(G) \pm s_y \quad \text{其中 } y = \ln(x)$$

□ 如果资料呈偏态分布 $Median(P_{25} \sim P_{75})$

□ 在确认性研究中, 往往统计描述采用均数, 中位数, 标准差, P_{25} , P_{75} , 最小值, 最大值。



定量资料的统计描述实例

医学统计 & SPSS

某市1995年110名7岁男童身高资料

| | | | | | | | | | |
|-------|-------|-------|-------|-------|-------|-------|--------------|-------|-------|
| 121.4 | 119.2 | 124.7 | 125 | 115 | 112.8 | 120.2 | <u>110.2</u> | 120.9 | 120.1 |
| 125.5 | 120.3 | 122.3 | 118.2 | 116.7 | 121.7 | 116.8 | 121.6 | 120.2 | 122 |
| 121.7 | 118.8 | 121.8 | 124.5 | 121.7 | 122.7 | 116.3 | 124 | 119 | 124.5 |
| 121.8 | 124.9 | 130 | 123.5 | 128.1 | 119.7 | 126.1 | 131.3 | 123.8 | 116.7 |
| 122.2 | 122.8 | 128.6 | 122 | 132.5 | 122 | 123.5 | 116.3 | 126.1 | 119.2 |
| 126.4 | 118.4 | 121 | 119.1 | 116.9 | 131.1 | 120.4 | 115.2 | 118 | 122.4 |
| 120.3 | 116.9 | 126.4 | 114.2 | 127.2 | 118.3 | 127.8 | 123 | 117.4 | 123.2 |
| 119.9 | 122.1 | 120.4 | 124.8 | 122.1 | 114.4 | 120.5 | 120 | 122.8 | 116.8 |
| 125.8 | 120.1 | 124.8 | 122.7 | 119.4 | 128.2 | 124.1 | 127.2 | 120 | 122.7 |
| 118.3 | 127.1 | 122.5 | 116.3 | 125.1 | 124.4 | 112.3 | 121.3 | 127 | 113.5 |
| 118.8 | 127.6 | 125.2 | 121.5 | 122.5 | 129.1 | 122.6 | <u>134.5</u> | 118.3 | 132.8 |

| | height |
|----|--------|
| 1 | 121.4 |
| 2 | 125.5 |
| 3 | 121.7 |
| 4 | 121.8 |
| 5 | 122.2 |
| 6 | 126.4 |
| 7 | 120.3 |
| 8 | 119.9 |
| 9 | 125.8 |
| 10 | 118.3 |
| 11 | 118.8 |
| 12 | 119.2 |
| 13 | 120.3 |
| 14 | 118.8 |
| 15 | 124.9 |
| 16 | 122.8 |
| 17 | 118.4 |
| 18 | 116.9 |
| 19 | 122.1 |
| 20 | 120.1 |
| 21 | 127.1 |
| 22 | 127.6 |
| 23 | 124.7 |
| 24 | 122.2 |

| 名称 | 类型 | 宽度 | 小数 | 标签 | 值 | 缺失 | 列 | 对齐 | 度量标准 |
|--------|-------|----|----|----|---|----|---|----|-------|
| height | 数值(N) | 6 | 1 | | 无 | 无 | 8 | 右 | 度量(S) |



定量资料的统计描述实例

医学统计 & SPSS

输出

- 日志
- 频率
 - 标题
 - 附注
 - 活动的数据集
 - 统计量
 - height

统计量

| height | | |
|--------|----|---------|
| N | 有效 | 110 |
| | 缺失 | 0 |
| 均值 | | 121.945 |
| 中值 | | 121.900 |
| 方差 | | 20.495 |
| 全距 | | 24.3 |
| 极小值 | | 110.2 |
| 极大值 | | 134.5 |
| 百分位数 | 25 | 119.075 |
| | 50 | 121.900 |
| | 75 | 124.800 |

集中趋势

- 均值(M)
- 中位数
- 众数(O)
- 合计

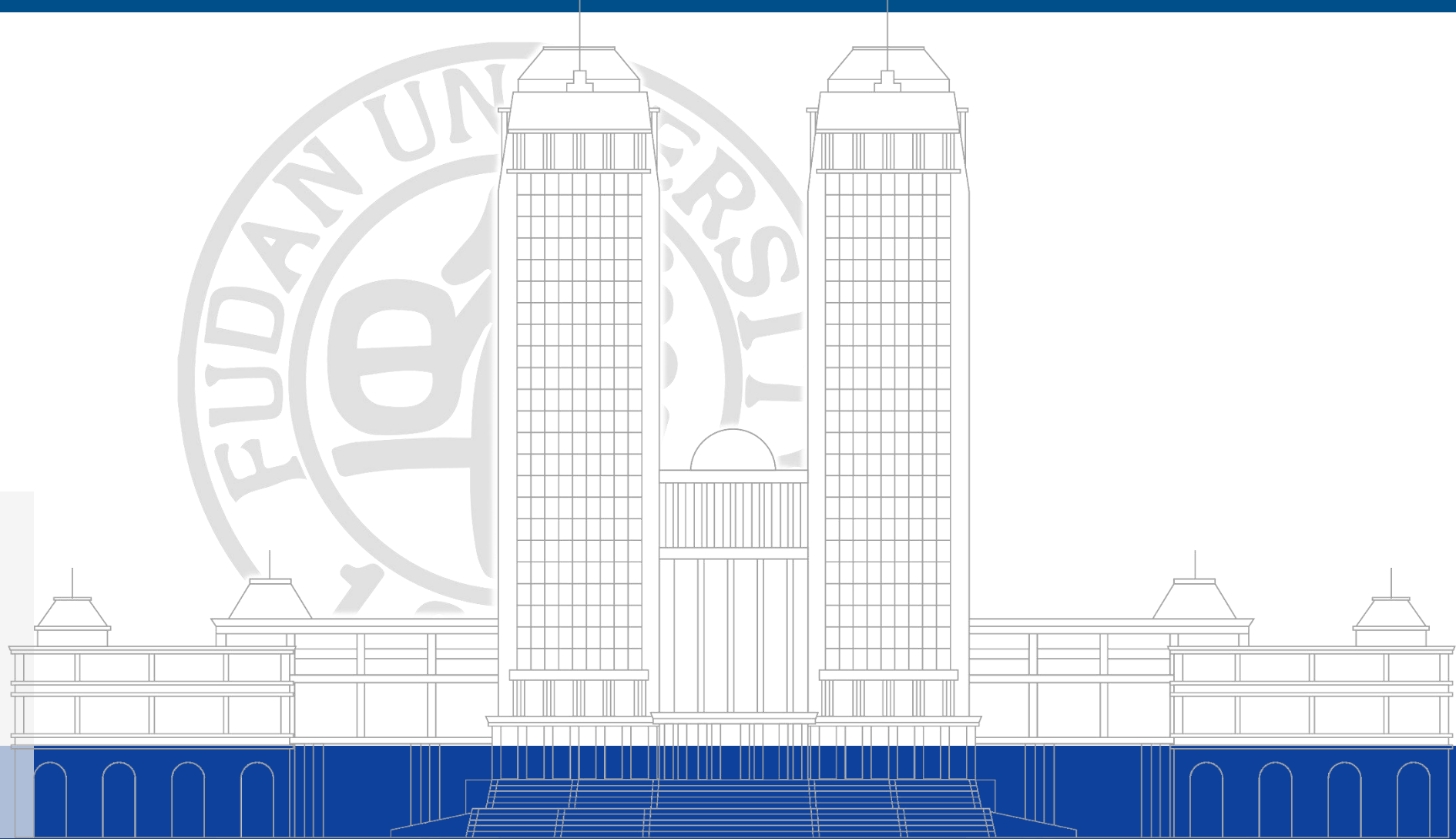
值为组的中点(L)

分布

- 偏度
- 峰度

帮助

04

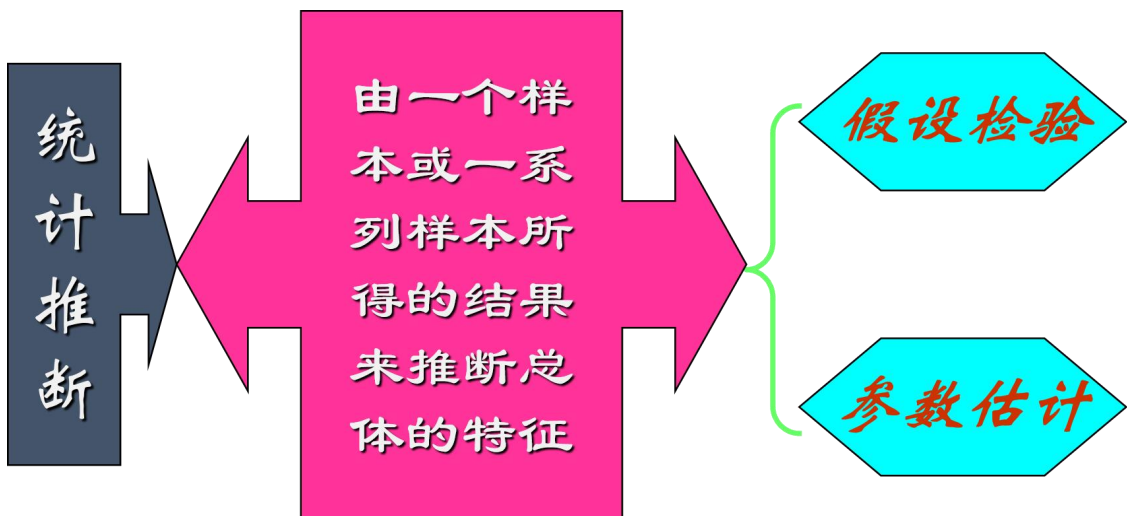


统计推断及SPSS实现



统计推断基础

医学统计 & SPSS



参数估计 (可信区间)

- 按一定的概率或可信度($1-\alpha$)用一个区间来估计总体参数所在的范围,该范围通常称为参数的可信区间或者置信区间,预先给定的概率($1-\alpha$)称为可信度或者置信度,常取95%或99%.

假设检验

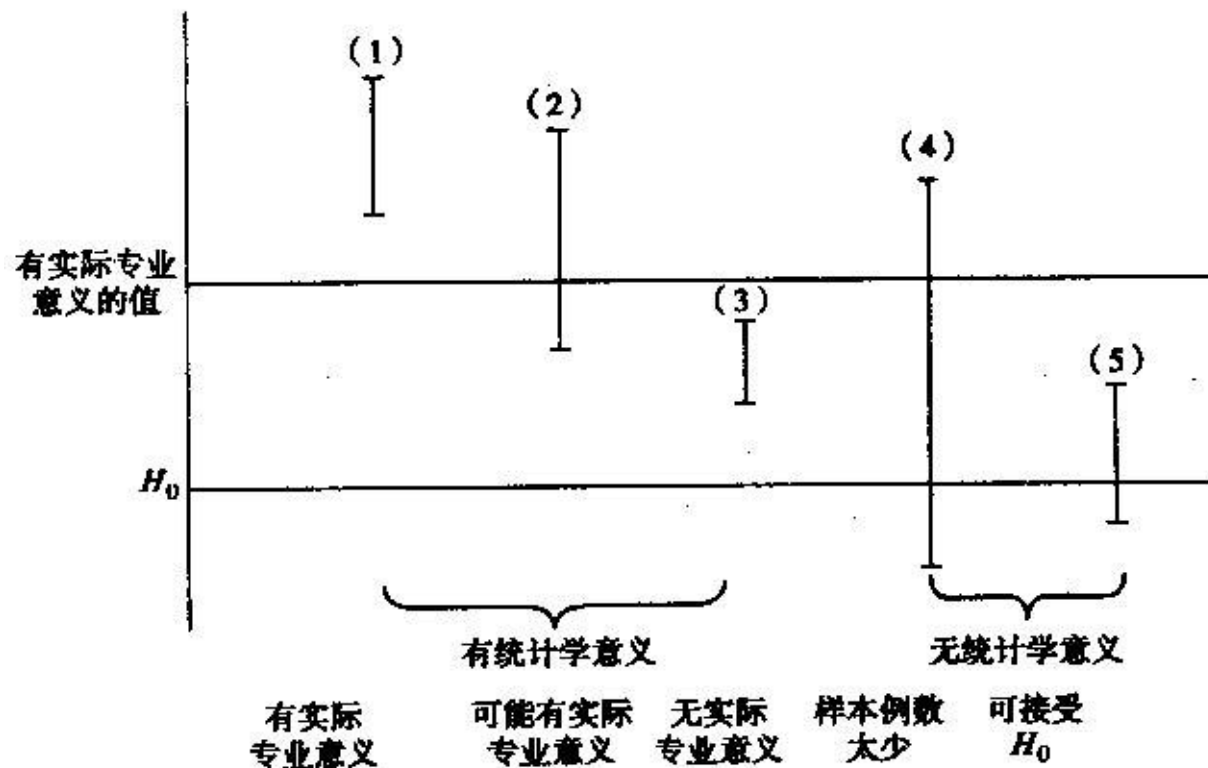
- 小概率思想是指小概率事件 ($P < 0.01$ 或 $P < 0.05$) 在一次试验中基本上不会发生。
- 假设检验的反证法思想: 先根据检验假设 H_0 , 建立适当的统计量, 确定假设 H_0 成立情况下服从某个概率分布, 定一个范围。 H_0 成立时, 统计量进入这个范围, 是一个小概率事件 ($P \leq 0.05$ 或更小), H_0 不成立时, 统计量进入这个范围的概率较大。
- 如果实际的抽样样本统计量进入这个范围, 对 H_0 成立情况下是一个小概率事件, 一般不会发生, 由此推断假设 H_0 不成立。这就是小概率反证法思想。



可信区间和假设检验

医学统计 & SPSS

- $P > 0.05$ 时的解读需慎重，不能拒绝无效假设不等同于接受无效假设（无有之据，非有无之据）
- 可信区间同样可用于回答假设检验的问题
- 可信区间比假设检验提供更多的信息，不但可以回答有无统计学意义，还可回答有无实际意义





统计学方法的选择

医学统计 & SPSS

• 统计分析原则

- 研究设计（完全随机分组、配对设计或随机区组设计、关联性分析）
- 数据类型（连续型变量、无序分类变量、有序分类变量）
- 适用条件（统计分析方法的优与劣，对与错）

• 统计分析的流程

- 从Table 1 to Table n
- 从单因素 to 多因素
- 从表 to 图



完全随机分组设计 两组定量资料统计分析的适用条件

- 大样本资料或服从正态分布的小样本资料
 - (1) 若方差齐性，则作成组t检验
 - (2) 若方差不齐，则作t' 检验或用成组的Wilcoxon秩和检验
- 小样本偏态分布资料，则用成组的Wilcoxon秩和检验



实例一 成组设计两组资料统计分析

医学统计 & SPSS

下面资料是关于18名单腿截肢者的健康足和18名正常健康人的足部相同部位组织切片毛细血管密度(/mm²)的测定结果, 试比较健康人和截肢者足部毛细血管密度有无差别?

➤ 正常人

| | | | | | | | |
|----|----|----|----|----|----|----|----|
| 16 | 30 | 29 | 33 | 28 | 28 | 36 | 29 |
| 27 | 33 | 37 | 38 | 40 | 41 | 39 | 39 |
| 39 | 48 | | | | | | |

➤ 截肢者

| | | | | | | | |
|----|----|----|----|----|----|----|----|
| 10 | 21 | 28 | 28 | 26 | 20 | 33 | 26 |
| 15 | 23 | 23 | 30 | 31 | 26 | 23 | 42 |
| 24 | 28 | | | | | | |



SPSS数据集样式

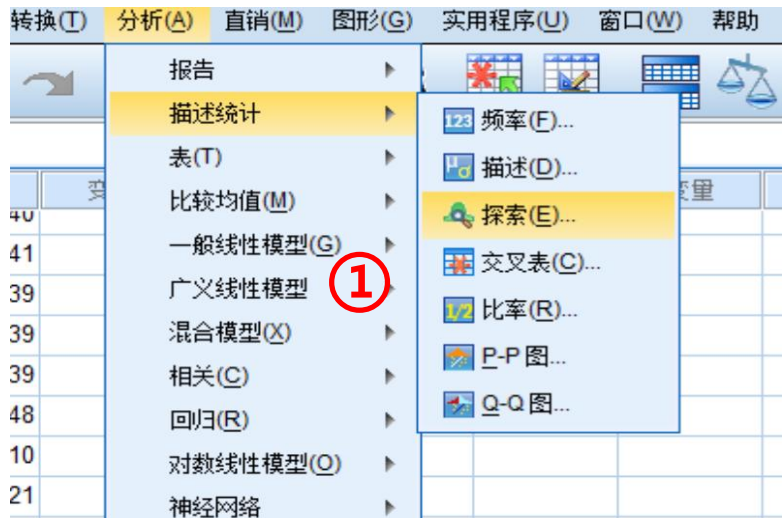
医学统计 & SPSS

| | group | y | |
|----|-------|----|--|
| 13 | 1 | 40 | |
| 14 | 1 | 41 | |
| 15 | 1 | 39 | |
| 16 | 1 | 39 | |
| 17 | 1 | 39 | |
| 18 | 1 | 48 | |
| 19 | 2 | 10 | |
| 20 | 2 | 21 | |
| 21 | 2 | 28 | |
| 22 | 2 | 28 | |
| 23 | 2 | 26 | |
| 24 | 2 | 20 | |
| 25 | 2 | 33 | |



正态性检验流程

医学统计 & SPSS



正态性检验结果 $\alpha=0.05$

正态性检验

| group | Kolmogorov-Smirnov ^a | | | Shapiro-Wilk | | |
|-------|---------------------------------|----|-------------------|--------------|----|------|
| | 统计量 | df | Sig. | 统计量 | df | Sig. |
| y 正常人 | .116 | 18 | .200 [*] | .949 | 18 | .404 |
| 截肢者 | .143 | 18 | .200 [*] | .958 | 18 | .564 |

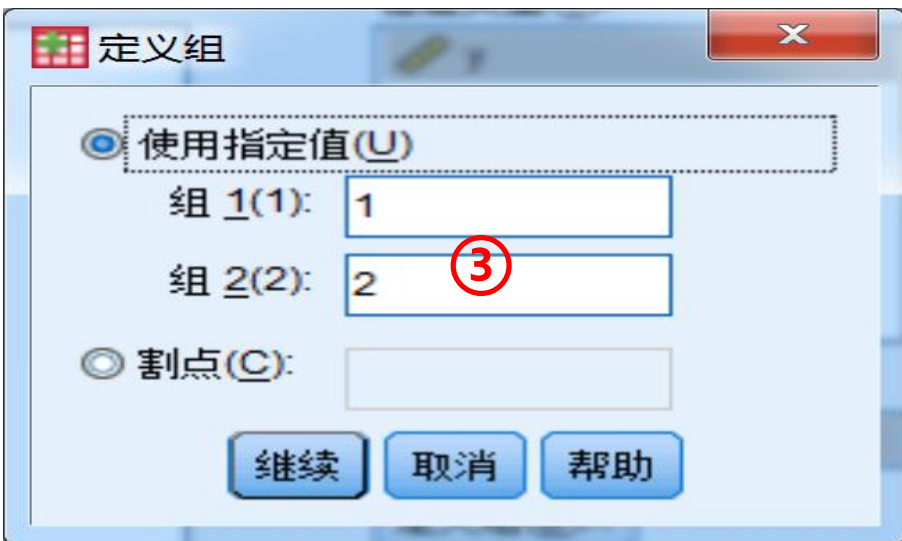
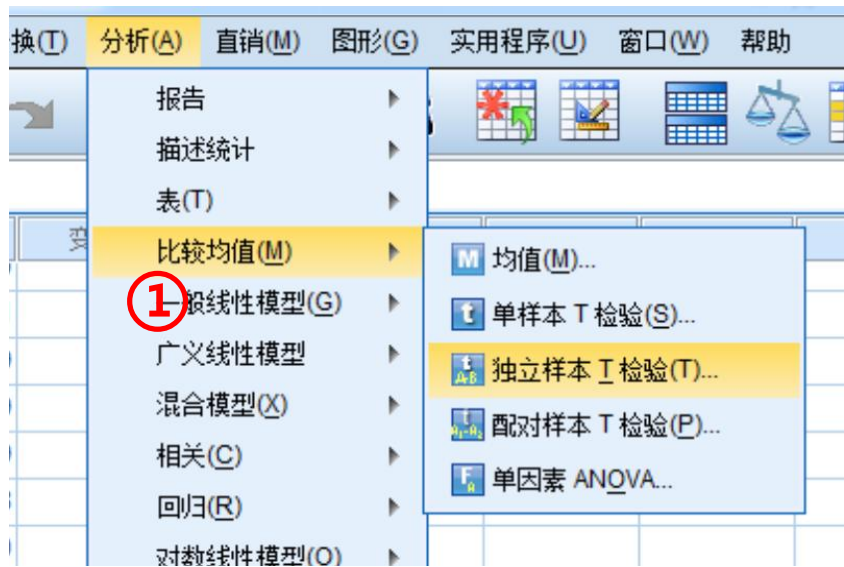
a. Lilliefors 显著水平修正

*. 这是真实显著水平的下限。



成组t检验

医学统计 & SPSS





成组t检验结果

医学统计 & SPSS

T检验

[数据集1]

组统计量

| group | N | 均值 | 标准差 | 均值的标准误 |
|-------|----|-------|-------|--------|
| y 正常人 | 18 | 33.89 | 7.267 | 1.713 |
| 截肢者 | 18 | 25.39 | 6.946 | 1.637 |

独立样本检验

| | 方差方程的 Levene 检验 | | 均值方程的 t 检验 | | | | | | |
|----------|-----------------|------|------------|--------|----------|-------|-------|--------------|--------|
| | F | Sig. | t | df | Sig.(双侧) | 均值差值 | 标准误差值 | 差分的 95% 置信区间 | |
| | | | | | | | | 下限 | 上限 |
| y 假设方差相等 | .343 | .562 | 3.587 | 34 | .001 | 8.500 | 2.370 | 3.685 | 13.315 |
| 假设方差不相等 | | | 3.587 | 33.931 | .001 | 8.500 | 2.370 | 3.684 | 13.316 |

结果解析:

1. 方差齐性检验的显著性水平为 0.562，大于0.10，两组方差齐性。
2. t检验显著性水平为0.001，两组毛细血管密度总体均数不同，由于正常人毛细血管密度均数更高，可以推断正常人的毛细血管密度高于截肢者。
3. 差值及95%CI: 8.50 (3.685-13.315)



适用条件

医学统计 & SPSS

配对设计

两组定量资料统计分析的适用条件

- 大样本资料或配对差值服从正态分布的小样本资料，作配对t检验
- 小样本并且差值呈偏态分布资料，则用Wilcoxon的符号配对秩检验



实例二 配对设计两组资料统计分析

医学统计 & SPSS

例 为比较两种方法对乳酸饮料中脂肪含量测定结果是否不同，某人随机抽取了 10 份乳酸饮料制品，分别用脂肪酸水解法和哥特里—罗紫法测定其结果如下表第(1)~(3)栏。

问两法测定结果是否不同？

表 两种方法对乳酸饮料中脂肪含量的测定结果(%)

| 编号 (1) | 哥特里—罗紫法 (2) | 脂肪酸水解法 (3) | 差值 d (4)=(2)-(3) |
|-----------|----------------|---------------|-----------------------|
| 1 | 0.840 | 0.580 | 0.260 |
| 2 | 0.591 | 0.509 | 0.082 |
| 3 | 0.674 | 0.500 | 0.174 |
| 4 | 0.632 | 0.316 | 0.316 |
| 5 | 0.687 | 0.337 | 0.350 |
| 6 | 0.978 | 0.517 | 0.461 |
| 7 | 0.750 | 0.454 | 0.296 |
| 8 | 0.730 | 0.512 | 0.218 |
| 9 | 1.200 | 0.997 | 0.203 |
| 10 | 0.870 | 0.506 | 0.364 |



SPSS数据集样式

医学统计 & SPSS

| | x1 | x2 | 变量 |
|----|-------|------|----|
| 1 | .840 | .580 | |
| 2 | .591 | .509 | |
| 3 | .674 | .500 | |
| 4 | .632 | .316 | |
| 5 | .687 | .337 | |
| 6 | .978 | .517 | |
| 7 | .750 | .454 | |
| 8 | .730 | .512 | |
| 9 | 1.200 | .997 | |
| 10 | .870 | .506 | |
| 11 | | | |
| 12 | | | |



生成差值

医学统计 & SPSS

数据(D) 转换(T) 分析(A) 直销(M) 图形(G) 实用程序(U)

1

- 计算变量(C)...
- 对个案内的值计数(O)...
- 转换值(E)...
- 重新编码为相同变量(S)...
- 重新编码为不同变量(R)...
- 自动重新编码(A)...
- 可视离散化(B)...

计算变量

目标变量(T): diff

数字表达式(E): x1 - x2

2

哥特里-罗紫法 [x1]
脂肪酸水解法 [x2]

函数组(G):
全部
算术
CDF 与非中心 CDF
转换
当前日期时间
日期运算
日期创建
函数和特殊变量(F):

如果(I)... (可选的个案选择条件)

确定 粘贴(P) 重置(R) 取消 帮助

| | x1 | x2 | diff |
|---|-------|------|------|
| | .840 | .580 | .26 |
| | .591 | .509 | .08 |
| | .674 | .500 | .17 |
| 3 | .632 | .316 | .32 |
| | .687 | .337 | .35 |
| | .978 | .517 | .46 |
| | .750 | .454 | .30 |
| | .730 | .512 | .22 |
| | 1.200 | .997 | .20 |
| | .870 | .506 | .36 |



差值正态性检验

医学统计 & SPSS



正态性检验

| | Kolmogorov-Smirnov ^a | | | Shapiro-Wilk | | |
|------|---------------------------------|----|-------------------|--------------|----|------|
| | 统计量 | df | Sig. | 统计量 | df | Sig. |
| diff | .100 | 10 | .200 [*] | .993 | 10 | .999 |

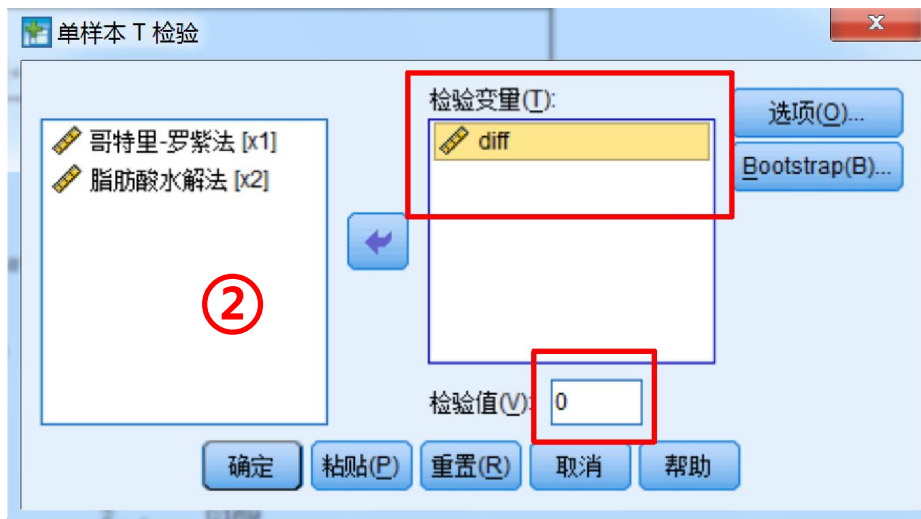
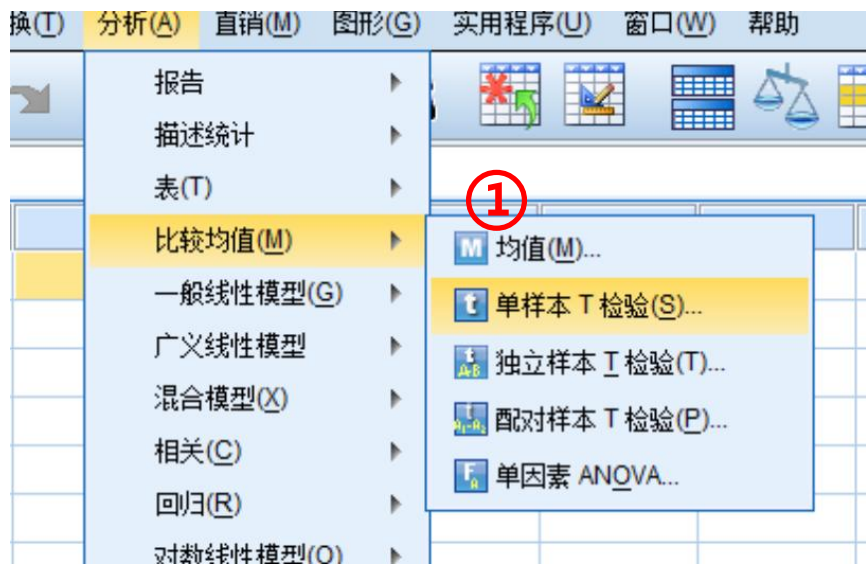
a. Lilliefors 显著水平修正

*. 这是真实显著水平的下限。



单样本t检验

医学统计 & SPSS



→ T检验

[数据集0]

③

单个样本统计量

| | N | 均值 | 标准差 | 均值的标准误 |
|------|----|-------|--------|--------|
| diff | 10 | .2724 | .10868 | .03437 |

单个样本检验

| 检验值 = 0 | | | | | | |
|---------|-------|----|----------|--------|--------------|-------|
| | t | df | Sig.(双侧) | 均值差值 | 差分的 95% 置信区间 | |
| | | | | | 下限 | 上限 |
| diff | 7.926 | 9 | .000 | .27240 | .1947 | .3501 |



完全随机分组设计 多组定量资料统计分析的适用条件

- 若大样本资料或服从正态分布，并且方差齐性，则作完全随机的方差分析。如果方差分析的统计检验为有统计学意义，则进一步作统计分析：选择合适的方法（如：LSD检验，Bonferroni检验等）进行两两比较。
- 如果小样本的偏态分布资料或方差不齐，则作Kruskal Wallis的统计检验。如果Kruskal Wallis的统计检验为有统计学意义，则进一步作统计分析：选择合适的方法（如：用成组的Wilcoxon秩和检验，但用Bonferroni方法校正P值等）进行两两比较。



实例三 成组设计多组资料统计分析

医学统计 & SPSS

例 8-1 为研究茶多酚保健饮料对急性缺氧的影响,某研究者将 60 只小白鼠随机分为低、中、高三个剂量组和一个对照组,每组 15 只小白鼠。对照组给予蒸馏水 0.25ml 灌胃,低、中、高剂量组分别给予 2.0g / kg、4.0g / kg、8.0g / kg 的饮料溶于 0.2~0.3ml 蒸馏水后灌胃。每天一次,40 天后,对小白鼠进行耐缺氧存活时间试验,结果如表 8-1。试比较不同剂量的茶多酚保健饮料对延长小白鼠的平均耐缺氧存活时间有无差别。

表 8-1 各组小白鼠耐缺氧时间

| 组别 | 耐缺氧时间 Y_{ij} (min) | | | | | | | | n_i | \bar{Y}_i | S_i |
|---------------------|----------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------------|-------|
| 对照组 ($i = 1$) | 20.79 | 22.91 | 27.21 | 19.34 | 17.85 | 23.79 | 22.60 | 18.53 | 15 | 21.32 | 3.40 |
| | 23.23 | 20.14 | 26.71 | 19.36 | 17.22 | 24.13 | 15.85 | | | | |
| 低剂量组 ($i = 2$) | 22.22 | 24.74 | 21.53 | 19.66 | 25.89 | 29.10 | 18.93 | 18.64 | 15 | 23.23 | 3.52 |
| | 26.39 | 25.49 | 20.43 | 22.69 | 29.67 | 20.36 | 22.74 | | | | |
| 中剂量组 ($i = 3$) | 28.56 | 28.67 | 25.28 | 30.38 | 23.13 | 23.47 | 28.88 | 29.62 | 15 | 28.14 | 4.00 |
| | 24.82 | 34.64 | 22.29 | 29.22 | 25.63 | 35.12 | 32.32 | | | | |
| 高剂量组 ($i = 4$) | 31.93 | 37.94 | 39.76 | 27.94 | 29.65 | 34.23 | 32.63 | 29.13 | 15 | 32.84 | 4.66 |
| | 39.62 | 36.15 | 28.85 | 24.07 | 29.29 | 35.24 | 36.13 | | | | |
| 合计 | | | | | | | | | 60 | 26.38 | 5.92 |



SPSS数据集样式

医学统计 & SPSS

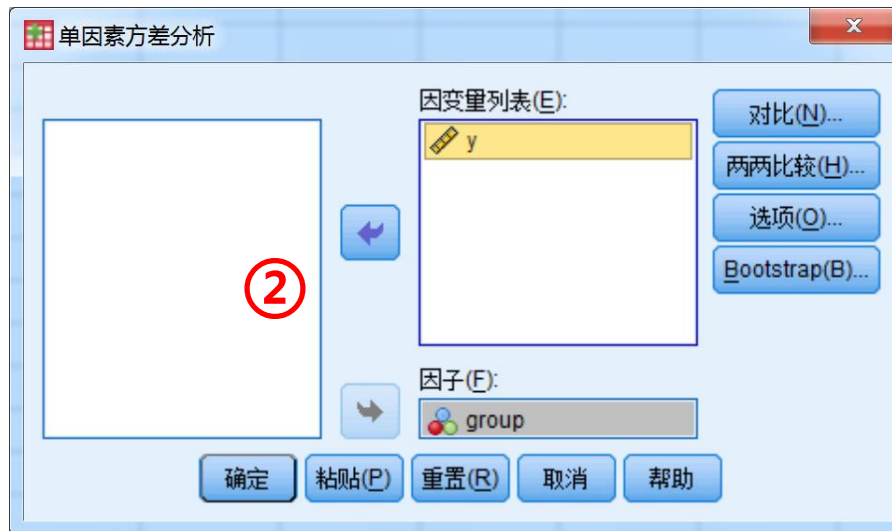
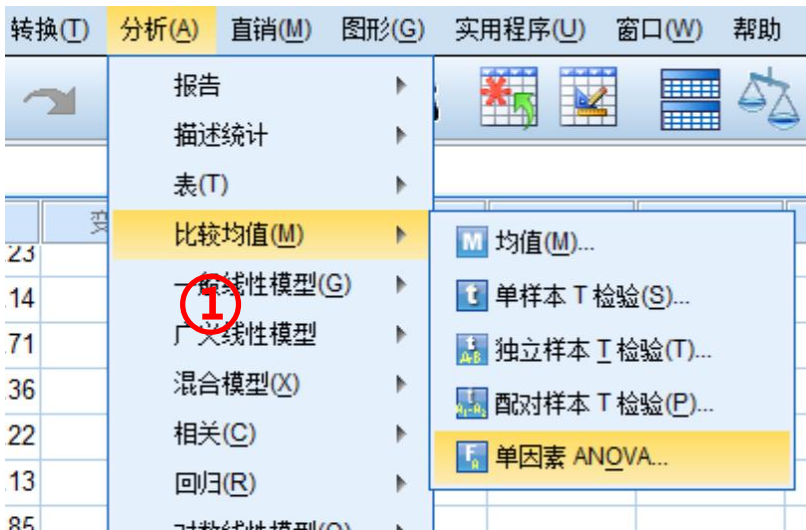
| | group | y |
|---|-------|-------|
| 9 | 1 | 23.23 |
| 0 | 1 | 20.14 |
| 1 | 1 | 26.71 |
| 2 | 1 | 19.36 |
| 3 | 1 | 17.22 |
| 4 | 1 | 24.13 |
| 5 | 1 | 15.85 |
| 6 | 2 | 22.22 |
| 7 | 2 | 24.74 |
| 8 | 2 | 21.53 |
| 9 | 2 | 19.66 |
| 0 | 2 | 25.89 |
| 1 | 2 | 28.10 |

| 名称 | 类型 | 宽度 | 小数 | 标签 | 值 | 缺失 | 列 | 对齐 | 度量标准 | 角色 |
|-------|-------|----|----|----|-------------|----|---|----|-------|----|
| group | 数值(N) | 1 | 0 | | {1, 对照组}... | 无 | 8 | 右 | 名义(N) | 输入 |
| y | 数值(N) | 8 | 2 | | 无 | 无 | 8 | 右 | 度量(S) | 输入 |



单因素方差分析

医学统计 & SPSS





单因素方差分析统计学检验结果

医学统计 & SPSS

方差齐性检验

y

| Levene 统计量 | df1 | df2 | 显著性 |
|------------|-----|-----|------|
| .830 | 3 | 56 | .483 |

ANOVA

y

| | 平方和 | df | 均方 | F | 显著性 |
|----|----------|----|---------|--------|------|
| 组间 | 1205.796 | 3 | 401.932 | 26.087 | .000 |
| 组内 | 862.797 | 56 | 15.407 | | |
| 总数 | 2068.593 | 59 | | | |

结果解析：

1. 方差齐性检验的显著性水平为 0.483，大于0.10，组间方差齐性。
2. 组间检验显著性水平为 <0.001 ，可以得出结论不同分组总体均数不全相等。
3. 接下来如需要可以做两两比较。



单因素方差分析两两结果

医学统计 & SPSS

| | (I) group | (J) group | 均值差 (I-J) | 标准误 | 显著性 | 95% 置信区间 | |
|------------|-----------|-----------|------------|---------|-------|----------|---------|
| | | | | | | 下限 | 上限 |
| LSD | 对照组 | 低剂量组 | -1.92133 | 1.43328 | .185 | -4.7925 | .9499 |
| | | 中剂量组 | -6.82467* | 1.43328 | .000 | -9.6959 | -3.9535 |
| | | 高剂量组 | -11.52667* | 1.43328 | .000 | -14.3979 | -8.6555 |
| | 低剂量组 | 对照组 | 1.92133 | 1.43328 | .185 | -.9499 | 4.7925 |
| | | 中剂量组 | -4.90333* | 1.43328 | .001 | -7.7745 | -2.0321 |
| | | 高剂量组 | -9.60533* | 1.43328 | .000 | -12.4765 | -6.7341 |
| | 中剂量组 | 对照组 | 6.82467* | 1.43328 | .000 | 3.9535 | 9.6959 |
| | | 低剂量组 | 4.90333* | 1.43328 | .001 | 2.0321 | 7.7745 |
| | | 高剂量组 | -4.70200* | 1.43328 | .002 | -7.5732 | -1.8308 |
| | 高剂量组 | 对照组 | 11.52667* | 1.43328 | .000 | 8.6555 | 14.3979 |
| | | 低剂量组 | 9.60533* | 1.43328 | .000 | 6.7341 | 12.4765 |
| | | 中剂量组 | 4.70200* | 1.43328 | .002 | 1.8308 | 7.5732 |
| Bonferroni | 对照组 | 低剂量组 | -1.92133 | 1.43328 | 1.000 | -5.8417 | 1.9990 |
| | | 中剂量组 | -6.82467* | 1.43328 | .000 | -10.7450 | -2.9043 |
| | | 高剂量组 | -11.52667* | 1.43328 | .000 | -15.4470 | -7.6063 |
| | 低剂量组 | 对照组 | 1.92133 | 1.43328 | 1.000 | -1.9990 | 5.8417 |
| | | 中剂量组 | -4.90333* | 1.43328 | .007 | -8.8237 | -.9830 |
| | | 高剂量组 | -9.60533* | 1.43328 | .000 | -13.5257 | -5.6850 |
| | 中剂量组 | 对照组 | 6.82467* | 1.43328 | .000 | 2.9043 | 10.7450 |
| | | 低剂量组 | 4.90333* | 1.43328 | .007 | .9830 | 8.8237 |
| | | 高剂量组 | -4.70200* | 1.43328 | .011 | -8.6223 | -.7817 |
| | 高剂量组 | 对照组 | 11.52667* | 1.43328 | .000 | 7.6063 | 15.4470 |
| | | 低剂量组 | 9.60533* | 1.43328 | .000 | 5.6850 | 13.5257 |
| | | 中剂量组 | 4.70200* | 1.43328 | .011 | .7817 | 8.6223 |

*. 均值差的显著性水平为 0.05。

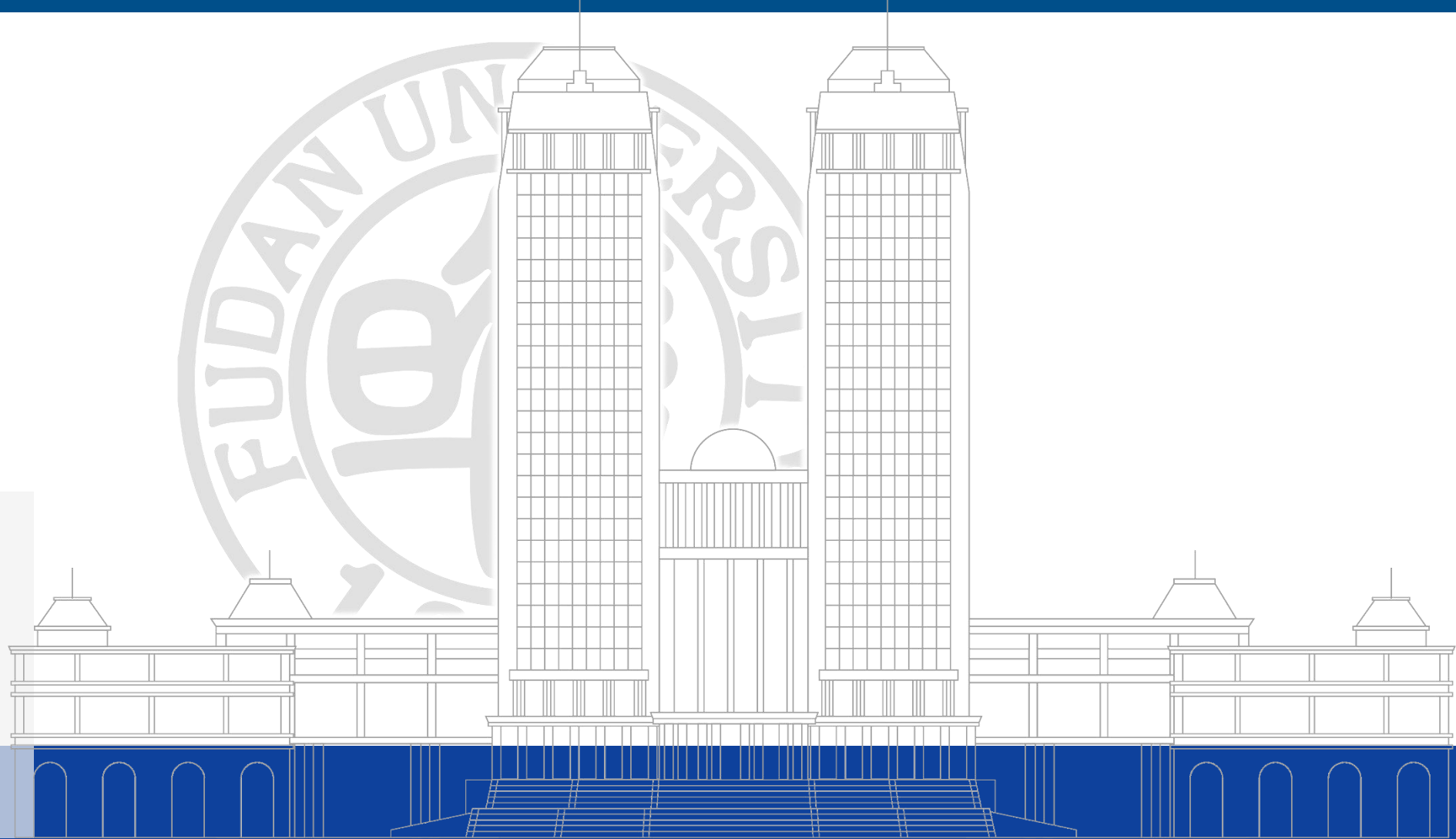


定量资料统计分析思维导图

医学统计 & SPSS



05



多因素统计分析方法拓展



多因素回归

医学统计 & SPSS

| 例号 | X_1 | X_2 | ... | X_p | Y |
|----|----------|----------|-----|----------|-------|
| 1 | X_{11} | X_{12} | ... | X_{1p} | Y_1 |
| 2 | X_{21} | X_{22} | ... | X_{2p} | Y_2 |
| ⋮ | ⋮ | ⋮ | ... | ⋮ | ⋮ |
| n | X_{n1} | X_{n2} | ... | X_{np} | Y_n |

Y 为定量变量——**Linear Regression**

Y 为二项分类变量——**Binary Logistic Regression**

Y 为多项分类变量——**Multinomial Logistic Regression**

Y 为有序分类变量——**Ordinal Logistic Regression**

Y 为生存时间与生存结局——**Cox Regression**

• 应用范围

- 校正混杂因素
- 筛选关联或者风险因素
- 预测与判断：预测事件发生概率，生存期或者生存概率等



多因素回归

医学统计 & SPSS

多因素回归方程的一般形式

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m + e$$

β_1 、 β_2 、 β_m 为偏回归系数 (Partial regression coefficient)

意义：如 β_1 表示在 X_2 、 X_3 …… X_m 固定条件下， X_1 每增减一个单位对 Y 的效应（ Y 增减 β 个单位）。