

普通高等教育“十二五”规划教材
全国高等医药院校规划教材

供临床、基础、预防、护理、药学、口腔医学类专业用

医学统计分析

主编 张永爱

清华大学出版社

普通高等教育“十二五”规划教材
全国高等医药院校规划教材

医学统计分析

主编 张永爱

清华大学出版社
北京

内 容 简 介

全书共分 16 章,主要介绍了医学统计分析的基础理论、统计分析的理论背景、临床实际科研资料的各种统计分析方法、 t -test 和方差分析(ANOVA)、平均数之间的比较、卡方检验(χ^2 test)、非参数检验、相关分析和回归分析、生存分析、因子分析和信度分析、医学论文中的统计学要求以及 SPSS 统计软件的使用方法。书中提供了 60 多个临床科研数据,采用 SPSS 格式建立数据文件,按照书上给出的 SPSS 操作步骤单击 SPSS 软件界面上的菜单,便可轻松得出书中所给结果,可运用于同类资料的分析中。

本书可供临床、药学、预防、护理专业学生使用。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

医学统计分析/张永爱主编.--北京:清华大学出版社,2015

普通高等教育“十二五”规划教材 全国高等医药院校规划教材

ISBN 978-7-302-39986-5

I. ①医… II. ①张… III. ①医学统计—统计分析—软件包—高等学校—教材 IV. ①R195.1-39

中国版本图书馆 CIP 数据核字(2015)第 086512 号

责任编辑:罗 健 王 华

封面设计:戴国印

责任校对:赵丽敏

责任印制:王静怡

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

课 件 下 载: <http://www.tup.com.cn>, 010-62770175-4134

印 装 者:清华大学印刷厂

经 销:全国新华书店

开 本:185mm×260mm 印 张:21.5

字 数:548 千字

版 次:2015 年 8 月第 1 版

印 次:2015 年 8 月第 1 次印刷

印 数:1~2500

定 价:49.80 元

产品编号:063934-01

编者名单

主 编

张永爱

副主编

韩湘淑 曹明芹 樊 霞

编 委

(按姓氏拼音顺序)

曹明芹 (新疆医科大学公共卫生学院)

樊 霞 (延安大学医学院)

韩湘淑 (韩国庆熙大学护理学院)

雷 静 (西安医学院)

李述刚 (石河子大学医学院)

庞宝华 (延安大学医学院)

芮东升 (石河子大学医学院)

唐争艳 (西安医学院护理学院)

杨 蕾 (新疆医科大学公共卫生学院)

姚雪梅 (新疆医科大学公共卫生学院)

张国珍 (新疆医科大学公共卫生学院)

张海苗 (西安医学院护理学院)

张景玉 (石河子大学医学院)

张永爱 (西安医学院护理学院)

前言

PREFACE

从事医疗工作的人经常会碰到统计分析的问题。每天经手大量的病历记录将成为数据文件，最后写成报告书。但是，如果不懂统计分析的方法，就很难写出正确的报告书。尤其对于在各医疗卫生单位及社区服务站独立从事各项临床、药学、预防、护理工作的应用型高级医务人才，必须要掌握统计分析的方法。

目前国内现有的很多统计学的教材，主要侧重于复杂的公式和统计学的各种基础原理的阐述。然而在临床工作中，医务工作者面临的最大问题是在处理数据时不知该选用何种分析方法，无从下手。作者从事临床教学工作10余年，除了临床专业课程以外，也曾担任统计学理论课程的教学，对统计分析产生了浓厚的兴趣。我一直有个梦想，希望能够把复杂的数学公式、难以理解的统计分析理论，借助SPSS软件并结合临床医疗工作及科研工作中收集的实例，用通俗易懂的语言，详细地介绍统计学方法，并对于SPSS输出的结果尽可能给出合理、详尽的解释，即统计学方法、SPSS操作、结果分析解释齐头并进，尤其希望强调常用的医学统计分析方法的介绍与分析结果的解释。从本书书名可想而知，这本书不是统计学理论教材，而是关于统计分析相关的书籍，它将教给初学者实用的医学统计分析原理及具体分析方法。书中提供了60多个临床科研数据，采用SPSS格式建立数据文件，按照书中给出的SPSS操作步骤选择SPSS软件界面上的菜单，便可轻松得出书中所给结果，亦可运用于同类资料的分析中。

本书由以下章节组成。首先，第1~5章主要介绍了关于医学统计分析的基础理论。阐述中尽可能避开复杂的数学公式和深奥的理论，只对统计分析方法所涉及的理论进行了简述。第1章先用简单的例题让学生理解统计分析的全部过程。第2~5章为统计分析的理论背景，介绍了调查设计、实验设计、描述性统计、推断性统计等。第6~12章为分析临床实际科研资料的各种统计分析方法。其中，第6章和第8章分别为 t -test和方差分析(ANOVA)，用于平均数之间的比较；第7章为 χ^2 检验(χ^2 test)，用于分析分类变量之间的独立性和从属性；第9章为非参数检验；第10~12章为相关分析和回归分析，用于分析因子之间的关系；第13章为生存分析；第14章介绍了因子分析和信度分析，以评价测量的效度和信度；第15章介绍了医学论文中的统计学要求；第16章则介绍了SPSS统计软件的使用方法，讲述使用SPSS软件进行数据分析的基本步骤和数据的编辑和整理。本书例题中相关数据资料可登录出版社网站(<http://www.tup.com.cn>)下载。

由于时间紧迫，加上作者水平有限，书中一定存在许多不尽如人意的地方，非常欢迎读者的批评指正，恳请各位读者通过E-mail(875019131@qq.com)等方式给予指正。

张永爱 韩湘淑

2015年5月 157

CONTENTS

目 录

第1章 绪论 1

- 第1节 什么是医学统计学? 1
- 第2节 统计学的若干基本概念 2
- 第3节 利用 SPSS 软件进行统计
分析的步骤 4

第2章 调查研究设计 10

- 第1节 调查研究概况 10
- 第2节 调查研究设计的基本内容 11
- 第3节 常用的抽样方法 14
- 第4节 调查研究常用的样本量
估计 17
- 第5节 调查研究的质量控制 19

第3章 实验研究设计 21

- 第1节 实验设计的基本要素 21
- 第2节 实验设计的基本原则 23
- 第3节 实验研究常用的样本含量
估计 26
- 第4节 常用的实验设计方案 29
- 第5节 临床试验设计 32

第4章 统计描述 35

- 第1节 定量资料的统计描述 35
- 第2节 定性资料的统计描述 54
- 第3节 统计表与统计图 61
- 课后练习题 77

第5章 推断性统计 80

- 第1节 参数估计 80
- 第2节 假设检验 86
- 第3节 假设检验与区间估计的关系 90

第6章 t 检验(t -test) 92

- 第1节 研究方法的理解 92
- 第2节 单个样本 t 检验 93
- 第3节 两独立样本 t 检验 96
- 第4节 配对样本 t 检验 99
- 课后练习题 103

第7章 χ^2 检验 106

- 第1节 研究方法的理解 106
- 第2节 配合度检验 114
- 第3节 独立性检验 117
- 第4节 同质性检验 121
- 课后练习题 124

第8章 方差分析 126

- 第1节 研究方法的理解 126
- 第2节 单因素方差分析 129
- 第3节 双因素方差分析 136
- 第4节 重复测定资料的方差
分析 143
- 课后练习题 157

第9章 非参数检验 160

第1节 Wilcoxon 符号秩和检验 161

第2节 两独立样本比较的秩和检验 166

第3节 成组设计多个独立样本比较的秩和检验 173

第4节 多个相关样本的秩和检验 181

课后练习题 184

第10章 相关性分析 186

第1节 相关关系的理解 186

第2节 简单相关分析 187

第3节 偏相关分析 193

课后练习题 196

第11章 回归分析 197

第1节 一元线性回归 199

第2节 多元线性回归 208

课后练习题 219

第12章 logistic 回归 220

第1节 logistic 回归模型的基本原理 220

第2节 二分类 logistic 回归模型的应用 225

课后练习题 239

第13章 生存分析 240

第1节 生存资料的特点及生存分析中的基本概念 240

第2节 生存率的寿命表法估计 242

第3节 生存率的 Kaplan-Meier 法估计 246

第4节 生存率的比较 251

第5节 Cox 比例风险回归模型 255

课后练习题 260

第14章 量表的效度和信度分析 261

第1节 效度和信度的定义 261

第2节 因子分析 263

第3节 信度分析 273

课后练习题 277

第15章 医学论文中的统计学要求 279

第1节 数据的质量控制 279

第2节 数据处理的几个常见问题 281

第3节 统计方法选择 282

第4节 统计分析的结果表达 286

第16章 SPSS 统计软件的应用 289

第1节 SPSS 统计软件的介绍 289

第2节 SPSS 进行数据分析的基本步骤 290

第3节 数据的编辑和整理 297

常用统计表 314

中英文名词对照 330

参考文献 334

第1章

绪论

第1节 什么是医学统计学?

日常生活中,我们经常会接触到大量的信息和数据,例如人均住房面积、国内生产总值等,对这些信息和数据的认识和理解将直接影响到我们对生活的认识和理解。同样,在医学研究领域,大量的信息和数据提示着医学科学的发展和进步。医务工作者每天都在接触患者的信息、治疗的信息、医院的信息、医学文献的信息等,医疗技术的提高和医学科学的发展离不开对这些信息和数据的分析和利用,这些都离不开统计学。

什么是统计学?统计学是收集和分析数据的科学或艺术。生物医学中充满了不确定性,比如,每个人的血压都不一样,每个高血压患者对同一种药物治疗的反应亦存在差异,这就是所谓的不确定性和个体差异,统计学中把不确定性和个体差异称为变异(variation)。因变异存在,医学需观测和分析一定数量的个体数据,从中探索和发现数据背后蕴含的规律。统计学正是处理数据中变异和不确定性的一门科学和艺术。所谓科学,指收集和利用的数据和信息应真实、可靠,实事求是,通过合理和科学的数据分析得到尽可能真实、可靠的结论;所谓艺术,指研究者需运用统计学观念和统计思维,基于数据的特点,选择适当、高效的统计分析方法,挖掘数据中蕴含的信息,从不确定性、机遇、风险和推断的角度,思考和解释问题。

为了总结经验、获取信息、发现知识,实现科学管理与决策,几乎所有的科学技术都需要统计学的帮助,医学研究更需如此。医学与统计学的结合是必然的,统计学是一门贯穿医学研究始终的学科。例如,为了检验某种新药对改善血液循环是否有帮助,涉及一系列的统计学知识:如何确定研究对象?研究对象又如何分组?什么对象作为试验组,什么对象作为对照组?不同组之间的治疗效果有无差异?为了节约成本,如何由较少对象的观察资料推断一般人群的疗效?为了获得有用、可靠的信息,每组至少需要多少人?等等。

医学统计学(medical statistics)是应用统计学的基本原理和方法来研究医学问题的一门学科,它包括了研究设计,数据收集、整理、分析以及分析结果的正确解释和表达。

第一次接触统计学的人大多数都会有恐惧的心理,因为一听到“统计”这个词的瞬间,脑子里首先出现的就是一大堆数字、复杂的公式及符号,加上置信区间、假设检验、 P 值等统计术语很难理解,以及周围朋友的负面影响,如“统计很难学懂,记住公式,不等于理解统计;学会计算,不等于会用统计”等,导致人们丧失学习统计学的信心。

随着科技的进步、计算机的普及、大量统计软件的涌现,统计公式和计算变得越来越容易。医学工作者利用统计软件可以进行各类数据的整理和汇总,完成各类简单或者复杂的统计分析,绘制各类统计图表表达和解释数据,探索和分析医学问题。计算机软件在给我们带来方便和快捷的同时,

亦存在一些问题,比如数据只要符合统计软件的格式和参数设置,就可以分析出结果,如果忽视医学统计学的基本原理和方法,势必导致统计学方法的误用和滥用。现实中,用于研究设计、数据收集、统计分析、结果解释以及结论报告任何一个环节中的缺陷或错误,已经断送了许多先进的甚至非常前沿的医学研究结果,即使是已经发表的研究成果或文献中,也存在大量的统计学错误和缺陷。

因此,本教材的目的是借助常用的 SPSS 统计软件,淡化统计公式和数据计算,介绍统计分析的基本原理和方法,根据研究目的和假设,选择适宜的统计分析方法,结合统计结果和医学背景,以期合理地表达和解释数据分析的结果,从而客观、科学地认识医学发展的规律。

第 2 节 统计学的若干基本概念

一、总体和样本

假定研究者欲了解一个城市居民高血压的患病现状,通常的做法是在这个城市调查一部分居民,利用这部分居民的高血压患病状况来反映整个城市的高血压患病状况。整个城市所有居民就是研究的总体。总体(population)是根据研究目的确定的所有同质观察单位某种观察值(即变量值)的集合。例如,研究某种药物对高血压患者的疗效,则所有服用该药物的高血压患者就是研究的总体,更确切地说,所有服用该药物的患者的疗效组成的集合即为总体。

研究和收集总体中的每一个观察个体是不现实也是不太可能实现的。例如,高血压患者总体随着时间和空间的延展,可以无限大。通常的做法是随机观察一部分服用药物的高血压患者,收集他们的治疗效果,统计分析,然后对药物治疗效果下结论。这一部分高血压患者被称为样本。样本(sample)是总体中随机抽取部分观察单位的观测值组成的集合。

样本是总体中的一部分,采用样本信息对总体信息进行推断即为统计推断。要保证统计推断的可靠性,样本需对总体具有代表性。样本必须从总体中随机抽取,保证每个个体被抽取到的机会均等,而且样本中应包含足够的观察单位数。

医学研究需在总体层面上得出结论,但收集和观测总体中的每个个体是不现实的,从总体中随机抽取一部分个体(样本),观察样本的信息和数据,依据样本信息对总体作出结论,称为抽样研究。医学研究中常采用的是抽样研究,抽取的样本为总体的一部分,且个体间存在差异(个体变异),这导致由样本信息推断总体信息时,存在着误差,即样本信息和总体信息不完全一致,统计学上将其称为抽样误差(sampling error)。抽样误差在抽样研究中是不可避免的,统计推断时必须考虑抽样误差的大小下结论,这就是统计学思维。

二、参数和统计量

总体的数值特征称为参数(parameter),用希腊字母表示。根据研究目的确定的总体,理论上,总体参数是客观存在和固定不变的。例如,研究某地健康男性的血红蛋白的平均水平,常用希腊字母 μ 表示总体均数。根据样本算得的某些数值特征称为统计量(statistic),常用英文字母表示。例如,随机抽取该地健康男性 100 例,实验室检测其血红蛋白含量,可计算出 100 名男性血红蛋白的平均水平,此为样本均数,常用 $\bar{X}$ 表示样本均数。因抽取的个体血红蛋白存在差异,抽取不同的样本,得到的样本均数可以不同。在抽样研究中,由个体变异产生,随机抽样引起的样本统计量与总体参数之间的差别,称为抽样误差。

三、变量的类型

研究观察对象的某种特征、属性或指标,称为变量(variable),变量的观测值称为变量值,也称为数据或资料。例如,该地 100 名健康男性的血红蛋白含量,变量为血红蛋白,100 个血红蛋白数值组成数据或资料(data)。根据变量值是数值的还是分类的,可将变量分为定量变量和定性变量。

1. 定量变量 亦称数值变量,其变量值是定量的,表现为数值大小,一般有度量衡单位。如调查某年某地 7 岁女童的生长发育状况,以人为观察单位,则女童的身高(cm)、体重(kg)、血红蛋白(g/L)等均属定量变量。

2. 定性变量 亦称分类变量,其观察值是定性的,表现为互不相容的类别或属性,分为以下两种情况。

(1) 无序分类变量:①二项分类。如调查吸毒者的 HIV 感染情况,结果分为阳性与阴性两类,表现为互不相容的两类属性。②多项分类。如人类的 ABO 血型,以人为观察单位,结果分为 A 型、B 型、AB 型与 O 型,表现为互不相容的多个类别。

(2) 有序分类变量:各类之间有程度的差别,给人以“半定量”的概念,亦称等级变量。如测定某人群某血清学反应,以人为观察单位,结果可分一、±、+、++ 4 级;又如观察某药治疗十二指肠溃疡的疗效,以每个患者为观察单位,治疗效果分为痊愈、显效、好转、无效 4 级。

数据分析时,常需要依据不同的分析目的,对变量类型进行转换。例如,患者的年龄(岁)为定量变量;而将患者的年龄分为<20 岁、20~29 岁、30~39 岁、40~49 岁、50~59 岁、≥60 岁等年龄段时,则为有序分类变量;若将肿瘤患者的年龄直接分为≤60 岁和>60 岁,则转换为二分类变量。再比如,血红蛋白值为定量变量;临床研究时,将女性的血红蛋白值依据几个界点划分为正常、轻度贫血、中度贫血、重度贫血,则转换为有序分类变量;进一步可以转换为正常、异常二分类变量。变量类型的区分在统计学分析中至关重要,不同的变量类型采用的统计分析方法不同,应学会依据不同资料选择适合的统计方法。

软件输入数据时,常见的数据格式为数值型和字符型。例如,血红蛋白具体的值为数值型变量。贫血情况(正常、轻度贫血、中度贫血、重度贫血)为有序分类变量,若直接输入汉字分类,则软件默认为字符型,为了输入数据快捷,常采用数值编码输入,可以输入 1、2、3、4 数值,分别表示正常、轻度贫血、中度贫血、重度贫血,此时软件默认为数值型。再例如,性别分为男性和女性,若直接在软件中输入“男”和“女”,则为字符型;若输入“1”、“2”分别表示男性和女性,则软件默认为数值型。统计方法是根据变量原有的类型选择的,而不是根据软件的数据格式。

四、概率

概率是描述随机事件发生可能性大小的数值,常用 P 表示。随机事件概率的大小在 0 与 1 之间, P 越接近 1,表示某事件发生的可能性越大; P 越接近 0,表示某事件发生的可能性越小。习惯上,将 $P \leq 0.05$ 的事件,称为小概率事件,表示在一次实验或观察中该事件发生的可能性很小,统计学上可视不太可能发生。

小概率事件是统计推断的基本依据,通常是依据一次抽样的信息对总体进行统计推断,若一次抽样样本信息支持某总体假设的可能性非常小,为小概率事件,则通常认为样本不是来自该总体,结论为放弃该假设。

五、统计工作的步骤

统计工作的步骤包括统计设计、收集资料、整理资料和分析资料。其中,统计设计是最首要、最

关键的一环,是后续步骤的依据。医学研究分为调查研究和实验研究,统计设计相应分为调查研究设计和实验研究设计。没有科学、严谨的统计设计,数据的收集以及分析常常是没有价值的。对于不准确或不可靠的数据,试图寻求统计方法加以弥补亦是徒劳无益的。

统计资料主要来自:①卫生统计报表;②经常性工作记录;③专题调查或实验研究。整理资料的目的是净化原始数据,使其系统化、条理化。分析资料即通过计算统计指标,反映数据的综合特征,阐明事物的内在联系和规律。

统计分析包括统计描述和统计推断两部分。统计描述(statistical description)指运用统计指标,如平均数、标准差、率以及统计表和统计图等,对数据的数量特征及其分布规律进行客观的描述和表达,不涉及样本推论总体的问题。例如,某护士想了解初一学生的身高,通常是先取得数据,对数据进行归纳分类,列表画图,再计算出这些数据的代表值以及衡量这组数据分散程度的数量指标,用以描述这组数据的特征。

然而,在实际工作中我们往往需要利用一组数据获得的信息做决策去推断更一般的情形。例如,我们想知道吸烟对新生儿的健康有无影响,实际上无法对所有的吸烟产妇所生的新生儿进行调查,而是选取样本,通过样本的描述来推断总体的特性,这种方法称为统计推断。统计推断(statistical inference)指在一定的可信程度或概率保证下,根据样本信息去推断总体特征。统计推断包括参数估计和假设检验两部分。参数估计指用样本统计量估计总体参数,例如,用样本均数 $\bar{X}$ 估计总体均数 μ ;采用 100 名某市健康男性的血红蛋白平均水平估计该市所有健康男性的血红蛋白的平均水平。假设检验是对总体进行假设,依据一次抽样的样本信息支持该假设的可能性作出推断结论,决定是否拒绝该假设。

第 3 节 利用 SPSS 软件进行统计分析的步骤

本节主要介绍统计分析的步骤,并以具体的临床实例来加深理解。

(1) 提出研究问题:根据研究设计过程先简要地整理出需要解决的问题。

(2) 选定变量,收集数据:确定研究对象的属性和指标,实验检测或问卷调查,收集数据。

(3) 录入数据:将原始数据录入到 SPSS 软件。

(4) 分析数据:利用 SPSS 统计软件分析。

(5) 撰写研究报告书:解释利用 SPSS 分析的结果,并写出报告书,将得出的结果应用在实际问题中。

一、提出研究问题

例 1-1 通常很多人认为吸烟会影响胎儿的生长发育。在 K 医院妇产科工作的某一位护士经常会遇到吸烟孕妇提出的问题:“我是否一定要戒烟?”这位护士为了验证“吸烟会影响胎儿的健康”这个事实,将吸烟产妇所产的新生儿作为研究对象,调查其胎龄。正常产妇的孕周为 40 周,所以正常新生儿的胎龄为 40 周。如果吸烟对胎儿的生长发育有影响而引起胎龄发生改变,则吸烟产妇所产的新生儿的胎龄不是 40 周。本研究的目的是分析吸烟产妇所产的新生儿的胎龄是否也是 40 周?

二、选定变量,收集数据

我们要研究的对象为吸烟产妇所生的新生儿,则要测量的变量为胎龄,其单位为孕周,这是个具体的数值,需要测量,故为连续性定量数据。在 SPSS 软件中,变量类型为数值变量(numeric),测量

尺度为定量尺度(scale)。

变量(variable)	定义(definition)	类型(type)	测量尺度(measure)
胎龄(即孕周)	孕妇的妊娠时间	数值变量	定量尺度(周数)

该护士收集了 40 名吸烟孕妇新生儿的孕周,数据如下:

36.5	37.2	37.6	38.1	38.4	38.4	38.6	39.1
37.0	37.4	37.6	38.2	38.4	38.5	38.6	39.1
37.1	37.4	38.0	38.2	38.4	38.5	38.6	39.1
37.1	37.6	38.0	38.2	38.4	38.5	38.6	39.1
37.1	37.6	38.1	38.4	38.4	38.6	39.0	39.1

三、SPSS 软件录入数据

SPSS 统计软件数据输入一般均遵循以下原则:

- (1) 每一行代表一个研究对象,同一研究对象的数据通常输入在同一行上。
- (2) 每一列代表一个变量或指标,同一变量的数据应输入在同一列上。
- (3) 输入的数据应包含原始数据的所有信息。
- (4) 输入数据时尽量采用数值型变量或编码,少用或不用文字或英文字母。

安装 SPSS 软件后,双击 SPSS 软件图标,会弹出任务窗口,见图 1-1。单击“取消”,可终止任务向导。

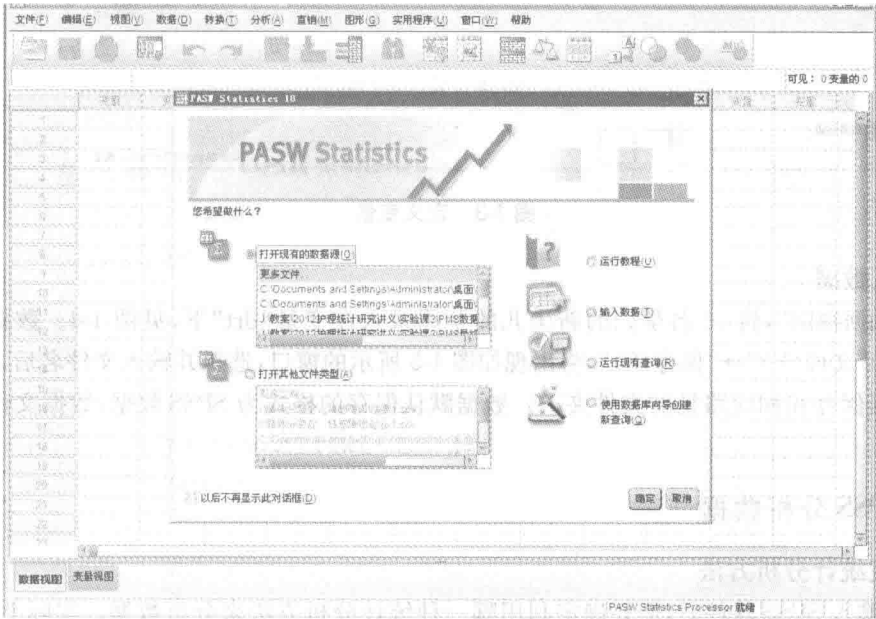


图 1-1 SPSS 软件界面

1. 打开“数据编辑”窗口

在任务窗口选择“输入数据”后,SPSS 会打开一个空白的数据编辑窗口,也称数据编辑器(PASW statistics),见图 1-2。该窗口为 SPSS 软件默认的常用窗口,可以录入、编辑、修改、浏览、保

存数据或者打开外部数据。数据编辑窗口由两个视窗组成(图 1-2 左下角):数据视图和变量视图。通过单击可以进行视窗切换。“变量视图”窗口内可以定义变量名、变量类型、宽度、小数位数、变量标签、变量值标签、度量标准等。“数据视图”窗口内可以录入、修改数据。

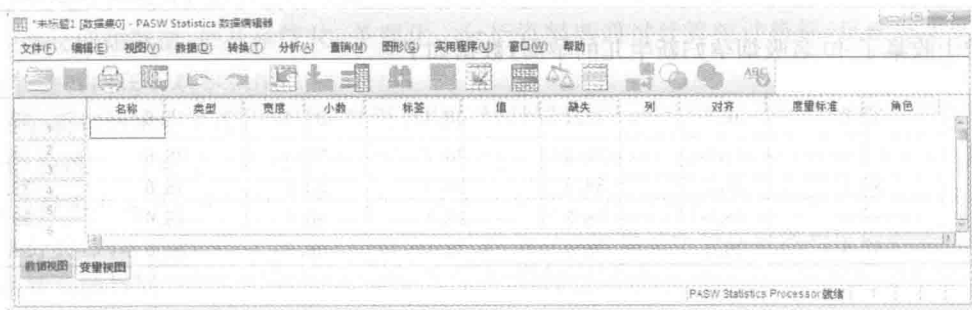


图 1-2 PASW Statistics 数据编辑器界面

2. 定义变量

单击“变量视图”,在该窗口内,定义变量名和变量属性。例 1-1 中,在名称中输入“drt”(duration 的简写),并在“标签”中输入“孕周”,见图 1-3。

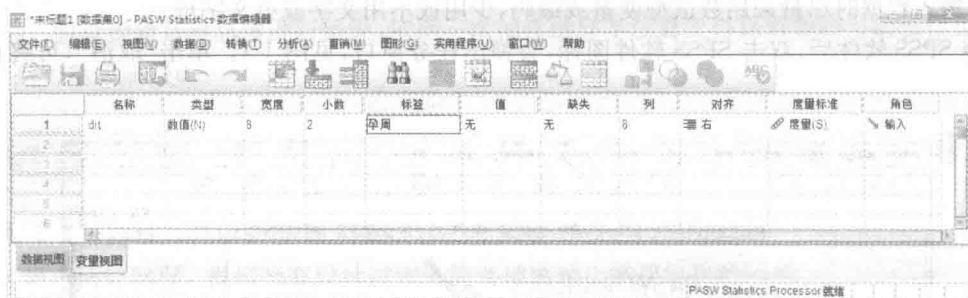


图 1-3 定义变量

3. 录入数据

单击“数据视图”,将 40 名孕妇的新生儿的孕周录入在变量“drt”下,见图 1-4。数据录入完成时,依次单击“文件(F)”→“保存(S)”,将出现如图 1-5 所示的窗口,选择并输入文件名后,单击“保存(S)”,将数据保存在相应路径的文件夹下。数据默认保存的格式为 SPSS 数据,数据文件的后缀为“.sav”。

四、SPSS 分析数据

1. 确定统计分析方法

把数据输入 SPSS 软件后,需要确定利用哪一种统计分析方法来分析数据。实际上,研究者建立数据库之前,已经确定好了分析方法。对于学习统计的人来说,这也是最难且最重要的部分。如果能够熟练地选择正确的统计分析方法,就能说明已经掌握了统计分析的方法。为此必须掌握统计学的基础理论。

确定统计分析方法首先要搞清研究目的,其次是数据的测量方法。例 1-1 为在吸烟妇女所生的新生儿群体中检验胎龄的平均值是否为 40 周,因此需要采用的分析方法是单样本的 t 检验(t -test)。



图 1-4 录入数据

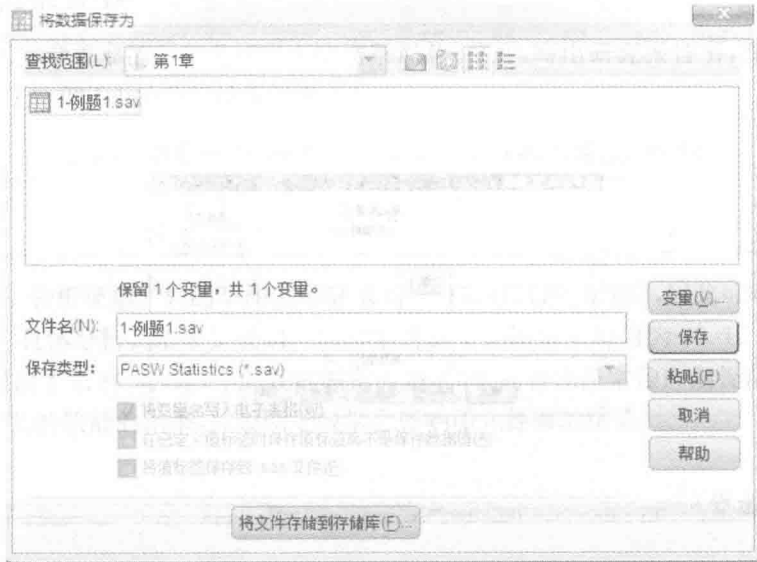


图 1-5 保存数据

2. 分析数据

(1) 在菜单中选择“分析(A)”→“比较均值(M)”→“单样本的 T 检验(S)”命令,见图 1-6。



图 1-6 分析数据界面 1

(2) 在“单样本 T 检验”对话框中,单击左边的“孕周[dt]”,再单击“”,变量就会移动到右边的“检验变量(T)”中,然后在“检验值(V)”文本框中输入“40”。单击“确定”按钮,见图 1-7。

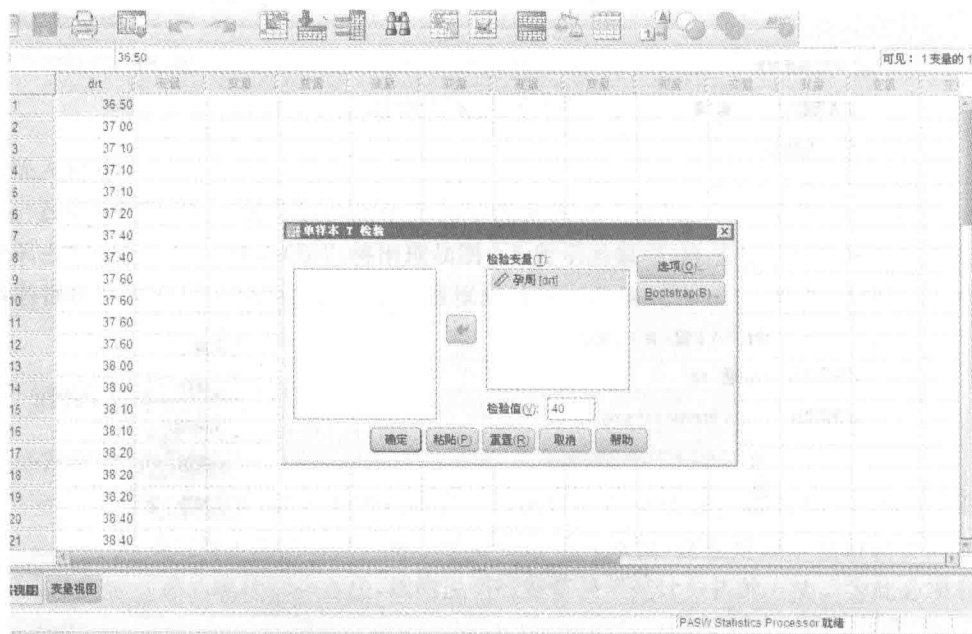


图 1-7 分析数据界面 2

(3) SPSS 分析结果显示于 SPSS 的结果浏览窗口中(SPSS Viewer)。例 1-1 的结果如下:

单个样本统计量¹

	N	均值	标准差	均值的标准误
孕周	40	38.1450	.67366	.10652

单个样本检验

	检验值 = 40				
	t	df	Sig.(双侧)	均值差值	差分的 95% 置信区间
					下限 上限
孕周	-17.415	39	.000	-1.85500	-2.0704 -1.6396

(4) 单击“文件(F)”→“保存(S)”,选择相应的路径,把分析结果存为 SPSS 结果文件,其后缀为“.spv”。

五、解释分析结果

通过 SPSS 得出的结果如何去解释? 分析结果时要注意的问题是“我们得出的数据不是所有吸烟产妇所生的新生儿中得到的结果,而是在通过 40 名吸烟产妇所生的新生儿样本中得出的结果”,继而从 40 名样本结果中推断所有吸烟产妇的孕周。这时需要的概念就是假设和概率,即我们所得到的结果是从样本中得出的,所以不能说结果是百分之百的正确。

通常统计学中利用小概率事件推断原理和 95% 可信区间进行假设推断。例 1-1 中研究假设为“ H_0 : 吸烟产妇的孕周是 40 周; H_1 : 吸烟产妇的孕周不是 40 周”。通常假设 H_0 成立,构造样本统计量 t 值。一次抽样不太可能出现小概率事件,若出现极端的 t 值,即概率 P 值(Sig. 双侧)较小,则有理由拒绝 H_0 。本例 $P < 0.001$,因此拒绝 H_0 ,接受 H_1 ,认为吸烟产妇的孕周不足 40 周。

六、写研究报告书

对于研究者,非常重要的一步是把以上的结果汇总、整理,写出研究报告书,并思考如何把结果应用在临床实际上。例 1-1 的分析结果见表 1-1。

表 1-1 吸烟产妇的新生儿孕周与正常孕妇 40 孕周的比较结果

	$\bar{X} \pm S$	t 值	P 值
孕周	38.15 $\pm$ 0.67	-17.415	0.000

其结果解释: 分析吸烟产妇的孕周结果显示 $t = -17.415 (P < 0.05)$,因此可以认为吸烟产妇的孕周不是 40 周。具体分析其结果是吸烟产妇所生的新生儿胎龄平均值为 38.15 周,比正常新生儿的胎龄(40 周)提前了 1.86 周,即,孕妇在妊娠期吸烟时,胎儿会比正常预产期提前出生。从而得出吸烟会缩短新生儿的胎龄的结果,并说明了吸烟会影响胎儿健康的结论。

(张永爱 曹明芹)

¹ 本书的软件输出结果表中数字表示形式等为软件设置特有形式。

第2章

调查研究设计

第1节 调查研究概况

一、调查研究的概念及其特点

调查(survey)一词来源于拉丁文 *supervidere*, 意为“观察”。调查研究(survey research)是指研究者在不施加任何干预措施的情况下, 客观地观察、记录和描述研究对象的特征和状态。医学领域中, 涉及了解知识、态度、信念、行为(如对艾滋病的认知程度和行为取向), 了解状况特点(如高血压的患病情况)、病因探讨(肺癌发生与吸烟的关系)等方面的研究, 较适合采用调查研究方法。

调查研究的特点: ①研究过程中无人为施加的干预措施, 研究对象及其相关特征(包括研究因素和非研究因素)是客观存在的; ②不能对研究对象进行随机分组, 也不能用随机化分组来平衡非研究因素对调查结果的影响。

调查研究的特点决定了研究者只是“被动”地观察、描述客观实际情况, 因此调查研究又称为观察性研究(observational study)。

二、调查研究的分类

1. 按照调查范围分类

可分为普查和非全面调查, 其中后者以抽样调查、典型调查最为常用。

(1) 普查(census): 亦称全面调查(overall survey), 是对总体中所有的观察单位都进行的调查, 如中国的六次人口普查。普查没有抽样误差, 所得资料全面, 但非抽样误差较大, 因此, 普查必须规定统一的标准时点、统一的普查期限、统一的普查项目和指标等。

(2) 抽样调查(sampling survey): 是遵循随机的原则, 从总体中抽取一定数量具有代表性的观察单位组成样本, 然后根据样本信息来推断总体特征的调查研究方法。医学研究的总体大多是无限总体, 因此只能进行抽样调查, 如药物疗效的观察、食品卫生质量的检查等。在抽样调查中, 要使样本具有代表性, 就必须针对调查对象的不同特点选择合理的随机化抽样方法。抽样调查在实际工作中应用最多, 是科学工作者应熟练掌握的调查研究方法。

(3) 典型调查(typical survey): 是研究者在对调查对象进行全面分析的基础上, 有目的地选择典型的观察单位进行调查。如调查某个计划生育先进县, 用于总结相关经验教训, 以便推广到其他地区。典型调查属于一种非概率抽样方法, 它没有遵循随机抽样的原则, 不是概率样本, 故不能用于估计总体参数。