

## SPSS基础与应用

### 读取数据

李丁 讲师

中国人民大学社会与人口学院

## 课程大纲

- SPSS可读取的数据格式
- Get data读取Excel格式(.xls)数据
- Get data读取文本格式(.txt)数据
- Get data读取数据库数据
- 复制数据属性
- Date List

## SPSS数据资料的获取

### ○ 数据的产生方式:

1. 开启已经事先建好的SPSS格式数据文件;
2. 录入自己所搜集的资料建立新数据文件;
3. 读取各种不同类型非SPSS格式数据文件。

### ○ SPSS的资料结构

	变数(variables)
个案(cases)	观察值(observed values)

## SPSS可直接读取的数据格式

- 启动SPSS, 选择“文件”→“打开”→“数据”

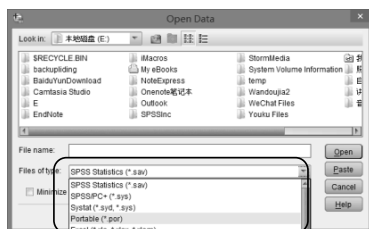
选择文件保存在哪里



默认为.sav文件

## SPSS可直接读取的数据格式

### ○ 设定数据文件类型



## SPSS可直接读取的数据格式

文件标识	文件类型
SPSS Statistics (*.sav)	SPSS各版本的数据文件
SPSS/PC+ (*.sys)	SPSS/PC+ 版本的数据文件
Systat (*.syd, *.sys)	Systat数据文件
便携 (*.por)	SPSS便携格式的数据文件
Excel (*.xls, *.xlsx, *.xlsm)	Excel各版本的数据文件
Lotus (*.W*)	Lotus各版本的数据文件
SYLK (*.slk)	以SYLK (符号链接) 格式保存的数据文件
dBASE (*.dbf)	dBASE系列数据文件 (从dBASE II - IV)
SAS (*.sas7bdat, *.sd7, *.sd2, *.ssd01, *.ssd04, *.xpt)	SAS各版本的数据文件
Stata (*.dta)	Stata各版本的数据文件
文本格式 (*.txt, *.dat)	纯文本格式的数据文件

## 读取Excel格式数据文件

- 以SPSS自带的文件demo.xls为例
  - 该数据文件位于SPSS安装目录下的Samples文件夹内
  - demo.xls的数据结构
- 注意：
- 包含几个数据表，应打开哪个数据表
  - 读取全部数据还是部分数据（比如：A2：F5）
  - 数据表的第一行是否是变量名

## 读取Excel格式数据文件

- 第1步：启动SPSS，文件→打开→数据，选择路径



## 读取Excel格式数据文件

- 第2步：选择文件类型“Excel(\*.xls, ...)”



## 读取Excel格式数据文件

- 第3步：选择目标数据文件，点击“打开”



建议点击“粘贴”，把所有的操作以语法文件的形式保存下来备查！

## 读取Excel格式数据文件

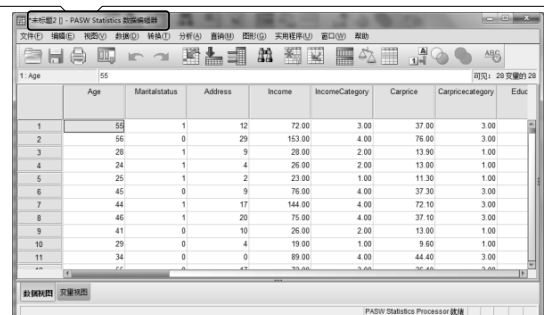
- 第4步：设定首行是否为变量名称、“工作表”和“范围”



数据被读入SPSS

## 读入Excel格式数据文件

赶紧：文件→另存为，将读取的Excel数据保存成SPSS格式数据



## 读取demo.xls的语法程序

```
GET DATA
  /TYPE=XLS
  /FILE='C:\Program Files\SPSSInc\PASWStatistics18\Samples\English\demo.xls'
  /SHEET=name 'demo'
  /CELLRANGE=full
  /READNAMES=on
  /ASSUMEDSTRWIDTH=32767.
```

**Get DATA**                      数据读取命令  
**/TYPE=XLS**                      指定文件类型  
**/FILE=' '**                      指定文件所在路径  
**/SHEET= name 'demo'**              指定要读取的表单  
**/CELLRANGE=full**              指定表单中要读区的数据范围  
**/READNAMES=on**              设定Excel文件第一行是否为变量名  
**/ASSUMEDSTRWIDTH=32767**      采用默认的字符串列的最大宽度

## 操作练习&提问时间

- 打开example1.xls数据，观察其基本结构
- 用SPSS读取example1.xls数据

## 读取文本格式(.txt)数据文件

- 以SPSS自带的文件demo.txt为例
  - 该数据文件位于SPSS安装目录下的Samples文件夹内
  - demo.txt的数据结构
- 注意：
- 第一行是否为变量名
  - 变量之间是固定宽度，还是以某符号（如，逗号）间隔

## 读取文本格式(.txt)数据文件

- 第1步：启动SPSS，文件→打开文本数据，选择路径



## 读取文本格式(.txt)数据文件

- 第2步：选择目标数据文件，点击“打开”

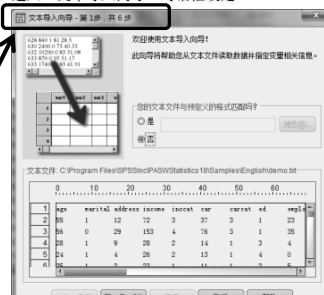


建议点击“粘贴”，把所有的操作以语法文件的形式保存下来备查！

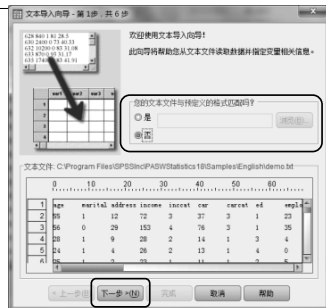
## 读取文本格式(.txt)数据文件

- 第3步：进入“文本导入向导”对话框设定

这一步包含六个小步骤

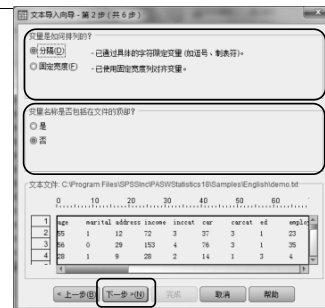


## 第3.1步：设定有无预定格式



如果有，选择“是”；  
如果没有，则保持SPSS  
的默认选择“否”。  
然后点击“下一步”

## 第3.2步：设定变量排列方式和名称



设定变量  
排列方式

文件顶部  
含变量名

## 第3.3步：设定开始行、个案数量等



通常采用  
默认设定

除非打算只读取  
部分案例，否则  
就采用默认设定

## 第3.4步：设定变量分隔符等



根据数据  
相应情况  
进行选择

若数据中的  
字符串变量  
使用限定符  
进行了分隔，  
在此指定

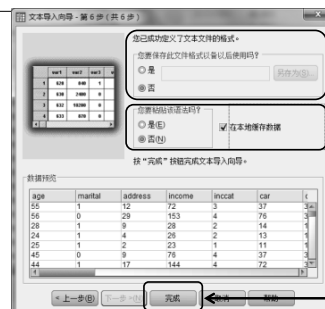
## 第3.5步：设定变量名及其类型



在此更改  
变量名称

在此设定  
变量类型

## 第3.6步：是否需要保留读取操作



若保留所定义  
的数据格式，  
在此进行保存

若保留本次数据读  
取操作，选择“是”  
粘贴相关语法

单击，完成读入

## 读入文本格式数据文件

赶紧：文件→另存为  
将读取的文本数据  
保存成SPSS格式数据

	age	marital	address	income	inccat	car	carcat	ed	employ	retire
1	55	1	12	72	3	37	3	1	23	
2	56	0	29	153	4	76	3	1	35	
3	28	1	9	26	2	14	1	3	4	
4	24	1	4	26	2	13	1	4	0	
5	25	1	2	23	1	11	1	2	5	
6	45	0	9	76	4	37	3	3	13	
7	44	1	17	144	4	72	3	2	23	
8	46	1	20	75	4	37	3	1	29	
9	41	0	10	26	2	13	1	1	8	
10	29	0	4	19	1	10	1	2	10	

## 读取demo.txt的语法程序

GET DATA  
/TYPE=TEXT  
/FILE=C:\Program Files\SPSS\Inc\PASW\Statistics18\Samples\English\demo.txt  
/DELIMITERS=" "  
/ARRANGEMENT=DELIMITED  
/FIRSTCASE=1  
/IMPORTCASE=ALL  
/VARIABLES= age F2.0 marital F1.0 address F2.0 income F3.0 inccat F1.0 car F2.0 carcat F1.0 ed F1.0 employ F2.0 retire F1.0 empcat F1.0 gender A1 reside F1.0 wireless F1.0 multiline F1.0 voice F1.0 pager F1.0 internet F1.0 callid F1.0 callwait F1.0 owntv F1.0 ownvcr F1.0 owncd F1.0 ownpda F1.0 ownpc F1.0 ownfax F1.0 news F1.0 response F1.0  
EXECUTE.

数据格式  
分隔符  
第一个案例开始的地方  
读入所有案例  
变量及格式列表

## GET DATA读文本数据：固定宽度

GET DATA  
/TYPE=TEXT  
/FILE="E:\EXAMPLE2.TXT"  
/RXCASE=1  
/ARRANGEMENT=FIXED  
/FIRSTCASE=1  
/IMPORTCASE=ALL  
/VARIABLES=  
/1) GENDER 4-4 F1.0  
AGE 5-5 F1.0  
EDU 6-7 F2.0  
CLASS 8-8 A1  
SCORE 9-11 F3.2.  
EXECUTE.

数据格式  
资料位置  
一行变量名  
在GET DATA 指令中，  
第一个字符位置0

27

SPSS 基础与应用(李丁)

## 操作练习&提问时间

- 打开记事本demo.txt数据，观察其基本结构
- 用SPSS读取demo.txt数据

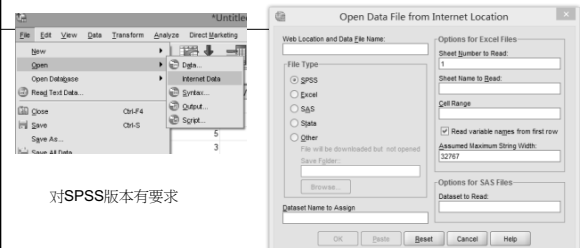
## 读取数据库中的数据

SPSS应用与基础(李丁)

2017/10/9

29

## 读取网络位置上的文件



对SPSS版本有要求

SPSS应用与基础(李丁)

2017/10/9

30

## 链接查询数据库



## 链接查询数据库

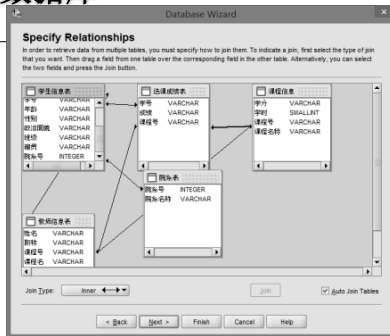


## 链接查询数据库

设置字段关系

连接：表的键  
连接有单向/双向

操作方式：  
•删除：选中连接线  
变粗后delete  
•连接：选中两变量  
+Join type +按键  
Join  
•连接：拖动一个变量  
放到另一个变量上

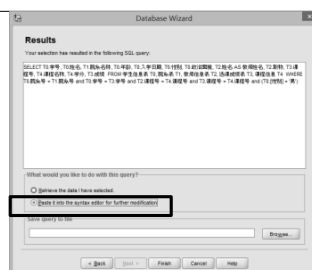


## 链接查询数据库

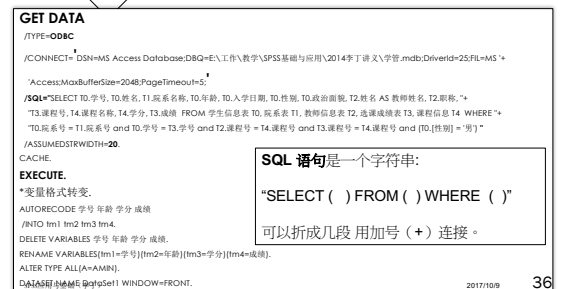


## 链接查询数据库

最后一步生成命令



## 链接查询数据库

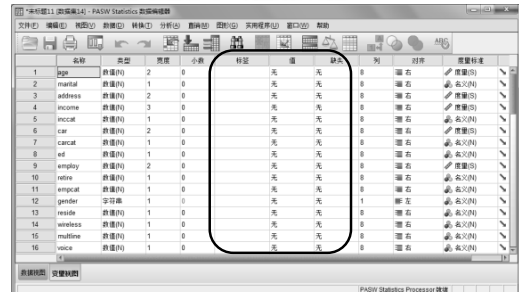


## 操作练习&提问时间

- 1、查询Demo.mdb中30岁以下男性案例的收入与教育水平。
- 2、用data list 命令输入自己与邻座的如下三个信息年龄、性别、爱好：  
age(f2.0) gender(f1.0) like(A6)

## 复制变量属性——有现成格式

Excel文本格式数据读入后，变量属性不完整



## 复制变量属性

- 将已有SPSS数据属性复制到读入数据中来
- 在数据窗口，数据→复制数据属性，启动“复制数据属性”对话框
- 经常用于如下情况：
  - 基于非最终文本数据建立了SPSS文件而非属性定义命令。
  - 此后发生了数据更新。



## 复制变量属性：5步走(1)



第1步：选择属性源动

点击“浏览”选择文件

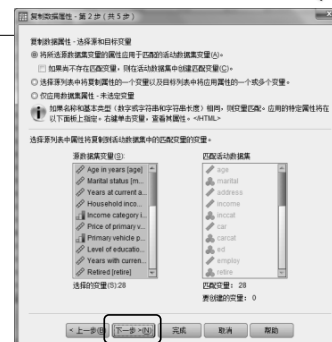
## 复制变量属性：5步走(1)



第1步  
选择属性源动

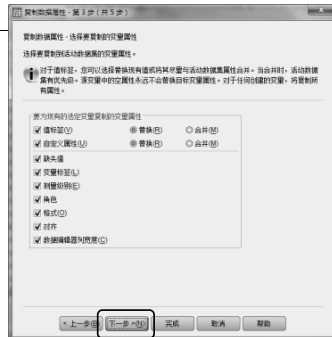
点击“浏览”选择文件

## 复制变量属性：5步走(2)



第2步  
选择源和目标变量

## 复制变量属性：5步走(3)



第3步  
选择要复制的变量属性

## 复制变量属性：5步走(4)



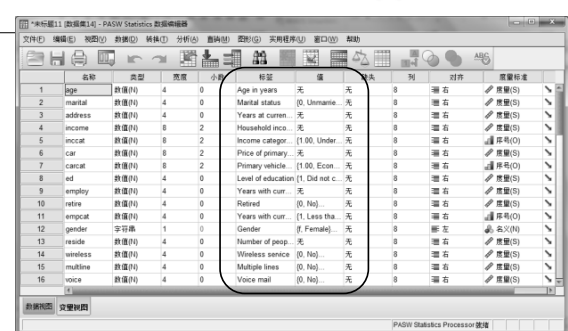
第5步  
是否保留相关操作的语法

## 复制变量属性：5步走(4)



第5步  
选择要复制的数据集属性

## 复制变量属性：5步走(5)



## 操作练习&amp;提问时间

## DATA LIST：更灵活的文本读入

\*示例(1) Fortran 格式。

DATA LIST FILE='d:\example2.txt'

FIXED RECORDS=1 tables(1)

idno (F4.0)

gender (F1.0)

age (F2)

educ (F1)

class (A1)

score (X,F3.1)

10011231A 893  
10020242A 234  
10030531A 783  
10040341A 765  
10051263C 679  
10061433C 798  
10071284C 890  
10080434B 694  
10090395B 798  
10100341B 901  
10111271B 824  
10121222B 836  
10130612B 847

第一行记录

列出所读变量摘要

每行一个案例

变量名和格式

48

SPSS 基础与应用 AUC 李丁



## DATA LIST——更灵活的文本读入

\*示例 (2) 设定栏位+格式

DATA LIST FILE='E:\工作\教学\SPSS基础与应用\2014李丁讲义\EXAMPLE2.txt'  
FIXED RECORDS=1 TABLES /1

idno 1-4  
age 6-7 可更改变量读入顺序  
gender 5  
educ 8  
class 9 (A) 第9列为字符  
score 11-13 (F,1).  
EXE.

第10列跳过，设定了一位小数。改了格式。

1001 231A 893  
10020242A 234  
10030531A 783  
10040341A 765  
10051263C 679  
10061433C 798  
10071284C 890  
10080434B 694  
10090395B 798  
10100341B 901  
10111271B 824  
10121222B 836  
10130612B 847

## DATA LIST——更灵活的文本读入

\*示例(4) 一个案例多行.

DATA LIST FILE='E:\工作\教学\SPSS基础与应用\2014李丁讲义\EXAMPLE2.txt' FIXED

RECORDS=2 TABLES /1

idno (F4.0)  
gender (F1.0)  
age (F2)  
/2  
educ (7X,F1)  
class (A1)  
score (1X, F3.1).  
EXE.

10011231A 893  
10020242A 234  
10030531A 783  
10040341A 765  
10051263C 679  
10061433C 798  
10071284C 890  
10080434B 694  
10090395B 798  
10100341B 901  
10111271B 824  
10121222B 836  
10130612B 847

SPSS基础与应用RUC 李丁

50

## DATA LIST 命令

FORTTRAN 式的格式定义方式：

Fw.d

F表示数字，w总位数，d小数位数

例如 ,2043为 (F4.0), 2043.43为(F6.2)

51

SPSS基础与应用RUC 李丁

## DATA LIST 命令

FORTTRAN的文本格式表示方式：

Aw

A代表文字格式，w为字母位数，汉字数\*2

例如 JOHN 为 (A4), B320012为(A7)

空格表示方式：nX

n表示空了n个空格

52

SPSS基础与应用RUC 李丁

## 用DATA LIST输入简短数据

DATA LIST =DATA LIST FIXED，即默认FIXED

/ID 1-3 SEX 5 (A) AGE 7-8  
OPINION1 TO OPINIONS 10-14.

BEGIN DATA

001 m 28 12212  
002 f 29 21212  
003 f 45 32145

...

128 m 17 11194

END DATA.

SPSS基础与应用 (李丁)

2017/10/9

53

## 用DATA LIST直接输入数据

前面都是FIXED，还可以是list、FREE

Data list list (" ")

/name(a16) total teacher

research(F5.3) student reputation

name2(a14).

Begin data

中国人民大学 96.5 87.93 97.02 100 98.6 中国人民大学  
北京大学 94.8 100 100 78.2 100 北京大学  
清华大学 83.77 41 93.01 42.78 93.4 清华大学  
南开大学 81.3 85.19 83.68 64.43 91.49 南开大学  
南京大学 79.1 77.52 73.18 76.01 90.86 南京大学  
复旦大学 74.2 77.33 66.78 75.85 80.32 复旦大学  
吉林大学 73.1 72.72 73.86 65.59 79.43 吉林大学  
End data.

一行一个案例，回车表示分隔  
分隔符：“空格”“逗号”，  
TAB等

文本，过短则  
删截字符

\*号表示接下来的变量不是默认格式，忽略将提示错误

54

SPSS基础与应用RUC 李丁

## 用DATA LIST直接输入数据

**DATA LIST** FREE / POSTPOS NWINS.

**BEGIN DATA**

2, 19, 7, 5, 10, 25, 5, 17, 8, 11, 3,, 6, 8, 1, 29

**END DATA.**

数据在一行

两个变量不断读下去

SPSS应用与基础（李丁）

2017/10/9

55

## 前面第一次数据录入的新做法

```
data list list
/ enrolled(F1.0) units(F2.0) gpa(F3.2).
begin data.
1 12 2.80
1 12 3.71
2 0 2.65
1 6 2.90
1 16 3.64
end data.
```

```
* Define Variable Properties.
VARIABLE LABELS enrolled 'Are you
currently enrolled in school'.
VALUE LABELS enrolled
1 'yes'
2 'no'.
VARIABLE LABELS units 'How many
units are you taking'.
MISSING VALUES gpa(999).
EXECUTE.
```

SPSS应用与基础（李丁）

2017/10/9

56

## 作业情况检查与讲解

- 数据是否读入了？
- 数据标签和属性是否设定并添加了？
- 发送到 liding@ruc.edu.cn

SPSS应用与基础（李丁）

2017/10/9

57

## 作业情况检查与讲解

- 数据读入过程中的问题
  - 乱码: 用记事本打开看看是否为utf-8(批转程序)
  - 死机: 变量名不合理(自定义变量名)
  - 如果导出的csv文件全部为文本怎么办?
    - 变量不多且数据不会再变: 手工替换.
    - 变量很多或数据可能变化: 命令转换.
  - 如何使用命令修改变量属性(示范)

SPSS应用与基础（李丁）

2017/10/9

58