

普通高等教育“十二五”规划教材

丛书主编 张景中院士

博雅

21世纪教育技术学精品教材

教育研究中定量数据的统计与分析

基于SPSS的应用案例解析

张屹 周平红 编著

清华大学出版社
TSINGHUA UNIVERSITY PRESS



丛书主编 张景中院士
执行主编 王继新



教育研究中定量数据的统计与分析

基于SPSS的应用案例解析

张 屹 周平红 编著
范福兰 周会萍 王攀花 马静思 参编



北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

教育研究中定量数据的统计与分析:基于 SPSS 的应用案例解析/张屹,周平红编著. —北京:北京大学出版社,2015.6

(21 世纪教育技术学精品教材)

ISBN 978 - 7 - 301 - 25777 - 7

I. ①教… II. ①张… ②周… III. ①教育统计—统计分析—软件包—高等学校—教材
IV. ①G40 - 051

中国版本图书馆 CIP 数据核字(2015)第 089599 号

书 名 教育研究中定量数据的统计与分析——基于 SPSS 的应用案例解析
著作责任者 张 屹 周平红 编著
丛书主持 唐知涵
责任编辑 李奕奕 唐知涵
标准书号 ISBN 978 - 7 - 301 - 25777 - 7
出版发行 北京大学出版社
地 址 北京市海淀区成府路 205 号 100871
网 址 <http://www.pup.cn>
电子信箱 zyl@pup.cn
新浪微博 @北京大学出版社
电 话 邮购部 62752015 发行部 62750672 编辑部 62767346
印 刷 者 北京宏伟双华印刷有限公司
经 销 者 新华书店
787 毫米 × 1092 毫米 16 开本 19.5 印张 440 千字
2015 年 6 月第 1 版 2015 年 6 月第 1 次印刷
定 价 48.00 元(含光盘)

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究

举报电话:010-62752024 电子信箱:fd@pup.pku.edu.cn

图书如有印装质量问题,请与出版部联系,电话:010-62756370

前 言

随着大数据时代的到来,教育研究中定量数据的分析显得尤为重要。在教育技术学领域中,定量数据分析与质的研究是教育技术研究两个主要范畴。随着信息科学的发展,定量数据分析逐步受到研究者的重视。从大量的研究论文中不难发现,在教育科学研究中学者们开始采用量化方法,注重对研究对象进行定量的观察与测量,注重用数学工具处理所得的资料和工具。

本书结合本人编写的《教育技术学研究方法》教材,在此基础上,面向教育研究者及一线教师在开展教育、教学研究时对定量数据分析的需求,精选教育研究案例,由浅入深地阐述不同情境中定量数据的统计分析方法。为此,本教材既可作为教育技术学本科生教材,也可作为专业研究生的选用教材。由于书中涉及大量丰富的教育和教育信息化领域的数据分析案例,因此,本书也可作为教育专业的学生、一线教师在进行数据统计分析时的参考教材。

本人作为华中师范大学教育信息技术学院的一名教师,长期从事教育技术学专业本科生和研究生的“教育技术学研究方法”“学与教的理论与方法”“信息技术环境下的教育研究理论与实践”等课程的教学。近10年来,我们的研究团队承担了20多项国家级、省部级纵向科研项目和20多项横向的科研项目,产生了一些有一定应用价值的研究成果,这些鲜活的、富有生命力的研究经验与研究成果,成为该教材编写极其宝贵的研究案例。

本书以我们研究团队的真实研究项目和研究课题为基础,遵循科学研究的规范研究过程与流程、数据统计分析的基本原理及操作流程,形成了如图1所示的内容知识导图。

全书基于SPSS软件的应用案例解析分为三大部分:第一部分,研究数据文件的建立及初级统计分析方法,介绍SPSS数据文件的建立及初级统计分析方法;第二部分,研究数据的高级统计分析方法,介绍了参数检验、非参数检验、方差分析、相关分析、因子分析、聚类分析、回归分析、结构方程模型等高级统计方法;第三部分,介绍教育测量与评价的质量特性分析方法,包括信度、效度、难度和区分度等质量特性指标。

本教材在章节体例的设计上,充分运用教学系统设计思想,力图有所创新。每一章按照如图2的体例编排,涵盖学习目标、关键术语、知识导图、内容简介、方法解读、课程学习、知识拓展、活动任务、参考文献等元素,在每一节的课程学习模块,包含问题导入、数据统计要求、理论讲解、案例解析等元素。

由于定量数据统计分析的方法比较多,且每种方法的适用条件也不同,为便于大家的学习,我们特别在课程学习模块设置了数据统计要求模块,在理论讲解和案例解析模块设置了丰富多彩的教学形式和学习方法,如理论讲解模块设置了核心概念、分析方法辨析、基本操作、知识卡片,案例解析模块设置了案例呈现、操作步骤、结果分

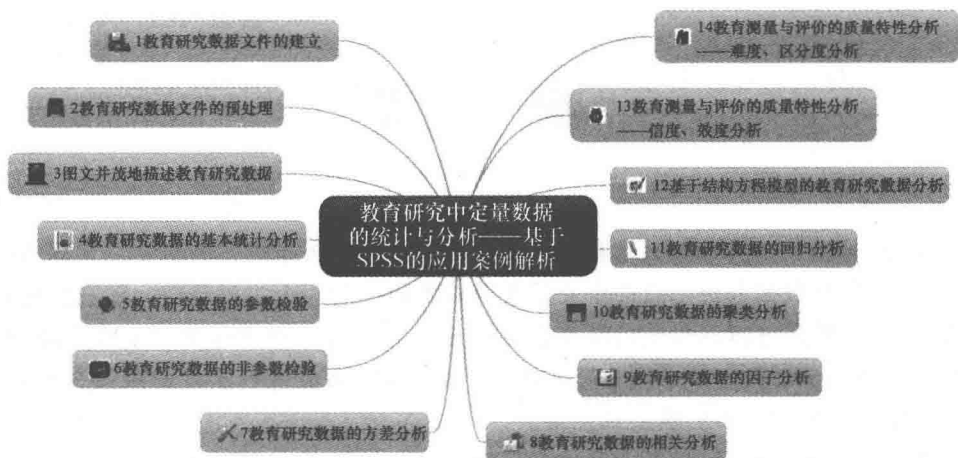


图1 内容知识导图

析、小贴士等,读者可以按照个人的需求进行选择。

本书的特色之处在于:

1. 内容框架的完整性与实用性

本书内容体系的设计以满足教育研究者和一线教师们的需求为目标,全面考虑教育研究者们在进行研究时会用到的各种定量统计分析方法,设计了一个由低级到高级数据统计分析的框架体系,由浅入深地介绍如何对教育研究中的数据进行分析,并通过遴选具有典型代表性的案例来展现数据统计分析方法的应用,具有较强的实用性。

2. 研究案例的丰富性和可操作性

本书每一个知识点的讲授都会以案例的形式呈现,共提供了将近50个案例,内容非常丰富。此外,本书还非常注重内容的可操作性,通篇都使用来源于真实教育情境中的丰富研究案例来讲述教育研究中定量数据的统计与分析。每一章都以问题的提出进行情境导入,以便读者能够在较快的时间内掌握基本的数据统计分析技巧和技能,并能将这些技能用于解决实际研究中的问题。

3. 知识讲解的生动性与翔实性

为了兼顾不同起点的读者,本书注重对每种统计方法的基本原理、适用条件以及统计分析结果的讲解,并对统计分析方法进行辨析,既有初级统计方法的介绍,又有高级统计方法的深入讲解。本书没有太多深奥繁琐的数学证明,但会有一些公式推导、操作流程来揭示统计指标之间的内在联系,以供读者理解每种方法作参考。

我们的研究团队在开展研究活动中,始终得到华中师范大学杨宗凯校长的大力支持,得到我校科技部曹青林部长的关心与支持,得到我院赵呈领教授、院长杨浩教授、常务副院长刘清堂教授、副院长杨九民教授和左明章教授、郭春娥书记的关心与帮助,在此我代表研究团队对各位的关心与支持表示深深的谢意!

在本书的编撰过程中,尤其感谢我们研究团队所有成员的辛劳付出,正是有大家毫不保留的付出,才有今天的书稿面世,正是有大家的精诚合作,才有今天的智慧结

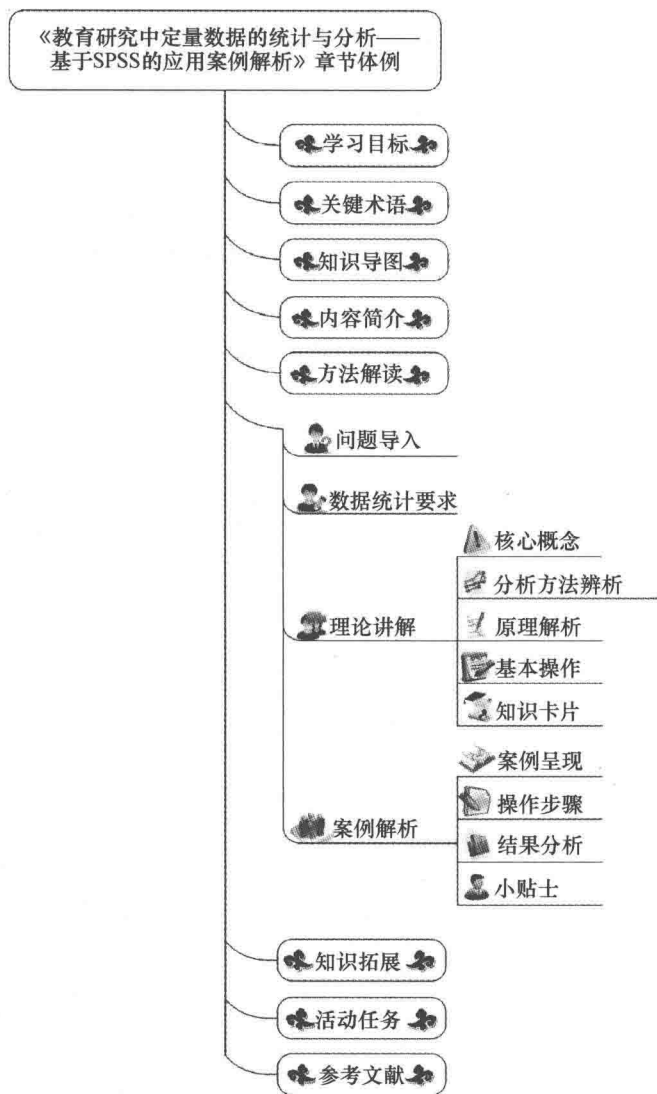


图2 章节的体例安排

晶,正是有大家的齐心协力,才有今天为之骄傲的成果。

本书从策划到出版,始终得到北京大学出版社的理解与支持,从当初我们与出版社的一拍即合,到如今书稿的出版,我们深感北京大学出版社的高效率工作。特别是唐知涵编辑,在本书的出版过程中,虽然她正孕育着一个小生命,但丝毫不减她对本书编辑工作的满腔热情,正是有她辛勤而卓有成效的工作,本书才得以迅速与读者见面。

科学研究永无止境,我们的研究团队在教育研究过程中,不断提升我们的教育研究水平,充分享受着科学研究带给我们的无穷魅力和乐趣!

张 屹 周平红

2014年6月于桂子山理科楼

目 录

第 1 章 教育研究数据文件的建立	001
1.1 引言	002
1.2 SPSS 数据的结构与定义	002
1.3 教育研究中单项选择题的录入	008
1.4 教育研究中多项选择题的录入	013
1.5 教育研究中非选择题的录入	024
第 2 章 教育研究数据文件的预处理	028
2.1 引言	029
2.2 插补数据文件中的缺失数据——数据的缺失值处理	029
2.3 按照一定的规则选择所需数据——数据的选取	034
2.4 重新排列变量数据的顺序——数据的排序	038
2.5 重新划分数据的组别——数据的分组	041
第 3 章 图文并茂地描述教育研究数据	046
3.1 引言	047
3.2 绘制变量的分布比例——饼图	047
3.3 绘制多变量的分布情况——条形图	051
3.4 绘制变量的正态分布情况——直方图	055
3.5 绘制变量的发展趋势——线图	058
3.6 绘制变量间的相关情况——散点图	061
第 4 章 教育研究数据的基本统计分析	067
4.1 引言	068
4.2 测算单变量的数据分布比例——频数分析	068
4.3 分析多变量的联合分布情况——交叉分组下的频数分析	071
4.4 测算多变量的数据分布比例——多选项分析	076
4.5 测算变量数据的中心发展趋势——集中趋势分析	081
4.6 测算变量数据的分散程度——离散趋势分析	085
4.7 测算变量数据的分布情况——分布形态	088
第 5 章 教育研究数据的参数检验	093
5.1 引言	093
5.2 推断样本来自的总体的均值是否与指定的检验值之间存在显著差异 ——单样本 T 检验	094

5.3	推断两个独立样本总体的均值是否存在显著差异——两独立样本 T 检验	099
5.4	推断两个配对样本总体的均值是否存在显著差异——两配对样本 T 检验	104
第 6 章	教育研究数据的非参数检验	109
6.1	引言	110
6.2	推断样本来自的总体的分布是否与已知的理论分布相吻合 ——单样本非参数检验	110
6.3	推断两独立样本来自的两个总体的分布是否存在显著差异 ——两独立样本非参数检验	115
6.4	推断两配对样本来自的两个总体的分布是否存在显著差异 ——两配对样本非参数检验	120
第 7 章	教育研究数据的方差分析	125
7.1	引言	126
7.2	单个因素对观测变量的影响——单因素方差分析	126
7.3	多个因素对观测变量的影响——多因素方差分析	132
7.4	调节协变量对因变量的影响效应——协方差分析	137
第 8 章	教育研究数据的相关分析	143
8.1	引言	143
8.2	测量事物间线性相关程度强弱——线性相关分析	144
8.3	控制其他变量的影响下分析两变量间的线性相关性——偏相关分析	153
第 9 章	教育研究数据的因子分析	158
9.1	引言	159
9.2	实测变量转化为少数几个不相关的综合指标——因子分析	159
第 10 章	教育研究数据的聚类分析	182
10.1	引言	183
10.2	按照一定层次对样本数据进行分类——层次聚类分析	186
10.3	通过反复迭代对样本数据进行分析——K-Means 聚类分析	195
第 11 章	教育研究数据的回归分析	202
11.1	引言	203
11.2	分析单个解释变量对被解释变量的影响——一元线性回归分析	203
11.3	分析多个解释变量对被解释变量的影响——多元线性回归分析	211
第 12 章	基于结构方程模型的教育研究数据分析	224
12.1	引言	225
12.2	验证某一因子模型是否与数据吻合——验证性因子分析	226
12.3	研究多个变量之间多层因果关系及其相关强度——路径分析	250
第 13 章	教育测量与评价的质量特性分析——信度和效度分析	263
13.1	引言	264

13.2	测度综合评价体系的稳定性和可靠性——信度分析	264
13.3	测量工具能测出其所要测量特质的程度——效度分析	279
第 14 章	教育测量与评价的质量特性分析——难度、区分度分析	290
14.1	引言	291
14.2	测算测验项目的难度系数——难度分析	291
14.3	测算测验题目的区分度——区分度分析	296

第 1 章

教育研究数据文件的建立

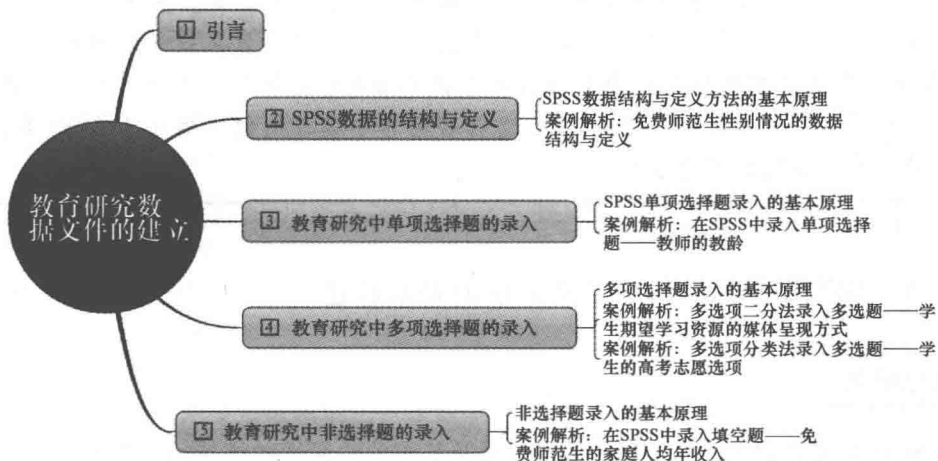
学习目标

1. 阐述教育研究数据文件建立的基本思路。
2. 掌握在 SPSS 中教育数据单项选择题、多项选择题和非选择题录入的方法。
3. 能够结合正在做的调查研究,熟练地应用 SPSS 软件进行教育研究数据文件的建立,并对已回收的问卷数据进行录入。
4. 归纳总结教育研究数据文件建立的基本知识。
5. 总结在进行教育研究数据文件的建立时遇到的困难和采取的解决措施。

关键术语

数据文件 SPSS 数据的结构 数据类型 变量名 变量名标签 变量值标签
单项选择题 多项选择题 非选择题

知识导图



1.1 引言

内容简介

通过对调查问卷的发放和回收,我们已经收集到了大量的第一手数据资料,例如免费师范生远程可视化学习平台需求的调查问卷、中国高校信息化应用质量与效果评价调查问卷(华中师范大学学生问卷)、湖北省农远工程学员能力现状和培训学员培训需求调查问卷等教育数据,那么怎样将回收到的问卷转化为 SPSS 数据,从而便于研究数据的分析呢?

本章主要学习使用 SPSS 软件对教育研究数据文件的建立,结合已回收的调查问卷的数据,详细介绍利用 SPSS 软件对调查问卷数据的结构与定义、单项选择题、多项选择题和非选择题的数据录入的方法。本章的重难点是如何在 SPSS 统计软件中设置教育数据的结构与定义的方法,以及单项选择题、多项选择题和非选择题录入的区别和技巧。

方法解读

教育研究数据文件的建立主要包括教育研究数据文件的建立和数据文件的录入。教育研究数据文件的结构定义包括对变量名、类型、宽度、列宽度、变量名标签、变量值标签、缺失值、度量尺度等信息的定义。教育研究数据的录入包括对单项选择题、多项选择题和非选择题的录入。

1.2 SPSS 数据的结构与定义

问题导入

通过调查问卷的发放和回收,我们已经收集到了大量的诸如免费师范生远程可视化学习平台需求的调查问卷、中国高校信息化应用质量与效果评价调查问卷(华中师范大学学生问卷)、湖北省农远工程学员能力现状和培训学员培训需求调查问卷的第一手教育数据资料,那么怎么将问卷中填写的数据转化为 SPSS 数据,以便于进行研究数据的分析呢?

1.2.1 SPSS 数据结构与定义方法的基本原理

核心概念

SPSS 数据的结构是对 SPSS 每列变量及其相关的属性的描述,它的定义是通过数

据编辑窗口中的变量视图实现的。

变量名 (Name) 是变量访问和分析的唯一标识。

数据类型 (Type) 是指每个变量取值的类型。SPSS 提供的变量类型共有 4 类 8 种,相应的类型都有默认的列宽 (Width) 或小数位宽 (Decimals)。

变量名标签 (Label) 是对变量名含义的进一步解释说明,它可增强变量名的可视性和统计分析结果的可读性。

变量值标签 (Values) 是对变量取值含义的解释说明信息,对于定类型数据和定序型数据尤为重要。

统计学依据数据的计量尺度 (Measure) 将数据划分为三大类:定距型数据 (Scale)、定序型数据 (Ordinal) 和定类型数据 (Nominal)。

原理解析

1. 变量名 (Name)

变量名一般不能多于 8 个字符,首字母应以英文字母开头,后面可以填写除了“?”“!”“*”之外的字母或数字;变量名不区分大小写字母,下划线、圆点不能为变量名的最后一个字符;变量名不能使用 SPSS 的内部特有的具有特殊含义的保留字;允许汉字作为变量名,汉字总数不超过 4 个。

2. 变量类型 (Type)

表 1.2.1 SPSS 提供的变量类型

变量类型	说明
Numeric	默认类型,标准数值型,同时需要定义值宽度和小数位数,默认分别为 8 和 2
Comma	加逗点的数值型,整数部分从右至左三位加一逗号,如 100,000,000
Dot	3 位加点数数值型,数值均以整数显示,每三位加一小点,小数位置都显示 0,小数点用“.”。如 1.2345 表显示为 12.345,00
Scientific notation	科学计数型
Date	日期型,可以选择相应的格式,如 mm/dd/yy、dd.mm.yy 等
Dollar	货币型,可以选择相应的格式,如 \$##,###等。“\$”符号自动显示
Custom currency	自定义类型
String	字符串类型,需要设置字符串的长度,默认为 8

3. 变量名标签 (Label)

变量名标签可用中文,总长度可达 120 个字符。变量名标签这个属性是可以省略的,但建议最好给出变量名标签。

在变量视图【Variable View】标签下,点击【Label】列下相应行的位置,输入变量名标签即可。

4. 变量值标签 (Values)

变量值标签是对变量取值的进一步说明。定义变量值标签可以更方便地对数据进行处理。

在变量视图【Variable View】标签下,单击【Value】列下相应行的位置,根据实际数据在弹出窗口中指定变量标签。

5. 计量尺度(Measure)

(1) 定距型数据(Scale)。

定距型数据通常是指诸如身高、体重等连续数值型数据,也包括诸如人数、频数等离散数值型数据。

(2) 定序型数据(Ordinal)。

定序型数据具有内在固有大小或者高低顺序,但它又不同于定距型数据,一般可以用数值或者字符表示。例如:职称变量可以有低级、中级和高级三个取值,可分别用1、2、3表示。年龄段变量可以有青、中、老三个取值,可以用A、B、C表示等。需要注意的是,无论是数值型数据还是字符型数据,都是有固有大小或者高低顺序的,但是数据之间却是不等距的。

(3) 定类型数据(Nominal)。

定类型数据是指没有内在固有大小或者高低顺序,一般以数值或者字符表示的分类数据。如性别变量中的男女取值,可以分别用1、2表示。这里的数值型数据或者字符型数据都只是一种名义上的指代,都不存在内在的固有大小或者高低顺序。

在SPSS数据编辑窗口的变量视图【Variable View】中,单击【Measure】列下相应行的位置,根据实际数据指定变量的计量尺度。



知识卡片

SPSS数据文件是一种有别于其他文件的特殊格式的文件。从应用角度理解,这种特殊性表现在两方面:第一,SPSS数据文件的扩展名是.sav;第二,SPSS数据文件是一种有结构的数据文件,它由数据的结构和内容两部分组成。其中,数据的结构记录了数据类型、取值说明、数据缺失情况等必要信息,数据的内容才是那些待分析的具体数据。

建立SPSS数据文件是利用SPSS软件进行数据分析的首要工作。建立SPSS数据文件时应完成两项任务:第一,描述SPSS数据的结构;第二,录入编辑SPSS的数据内容。这两部分分别在SPSS数据编辑窗口的变量视图和数据视图完成。

基本操作

数据结构与定义方法的主要操作以“免费师范生远程可视化学习平台需求调查”问卷的“A个人情况”部分的第2题为例,在SPSS中设置该题数据的结构与定义。

A.2. 您的民族是()

汉族 少数民族

(1) 打开“PASW Statistics 18”统计软件,单击【File】→【New】→【Date】命令,在变量视图【Variable View】下进行数据的结构定义。

(2) 单击【Name】列下对应行的位置输入“A.2 民族”,定义变量名。

(3) 单击【Type】方框右侧,出现【Variable Type】对话框定义数据类型,在对话框中选择“Numeric(标准数值型)”,设置右侧的“Width”为8,“Decimal Place”为2。

(4) 单击【Label】列下对应行的位置输入“民族”,定义变量名标签。

(5) 单击【Value】列下对应行的位置,出现【Value Labels】对话框设置变量值标签,进行选项的设置定义,在“Value”中输入数值1,在“Label”中输入“汉族”单击“Add”后,继续在“Value”中输入数值2,在“Label”中输入“少数民族”单击“Add”,设置变量值标签为“1 = 汉族、2 = 少数民族”。

(6) 单击【Missing】、【Columns】和【Align】列下对应行的位置,设置数据的缺失值、列和对齐方式。

(7) 单击【Measure】列下对应行的位置根据实际数据选择,定义计量尺度。本例定义为“Nominal(定类型数据)”。

其他题目的数据结构定义可参照上述操作步骤定义,当将整份问卷的所有题目定义完成之后即可按照定义的顺序依次录入回收到的问卷数据。

1.2.2 案例解析:免费师范生性别情况的数据结构与定义

案例呈现

选择“免费师范生远程可视化学学习平台需求调查问卷”的A“个人情况”部分的第1题,在SPSS中设置免费师范生性别的数据文件的结构定义。

A.1. 您的性别是()

男 女

操作步骤

(1) 打开“PASW Statistics 18”统计软件,单击【File】→【New】→【Date】命令,在变量视图【Variable View】下进行数据的结构定义,如图1.2.1所示。

(2) 在【Name】列下相应行的位置输入“A.1 性别”,定义变量名,如图1.2.2所示。

(3) 单击【Type】方框右侧,打开“Variable Type”对话框,选择“Numeric(标准数值型)”,设置“width”为8,“Decimal Places”为2。如图1.2.3所示。

(4) 单击【Label】列下相应行的位置输入“性别”,定义变量名标签。如图1.2.4所示。

(5) 单击【Value】相应行的位置,会出现“Value Labels”对话框,在“Value”中输入数值“1”,在“Label”中输入“男”单击“Add”后,继续在“Value”中输入数值“2”,在

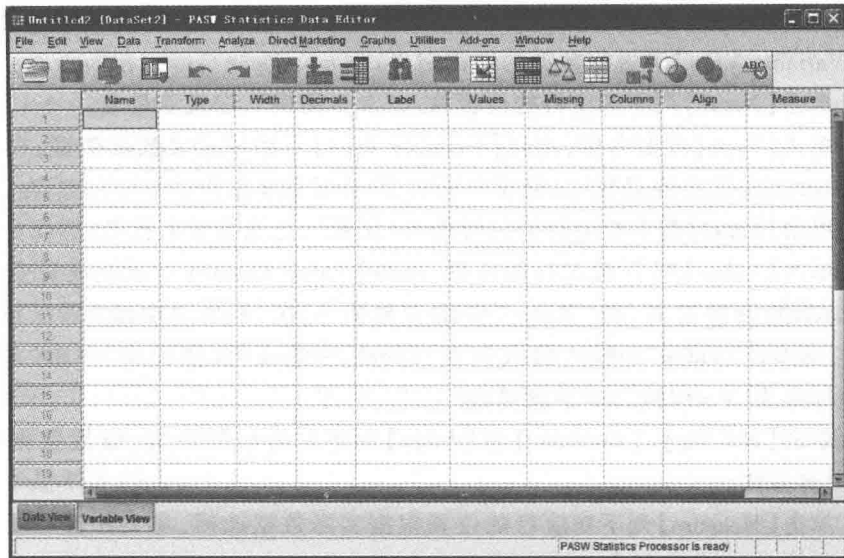


图 1.2.1 “Variable Views”视图窗口



图 1.2.2 变量名的定义

“Label”中输入“女”单击“Add”，设置变量值标签为“1 = 男、2 = 女”。如图 1.2.5 所示。

(6) 单击【Missing】列下对应行“A.1 性别”按钮在打开的“Missing Values”对话框中选择默认“No Missing Values”选项。如图 1.2.6 所示。

(7) 单击【Measure】相应行，在出现的下拉菜单中根据实际数据指定变量的计量尺度即可。本例为“Nominal(定类型数据)”，如图 1.2.7 所示。

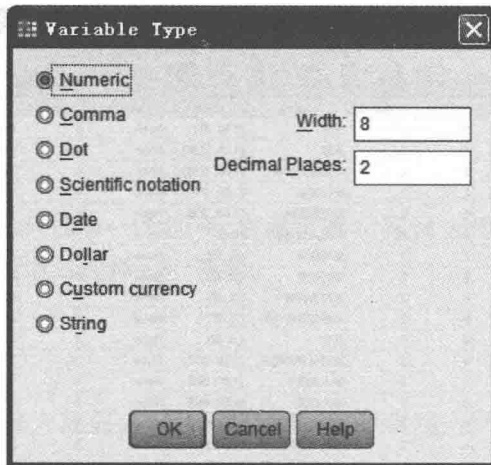


图 1.2.3 定义数据类型

	Name	Type	Width	Decimals	Label	Values	Missing
1	A.1性别	Numeric	8	2	性别	{1.00, 男}...	None
2	A.2民族	Numeric	8	2	民族	{1.00, 汉族}...	None
3	A.3专业所...	Numeric	8	2	专业所属学科	{1.00, 文科}...	None
4	A.4政治面貌	Numeric	8	2	政治面貌	{1.00, 中共}...	None
5	A.5家庭所...	Numeric	8	2	家庭所在地	{1.00, 直辖}...	None
6	A.6家庭人...	Numeric	8	2	家庭人均年收入	None	None
7	B.1.1新闻媒体	Numeric	8	2	新闻媒体	{0, 否}...	None
8	B.1.2相关政策	Numeric	8	2	相关政策	{0, 否}...	None
9	B.1.3老师告诉	Numeric	8	2	老师告诉的	{0, 否}...	None
10	B.1.4亲戚朋...	Numeric	8	2	亲戚朋友告诉的	{0, 否}...	None
11	B.1.5其它	Numeric	8	2	其它	{0, 否}...	None
12	B.2高中同学	Numeric	8	2	选读免费师范生	{1.00, 很多}...	None
13	B.3家人态度	Numeric	8	2	家人态度	{1.00, 很支}...	None

1.2.4 定义变量名标签

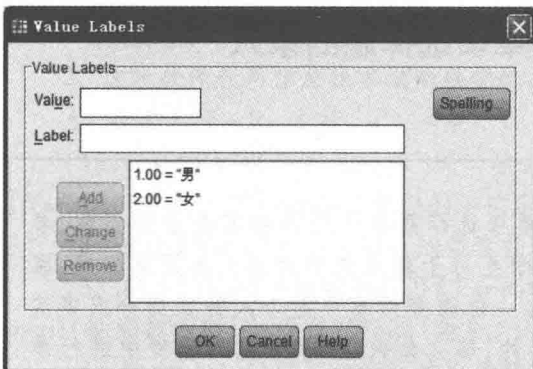


图 1.2.5 定义变量值标签

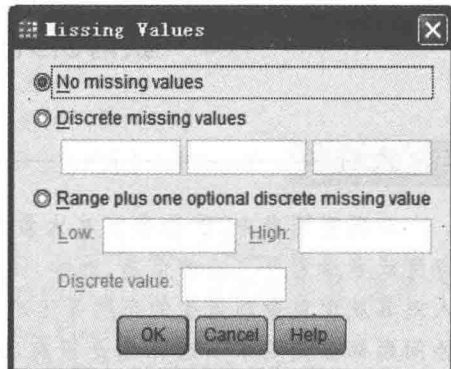


图 1.2.6 定义数据的缺失值

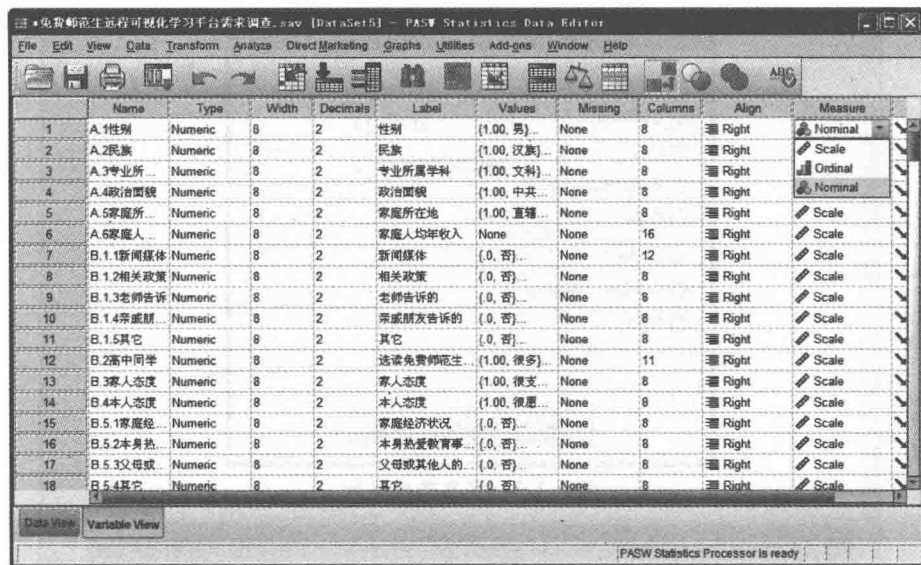


图 1.2.7 定义变量的计量尺度

结果分析

通过上述操作步骤依次完成对变量名、数据类型、变量值标签和变量的计量尺度的定义,对于问卷中的其他题目也可参考上述操作步骤设置,实现在 SPSS 中定义调查问卷数据文件结构。

小贴士

了解了 SPSS 数据的结构之后,就需要掌握描述和定义结构的操作方法。定义 SPSS 数据结构的操作是在数据编辑窗口的变量视图进行的。定义结构的具体操作方法完全遵从 Windows 下的其他常用软件的操作方法,如使用下拉框、单选框等。

1.3 教育研究中单项选择题的录入

问题导入

在调查问卷中,往往会存在大量的单项选择题用于调查被调查者的态度、参与度或者满意度之类的信息,例如,调查学生对于就读免费师范生的愿意程度、家人对其就读免费师范生的态度等信息,在一份调查问卷中的个人特征部分或事实性问题部分使用单项选择题是很有必要的,那么如何在 SPSS 统计软件中进行单项选择题的录入呢?在数据结构的定义时有哪些注意事项呢?