



南京大學  
NANJING UNIVERSITY

## 第2章 数据录入与数据获取

- 用于分析的数据资料有两种：
  - **原始资料**。如调查问卷中的数据，需要将它们录入进SPSS，建立数据文件。
  - **其他数据格式的资料**。需要将它们导入SPSS，如Excel数据格式。

# 学习目标

- 熟悉SPSS的数据格式
- 能将问卷数据录入成SPSS数据文件
- 能把文本数据文件、Excel数据文件、数据库文件转化成SPSS数据文件
- 能把SPSS数据文件保存为文本数据文件、Excel数据文件和数据库文件

# 主要内容

- 1. 数据格式概述
- 2. 数据的直接录入
- 3. 外部数据的获取
- 4. 数据的保存

# 1. 数据格式概述



	time	id	s0	s2	s3	s4	s5	s7	s9	c0_1	c0_2	c0_3	O1
1	200704	1	100北京	男	20	本科	公司普通职...	未婚	2000-2999元	2 无	2 无	2 无	没有
2	200704	2	100北京	男	24	高中/中专	公司普通职...	未婚	6000-7999元	2 无	2 无	2 无	没有
3	200704	3	200上海	男	20	高中/中专	无业/待业/...	未婚	1000-1499元	2 无	2 无	2 无	没有
4	200704	4	100北京	女	65	大专	企/事业管...	已婚	1000-1499元	2 无	2 无	2 无	没有
5	200704	5	200上海	女	40	高中/中专	公司普通职...	已婚	99拒绝回答	2 无	2 无	2 无	没有
6	200704	6	100北京	男	50	大专	公司普通职...	已婚	4000-4999元	2 无	2 无	2 无	没有
7	200704	7	100北京	女	53	初中/技校...	退休	已婚	5000-5999元	2 无	2 无	2 无	有
8	200704	8	300广州	男	44	高中/中专	无业/待业/...	已婚	99拒绝回答	2 无	2 无	2 无	没有
9	200704	9	200上海	女	35	大专	公司普通职...	已婚	3000-3999元	2 无	2 无	2 无	没有
10	200704	10	200上海	男	21	本科	学生	未婚	5000-5999元	2 无	2 无	2 无	没有
11	200704	11	300广州	男	34	本科	企/事业管...	已婚	6000-7999元	2 无	2 无	2 无	没有
12	200704	12	200上海	女	28	高中/中专	公司普通职...	已婚	5000-5999元	2 无	2 无	2 无	有
13	200704	13	100北京	女	63	本科	退休	已婚	3000-3999元	2 无	2 无	2 无	没有
14	200704	14	100北京	女	46	高中/中专	公司普通职...	已婚	4000-4999元	2 无	2 无	2 无	没有
15	200704	15	100北京	男	28	初中/技校...	公司普通职...	未婚	5000-5999元	2 无	2 无	2 无	没有
16	200704	16	200上海	男	58	硕士或以上	教师	已婚	10000-149...	2 无	2 无	2 无	没有
17	200704	17	100北京	男	35	本科	其他职业	已婚	8000-9999元	2 无	2 无	2 无	有
18	200704	18	100北京	男	22	高中/中专	公司普通职...	未婚	3000-3999元	2 无	2 无	2 无	没有
19	200704	19	100北京	男	30	硕士或以上	企/事业管...	已婚	6000-7999元	2 无	2 无	2 无	没有
20	200704	20	200上海	女	30	高中/中专	国家公务员	已婚	15000-199...	2 无	2 无	2 无	有
21	200704	21	200上海	女	31	高中/中专	公司普通职...	已婚	4000-4999元	2 无	2 无	2 无	没有

数据视图

	名称	类型	宽度	小数	标签	值	缺失	列	对齐	测量	角色
1	time	数值	6	0	月份	无	无	8	右	有序(O)	输入
2	id	数值	4	0	ID	无	无	8	右	有序(O)	输入
3	s0	数值	3	0	S0. 城市	{100, 100北...	无	8	右	名义(N)	输入
4	s2	数值	1	0	S2. 性别	{1, 男}...	无	8	右	名义(N)	输入
5	s3	数值	4	0	S3. 年龄	无	无	8	右	度量	输入
6	s4	数值	1	0	S4. 学历	{1, 初中/技...	无	8	右	有序(O)	输入
7	s5	数值	8	2	S5. 职业	{1.00, 企/事...	无	8	右	名义(N)	输入
8	s7	数值	1	0	S7. 婚姻状况	{1, 已婚}...	无	8	右	名义(N)	输入
9	s9	数值	2	0	S9. 家庭月收入	{1, 999元或...	99	8	右	名义(N)	输入
10	c0_1	数值	2	0	C0. 请问您的家...	{1, 1 有}...	无	8	右	名义(N)	输入
11	c0_2	数值	2	0	C0. 请问您的家...	{1, 1 有}...	无	8	右	名义(N)	输入
12	c0_3	数值	2	0	C0. 请问您的家...	{1, 1 有}...	无	8	右	名义(N)	输入
13	O1	数值	1	0	O1. 是否拥有家...	{1, 有}...	无	8	右	名义(N)	输入
14	a3	数值	1	0	A3. 首先, 请问...	{1, 1 明显好...	无	8	右	有序(O)	输入
15	a3a_1	数值	3	0	A3a. 您为什么...	{0, 中性原因...	无	8	右	名义(N)	输入
16	a3a_2	数值	3	0	A3a. 您为什么...	{0, 中性原因...	无	8	右	名义(N)	输入
17	a4	数值	1	0	A4. 那么与现在...	{1, 1 明显好...	无	8	右	有序(O)	输入
18	a8	数值	1	0	A8. 那么与现在...	{1, 1 非常好...	无	8	右	有序(O)	输入
19	a9	数值	1	0	A9. 那么您认为...	{1, 1 明显改...	无	8	右	有序(O)	输入
20	a10	数值	1	0	A10. 那么与现...	{1, 1 明显繁...	无	8	右	有序(O)	输入
21	a16	数值	1	0	A16. 对于大宗...	{1, 1 很好的...	无	8	右	有序(O)	输入
22	Ts9	数值	8	2	家庭收入2级	{1.00, Belo...	9.00	8	右	名义(N)	输入

# 数据的录入格式

- SPSS数据文件是一种有结构的数据文件。
- 数据格式要求：
  - 不同观察对象的数据不能在一条记录中产生，即同一个案的数据应当独占一行；
  - 每一个测量指标或影响因素只能占据一列的位置，即同一个指标的测量值都应当录入到同一个变量中去。



# 变量属性（1）

- 变量的存储类型（Type）
  - 数值型、字符型、日期型等
- 变量的测量尺度（Measure）
  - 名义尺度：性别、民族等；
  - 有序尺度：满意度等；
  - 定距尺度：温度等，能进行加减运算，但不能进行乘除运算；
  - 定比尺度：薪水、销售额等，能进行加减乘除运算。

## 变量属性（2）

- 变量的名标签（Label）与值标签（Values）：对变量名及变量值进行补充说明，使结果更易于阅读。
  - 变量名s2，它的名标签：性别；它的值标签：1=“男”，2=“女”
- 缺失值（Missing）
  - 系统缺失值
  - 自定义缺失值

# 一个数据录入实例（1）

- 欲比较A公司与B公司普通员工的收入（月薪）有无差距，随机各调查了五人，数据如下：  
A公司：2000，3000，4000，3000，4500  
B公司：3000，3000，3500，4000，5000  
请建立数据文件。

## 一个数据录入实例（2）

- 操作步骤：
  - 定义各变量名；
  - 指定每个变量的各种属性：存储类型、长度、测量尺度等；
  - 录入数据。

# 一个数据录入实例（3）

- 定义变量的基本原则
  - 不同观察对象的数据不能在一条记录中出现；
  - 每一个测量指标/影响因素只能占据一列的位置，即同一个指标的测量数值都应当录入到同一个变量中去；
  - 最终的数据集应当能够包含原始数据的所有信息。

## 一个数据录入实例（4）

- 本例需要建立两个变量，一个变量代表收入的具体数值，取名为X；另一个变量代表观察对象所在的是哪个公司，取名为GROUP；
- 对于调查问卷数据的录入，问题就要简单得多，往往一张调查问卷对应了一个受访者，大多数情况下也就对应了一条记录。

## 一个数据录入实例（5）

- 指定各变量的属性
  - 确定变量**存储类型**：数值型应用最广泛，字符型和日期型尽量少用；
  - 确定**测量尺度**；
  - 变量的**名标签**与**值标签**的补充说明；
  - **缺失值**的处理：除非问卷中有特殊编码，否则按默认情况处理；
  - 其他设置：默认设置。

# 一个数据录入实例（6）

- 输入数据





南京大學  
NANJING UNIVERSITY

## 2. 数据的直接录入

## 中国消费者信心指数研究问卷

S0 受访者所在城市:

单选题

100 北京 200 上海 300 广州

S1 请问您贵姓是? \_\_\_\_\_

开放题

S2 记录被访者性别:

1 男性 2 女性

S3 请问您的年龄是? \_\_\_\_\_

单选题

S4 请问您的学历是?

单选题

1 初中/技校或以下 2 高中/中专 3 大专 4 本科 5 硕士或以上

S5 请问您的职业是?

单选题

1 企/事业管理人员 2 工人/体力工作者(蓝领) 3 公司普通职员(白领)

4 国家公务员 5 个体经营者/私营业主 6 教师

7 学生 8 专业人士(医生、律师等) 9 无/待/失业、家庭主妇

10 退休 11 其他职业

S7 请问您的婚姻状况是?

单选题

1 已婚 2 未婚 3 离异/分居/丧偶

S9 请问您的家庭月收入(包括工资、奖金和各种外快收入)大约在什么范围呢?

- |                      |                      |                      |
|----------------------|----------------------|----------------------|
| 1 999 元或以下           | 2 1 000 ~ 1 499 元    | 3 1 500 ~ 1 999 元    |
| 4 2 000 ~ 2 999 元    | 5 3 000 ~ 3 999 元    | 6 4 000 ~ 4 999 元    |
| 7 5 000 ~ 5 999 元    | 8 6 000 ~ 7 999 元    | 9 8 000 ~ 9 999 元    |
| 10 10 000 ~ 14 999 元 | 11 15 000 ~ 19 999 元 | 12 20 000 ~ 29 999 元 |
| 13 30 000 以上         | 98 无收入               | 99 拒答                |

单选题

C0 请问您的家庭目前有下列还贷支出吗?

多选题—类型1

C0\_1 房贷 1 有 2 无 99 拒答

C0\_2 车贷 1 有 2 无 99 拒答

C0\_3 其他一般消费还贷 1 有 2 无 99 拒答



单选题

C0\_2 车贷 1 有 2 无 99 拒答

C0\_3 其他一般消费还贷 1 有 2 无 99 拒答

单选题

O1 请问您家里有家用轿车吗?

单选题

1 有 2 没有

A3 首先,请问与一年前相比,您的家庭现在的经济状况怎么样呢? 是变好、基本不变还是变差?  
1 明显好转 2 略有好转 3 基本不变 4 略有变差 5 明显变差 9 说不清/拒答

单选题

A3a 为什么您这样说呢?(最多限选两项)——

多选题—类型2

0 中性原因 90 不知道/拒答

10 改善:收入相关 110 恶化:收入相关

20 改善:就业状况相关 120 恶化:就业状况相关

30 改善:投资相关 130 恶化:投资相关

40 改善:家庭开支相关 140 恶化:家庭开支相关

50 改善:政策/宏观经济 150 恶化:政策/宏观经济相关

A4 那么与现在相比,您觉得一年以后您的家庭经济状况将会发生什么变化?  
1 明显好转 2 略有好转 3 基本不变 4 略有变差 5 明显变差 9 说不清/拒答

单选题

A8 那么与现在相比,您认为一年以后本地区的经济发展状况将会如何?  
1 非常好 2 比较好 3 保持现状 4 比较差 5 非常差 9 说不清/拒答

单选题

A9 您认为一年之后本地区的就业状况将会如何变化?  
1 明显改善 2 略有改善 3 保持现状 4 略有变差 5 明显变差 9 说不清/拒答

单选题

A10 那么与现在相比,您认为5年之后,本地区的经济将会出现怎样的变化?  
1 明显繁荣 2 略有改善 3 保持现状 4 略有衰退 5 明显衰退 9 说不清/拒答

单选题

A16 对于大宗耐用消费品的购买,如家用电器、家用计算机以及高档家具之类的,您认为当前是购买的好时机吗?

单选题

1 很好的时机 2 较好时机 3 很难说,看具体情况而定 4 较差时机 5 很差的时机 9 不知道/拒答

## 一、开放题的录入（1）

- 开放题：需要被调查者填写内容的问题。
- 先定义变量名，再定义变量属性。
- 大多数时候，开放题的答案可能是一个句子或者一段话，此时要更改该变量的长度（Width）。

## 一、开放题的录入（2）

- S1 请问您贵姓是? \_\_\_\_\_
  - 变量名：姓
  - 变量属性：类型——字符型；测量尺度——名义尺度（Norminal）

## 二、单选题的录入（1）

- 一般情况下，一个单选题设置一个变量。
- 需定义变量的值标签。
- 采用 数值代码+值标签 的录入方式。
- 如果问卷中含“其他，请指出”选项的单选题，则在录入时可以采取两个变量对其进行定义，在第一个变量中，“其他，请指出”作为选项中的一个可进行选择，第二个变量将“其他，请指出”看成一个独立的开放题，按照开放题的录入方式进行数据录入，将没有选择该选项的被访者作为缺失值处理。



## 二、单选题的录入（2）

- S0 受访者所在城市：

100 北京    200 上海    300 广州

— 变量名：s0

— 变量属性：

- 类型——数值型；
- 名标签——S0. 城市；
- 值标签——100= “100北京” ， 200= “200上海” ， 300= “300广州” ；
- 测量尺度——名义尺度

## 二、单选题的录入（3）

- S2 记录被访者性别：
  - 1 男性 2 女性
- 变量名：性别
- 变量属性：
  - 类型——数值型；
  - 值标签——1=“男”，2=“女”；
  - 测量尺度——名义尺度

### 三、多选题的录入（1）

- 1、多重二分法

- 在编码的时候，对应每一个选项都要定义一个变量，这些变量均为二分类（该变量只取两个值：“选中”与“未选”）。

注：（1）如果选项不是很多可以用这种方法！

（2）后续对多选题进行分析时，必须把这些变量捆绑成一个变量集，再进行分析！

Analyze->Multiple Response->Define Variable Sets

### 三、多选题的录入（2）

- C0 请问您的家庭目前有下列还贷支出吗？

C0\_1 房贷 1 有 2 无 99 拒答

C0\_2 车贷 1 有 2 无 99 拒答

C0\_3 其他一般消费还贷 1 有 2 无 99 拒答

– 变量名： c0\_1 ； c0\_2 ； c0\_3 ； （共3个变量）

– 变量属性（以“c0\_1”为例）：

- 类型——数值型；
- 名标签——C0. 请问您的家庭目前有下列还贷支出吗：  
： 房贷；
- 值标签——1= “有”， 2= “无” ； 99= “拒答”
- 测量尺度——名义尺度。

C0 请问您的家庭目前有下列还贷支出吗？

A. 房贷 B. 车贷 C. 其他一般消费还贷

### 三、多选题的录入（3）

- 2、多重分类法

- 与多重二分法相似，也是利用多个变量对选项进行定义，但是变量的个数由被访者实际可能给出的最多答案数而定，并不是每个选项对应一个变量。
- 若选项中出现“其他，请指出”选项，处理方法与单选题类似。

注：如果题干后附：最多选3（4、5、…）项，  
可以用这种方法！

### 三、多选题的录入（4）

- A3a 为什么您这样说呢？（**最限定选两项**）
  - 变量名：A3a\_1； A3a\_2； 共2个变量。
  - 变量属性（“A3a\_1”）：
    - 类型——数值型；
    - 名标签——您为什么这样说呢？；
    - 值标签——0=“中性原因”， 10=“改善：收入相关”；  
20=“改善：就业状况相关”； 30=“改善：投资相关”；  
.....
    - 测量尺度——名义尺度。

A3a 为什么您这样说呢？（最限定选两项）——

0 中性原因 90 不知道/拒答

10 改善:收入相关 110 恶化:收入相关

20 改善:就业状况相关 120 恶化:就业状况相关

30 改善:投资相关 130 恶化:投资相关

40 改善:家庭开支相关 140 恶化:家庭开支相关

50 改善:政策/宏观经济 150 恶化:政策/宏观经济相关



南京大學  
NANJING UNIVERSITY

### 3. 外部数据获取

- 三种方式

- 直接打开Excel数据文件

- 利用文本向导读取文本数据

- 利用ODBC接口读取数据库文件



## 4. 数据的保存

- 可以存为SPSS的数据格式（.sav）
- 也可以存为其他类型的格式
  - 文本格式、Excel格式，dbf格式等

# 练习

- 打开CCSS\_Sample.sav文件，观察其变量的各种属性设置。
- 尝试打开主文件名为demo的几个非sav格式数据文件，并将其导入SPSS。将CCSS\_Sample.sav另存为txt、xls和mdb格式的数据文件。

**THE END**