

## SPSS基础与应用

### 数据录入

李丁  
社会与人口学院

## 课程大纲

- 问卷信息
- SPSS数据基本格式
- 数据录入

## SPSS数据的基本格式

变量列

案例行

	time	id	v1	v2	v3	v4	v5	v6	v7	v8	v9
1	2007/04/01	1	100	1	20	4	3.00	2	4		
2	2007/04/02	2	100	1	24	2	3.00	2	8		
3	2007/04/03	3	200	1	20	2	9.00	2	2		
4	2007/04/04	4	100	2	65	3	1.00	1	2		
5	2007/04/05	5	200	2	40	2	3.00	1	99		
6	2007/04/06	6	100	1	50	3	3.00	1	6		
7	2007/04/07	7	100	2	53	1	10.00	1	7		
8	2007/04/08	8	300	1	44	2	9.00	1	99		
9	2007/04/09	9	200	2	35	3	3.00	1	5		
10	2007/04/10	10	200	1	21	4	7.00	2	7		
11	2007/04/11	11	300	1	34	4	1.00	1	8		
12	2007/04/12	12	200	2	39	2	3.00	1	7		

规则：一般地，一个案例占一行，一个变量占一列。

## 定义变量属性

○ 为了对变量进行统计分析，还需要定义变量属性

名称	类型	宽度	小数	格式	值	标度	缺失	度量
time	日期型(D)	8	0	短日期	无	无	无	日期
id	数值型(N)	4	0	ID	无	无	无	标度
v1	数值型(N)	4	0	ID	无	无	无	标度
v2	数值型(N)	3	0	ID	无	无	无	标度
v3	数值型(N)	1	0	ID	无	无	无	标度
v4	数值型(N)	4	0	ID	无	无	无	标度
v5	数值型(N)	1	0	ID	无	无	无	标度
v6	数值型(N)	4	0	ID	无	无	无	标度
v7	数值型(N)	8	2	ID	无	无	无	标度
v8	数值型(N)	1	0	ID	无	无	无	标度
v9	数值型(N)	2	0	ID	无	无	无	标度
v10	数值型(N)	3	0	ID	无	无	无	标度
v11	数值型(N)	2	0	ID	无	无	无	标度
v12	数值型(N)	2	0	ID	无	无	无	标度
v13	数值型(N)	1	0	ID	无	无	无	标度
v14	数值型(N)	1	0	ID	无	无	无	标度
v15	数值型(N)	3	0	ID	无	无	无	标度

## 问卷信息

- 一份问卷即一个个案即一名被调查者
- 一个问题通常对应着一个变量
- 问题的答案被称作变量的取值

## 问卷及其编码

- 问卷是抽样调查收集信息的基本工具。
- 问卷是什么？
  1. 是需要被调查对象来回答的一组问题。
  2. 问题需要经过科学严谨的设计、逻辑合理的编排。
  3. 问卷可区分为不同的类型：自填问卷和访问问卷
  4. 问卷具有较为固定的结构。

## 问卷及其编码：问卷结构

○ 通常，一份问卷应包括以下几个部分：

1. 封面信：Who、What、Why、How
2. 指导语：如何作答的说明
3. 问题与答案：开放式 v.s. 封闭式；特征、行为和态度
4. 编码：给每个问题的每一答案赋予一个数字，只有这样将问卷数字化之后，才能用计算机进行统计分析和统计。

## 问卷编码举例

问	卷	题位
问卷编号：0001		1-4
1. 专业： 汉语言文学		1 5-6
2. 省份： 江苏		32 7-8
3. 性别： A男 B女		2 9
4. 城市： A省会城市 B中等城市 C县级 D乡镇 E农村 F其他		5 10
5. 职业： A工人 B农民 C企事业单位人员 D公务员 E个体户 F其他		2 11
6. 独生子女： A是 B否		2 12
7. 家庭月收入： A5000元以上 B3000-5000元 C1000-3000元 D1000元以下		3 13
8. 学费来源： A父母工资及家庭储蓄 B亲朋好友资助 C借债 D自己勤工助学 E无来源		2 14
9. 高中上网条件： A有 B无		2 15
10. 毕业中学： A省级重点中学 B地区(市)级重点中学 C普通中学 D其他		1 16
11. 填报我校信息来源： A电视 B报纸 C网络 D学校资料 E招生现场咨询 F他人推荐		6 17-19
12. 知道或了解我校的途径： A电视 B报纸 C网络 D学校资料 E招生咨询会 F我校师生 G他人推荐		7 20-22
13. 对我校及专业的了解： A非常了解 B基本了解 C知道一些，但不特别清楚 D不了解		3 23
14. 我校知名度影响力： A很好 B较好 C一般 D不清楚，还要进一步了解		2 24
15. 填报我校的决定： A家人 B班主任老师 C自己 D亲友 E权威专家或机构指导 F其他		4 25
16. 报考专业原因： A爱好所致 B父母决定 C老师或亲属的推荐 D此学科有发展潜力 E就业前景		3 26
17. 报考时，重点考虑学校：		
① 学校所属地理位置 ② 学校知名度 ③ 专业 ④ 学校环境 ⑤ 毕业生就业		12 27-32
F教学质量 G教学设施 H学校管理 I学费 J奖学金 K其他		4 33
18. 采取的态度： A非常满意 B满意 C不满意 D无所谓		3 34
19. 就读专业的态度： A非常满意 B满意 C不满意 D无所谓		2 35
20. 愿意介绍我校： A愿意 B不愿意 C无所谓		2 36
21. 读大学的态度： A考研 B直接参加工作 C出国深造 D尚未考虑 E其他		2 37
22. 最好的招生宣传方式：		
A社会新闻媒体宣传 B省招生考试报刊 C向中学寄送招生简章 D现场招生咨询活动 E学校主页 F其他形式		6 37
23. 对学校迎新工作评价： A非常满意 B满意 C一般 D不满意		3 38
24. 我校招生网站内容： A很好 B好 C一般 D不好		2 39
25. 填报志愿时，上学校招生网查询： A仔细阅读 B一般了解 C不看		4 40

## 数据编码方案

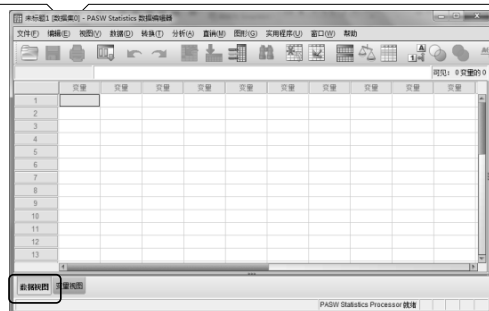
变量名(标题)	变量名	题位
编号	id	1-4
专业	zy	5-6
省份	sf	7-8
性别	sxx	9
城市	cs	10
职业	v1	11
独生子女	v2	12
家庭月收入	v3	13
学费来源	v4	14
上网条件	v5	15
信息来源	v71-v73	17-19
了解途径	v81-v83	20-22
专业了解	v9	23
知名度影响力	v10	24
报考决定	v11	25
报考专业原因	v12	26
报考学校原因	v131-v133	27-32
采取态度	v14	33
专业态度	v15	34
介绍我校	v16	35
志向	v17	36
招生宣传方式	v18	37
迎新工作评价	v19	38
招生网站内容	v20	39
上校网查询	v21	40

## 从问卷到SPSS数据

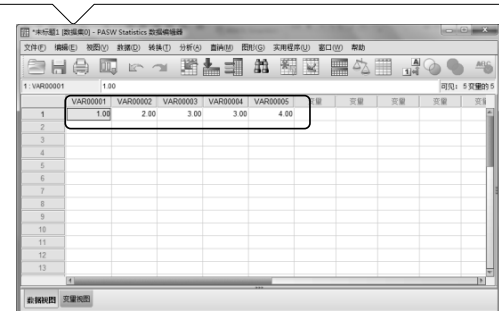
○ 一份问卷即一个个案(case)，即一名被调查者

○ 一个问题通常对应着一个变量(variable)

## 数据录入：打开新的数据窗口



## 实验：在第一行随便输入几个数

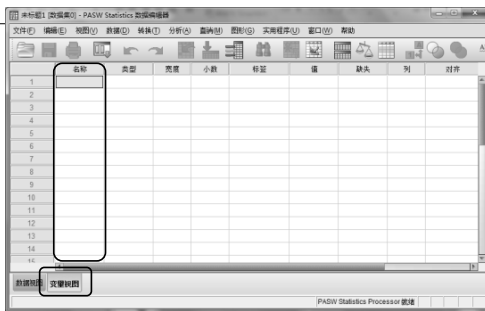


- SPSS对数据的处理是以变量为前提的。
- 所以，数据录入前最好先定义变量属性，包括指定名称、（存储）类型、宽度、小数、标签、值、缺失、列（宽）、对齐、度量标准和角色。
- 方法：单击数据窗口左下方的“变量视图”标签或双击第一列的题头 **变量** 进入变量视图，在此窗口完成变量属性相关内容的设定。

## 数据录入：切换到变量视图



## 数据录入：指定变量“名称”



## 指定变量“名称”注意事项

- SPSS默认变量名称为VAR00001、VAR00002等。但，宽度不会超过8个字符。
- 用户可以根据自己方便来修改或指定变量名称，其中有一些规则或经验。

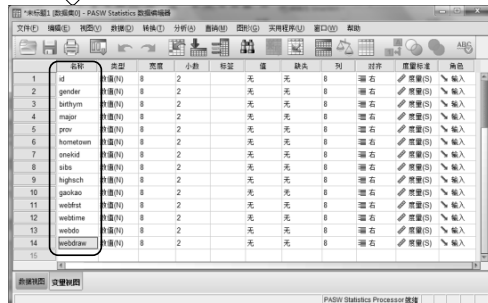
## 指定变量名的规则和经验

- 变量名必须以字母、汉字或字符@开头，其他字符可以是任何字母、数字或\_、@、#、\$等符号。
- 变量名中不能有空白字符或其他特殊字符（如“！”、“？”、“\*”等）。
- 变量名最后一个字符不能是句号，建议使用下划线\_作为最后一个字符。
- 在SPSS中不区分大小写。例如，HXH、hxx或Hxx对SPSS而言，均被视为同一个变量。

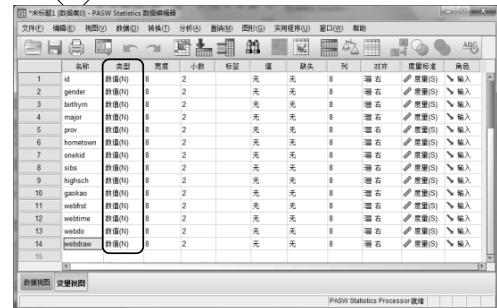
## 指定变量名的规则和经验

- SPSS的保留字不能作为变量的名称，如ALL、AND、BY、EQ、GE、GT、LT、NE、NOT、OR、TO、WITH等。
- 变量命名必须惟一，不能有两个相同的变量名。
- 建议：变量名总长度不超过8个字符（即4个汉字）。
- 建议：不使用中文字符作为变量名。

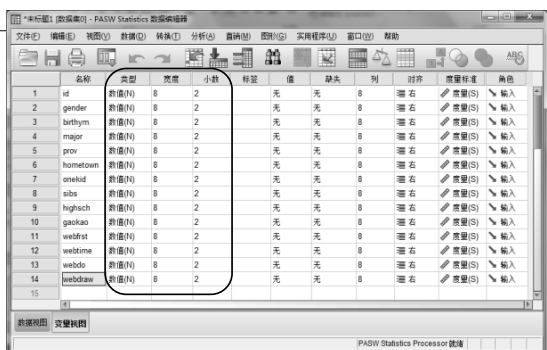
## 数据录入：指定变量“名称”



## 数据录入：指定变量“类型”

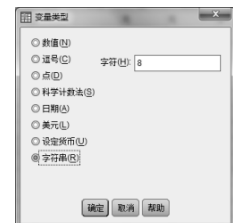


## 变量（取值的存储）类型



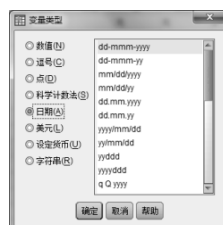
## 变量（取值的存储）类型

- **字符型**或者**字符串**也是SPSS较常用的数据类型。
- 字符串默认的显示宽度为8个字符位（即4个汉字），且区分大小写。
- 字符串不能进行数学运算。
- 字符串在数据处理过程中采用一对双引号引起来；但在数据录入时，不应加上双引号，否则将被作为字符型数据的一部分。



## 变量（取值的存储）类型

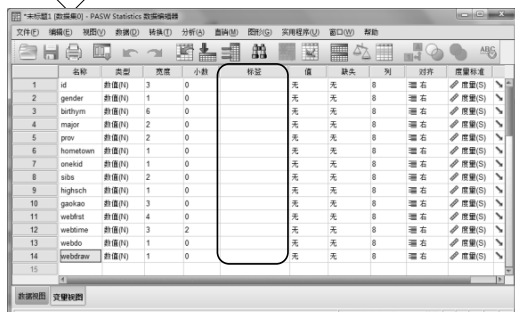
- **日期型**或者**日期**表明对应的取值是表示时间或日期的。
- 日期型数据不允许自由设定宽度和小数位，但可以选择不同的显示格式。
- 日期型数据主要在时间序列分析较为有用，在通常的情况中，可以将其设定为数值型即可。



## 变量（取值的存储）类型



## 数据录入：指定变量“标签”



## 指定变量“标签”



## 数据录入：指定变量“值”标签



## 变量“值”标签

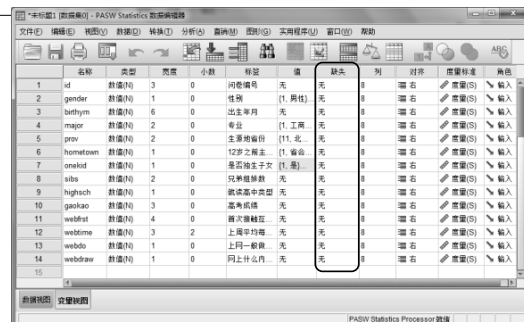
- 变量的值标签是对变量取值含义的进一步解释说明。
- 对于度量标准为“名义”和“序号”的变量（即定类变量和定序变量），变量值标签是必不可少的。
- 变量值标签不但是定类变量和定序变量的数据录入更加方便，且明确了取值含义，也便于结果的解读。



## 数据录入：指定变量“值”标签

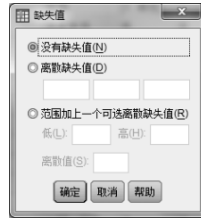


## 数据录入：指定变量“缺失值”



## 指定变量“缺失值”

- SPSS有两类缺失值：系统默认缺失值和用户定义缺失值。
- 对于数值型变量，系统默认缺失值为圆点“.”，而字符型变量值的系统默认缺失值为空字符串。
- 通过缺失值对话框，有两种定义缺失值的方式。
  - 可以定义3个单独的缺失值。
  - 可以定义一个缺失值范围和一个单独的缺失值。



## 数据录入：指定“列”和“对齐”



## 数据录入：指定变量“度量标准”



## 变量“度量标准”

- 对于取值类型相同的变量，如同为数值型，其数值的具体含义不同，所携带的信息量不同，适用的统计方法也就不同。为此，需要增加变量的“度量标准”属性做进一步区分。
- 统计学中，按照对事物描述的精确程度，将度量标准从低到高区分为4种类型：
  - 定类尺度：nominal scale,
  - 定序尺度：ordinal scale,
  - 定距尺度：interval scale,
  - 定比尺度：ratio scale



## 变量的4种“度量标准”

- 定类尺度/名义
  - 对事物的类别或属性的一种测度，按照事物的某种属性对其进行分类或分组。
  - 定类变量的特点是：其取值仅代表事物的类别或属性，仅能测定类别差，不能比较各类之间的大小，所以各类之间没有顺序或等级。
  - 定类变量的取值既可以是数值型的，也可以是字符型；但都只能计算频数、频率、百分比。
  - 定类变量的取值类别必须满足完备性和互斥性原则。

## 变量的4种“度量标准”

- 定序尺度/序号
  - 对事物之间等级或顺序差别的一种测度，可比较优劣或排序。
  - 定序变量比定类变量包含更多的信息，不但可以分类，还可以排序次。
  - 但是，因只测度了类别之间的序次，而无法反映类别之间的准确差值，故定序变量的取值不没有绝对的数量含义。因此，只能分类、排序，仍不能加减。
  - 与定类变量一样，其取值也既可以是数值型的，也可以是字符型的。
  - 定序变量也可以计算频率，但还可以计算累积频率。

## 变量的4种“度量标准”

- 定距尺度
  - 对事物类别、序次之间间距的测度。
  - 定距变量的取值不但能分类、排序，还能进行加减。
  - 但是，因为其0值没有物理含义，即没有绝对的“0”点，故仍不能做乘、除运算。
  - 生活中最典型的定距变量是温度。

## 变量的4种“度量标准”

- 定比尺度
  - 定比尺度是能够测算两个测量值之间比值的一种度量尺度。
  - 其结果和定距变量一样也表现为数值，比如高考成绩。但差别在于，定比尺度变量具有绝对“0”点，即0表示没有。因此，对于定比变量，可以进行乘除运算。
  - 定比尺度是最高层次的度量标准。SPSS中默认的是变量度量标准是定比尺度，不过由于定距和定比尺度变量在绝大多数统计分析中没有本质差别，故SPSS将定距和定比变量合称为“度量”。

## 变量“度量标准”

- 度量标准的本质在于变量取值所蕴含的信息量不同。
- 信息量的多少决定了适用统计方法的不同。
- 一般地，可以将属于更高度量标准的变量降级为更低度量标准的变量来处理，比如将“度量（S）”作为“名义（N）”来处理，不过这样做会损失原来变量所包含的一部分信息。
- 但是，一般**不能**将更低度量标准的变量升级为更高度量标准的变量来处理。

## 指定变量属性中不同情形的处理

- 单选题和开放题的处理
  - 简单直接，没有什么需要特别注意的地方。
- 半开放题的处理
  - 半开放题是指含有“其他（请注明： ）”选项的单选题。
  - 指定变量时，可以使用两个变量，第1个变量中，“其他”作为一个选项；第2个变量将“其他”中“请注明”的内容作为一个单独的开放题，而将没有选择“其他”一项的案例在此变量上的取值作为系统缺失值。

## 指定变量属性中不同情形的处理

- 多选题的处理
  - 多选题可能存在不止一个答案，因此无法直接用一个变量来直接编码。常见的方式有两种：多重二分法，和多重分类法。
  - **多重二分法**：编码时，将每一个选项定义成一个变量，有几个选项就有几个变量，且均以取值等于1表示选了该项、以取值等于0表示未选该项。
  - 这时，需要注意的是，这一组变量最好用后缀加以标识，比如 v02a、v02b、v02c。
  - 多重二分法是多选题的标准处理方式，但如果选项比较多，则会增加录入工作量。

## 指定变量属性中不同情形的处理

- 多选题的处理
  - **多重分类法**：也是利用多个变量来对一个多选题的答案进行编码。应该用多少个变量，取决于实际可能给出的最多答案的数目而定。
  - 这多个变量必须为数值型变量，所有变量采用同一套取值标签。
  - 这时，需要注意的是，这一组变量最好用后缀加以标识，比如 v02a、v02b、v02c。

## 数据录入过程中的编辑与修改

- **单元格值的修改**：由于各种原因，已经输入的数据有时需要修改，这就需要进行编辑，可用方向键或鼠标将黑框移动到要修改的单元，键入新值。
- 增加或删除一个新的个案（即一个新的行）
- 增加或删除一个新的变量（即一个新的列）

## 数据另存



## 小结

- **数据录入**：就会把每个被调查者填答的问卷信息经过编码，以变量及其取值的形式录入到SPSS中的过程
- 包括三个步骤：
  1. 定义变量名、指定变量属性（**定规矩**）
  2. 确定编码规则（**定规矩**）
  3. 录入数据（**要细致**）
- 单选题、多选题、开放题的处理方式。

## 练习：问卷数据录入

- **第一题：分组完成 Working in SPSS 的课后作业**
- **第二题**：根据《商学院学生社会科学素养调查问卷》设计一个SPSS录入模板。要求：
  - 使用简洁易懂的变量名
  - 设定好变量标签
  - 设定好取值标签

## 操作练习&提问时间

## 本周预习内容：

- 预习第3章变量级别的数据管理3.1，3.2，3.3
- 可以**参考** C H A P T E R 1 0 Supplemental SPSS Operations