

人口和就业 统计分析技术

RENKOU HE JIUYE TONGJI FENXI JISHU

国家统计局人口和就业统计司 编
中国人民大学社会与人口学院



 中国统计出版社
China Statistics Press

人口和就业 统计分析技术

RENKOU HE JIUYE TONGJI FENXI JISHU

国家统计局人口和就业统计司 编
中国人民大学社会与人口学院



 中国统计出版社
China Statistics Press

(京)新登字 041 号

图书在版编目(CIP)数据

人口和就业统计分析技术 / 国家统计局人口和就业统计司, 中国人民大学社会与人口学院编. —— 北京: 中国统计出版社, 2012. 2

ISBN 978-7-5037-6475-2

I. ①人… II. ①国… ②中… III. ①人口统计—统计分析—中国②就业统计—统计分析—中国 IV. ①C924. 25; D669. 2

中国版本图书馆 CIP 数据核字(2012)第 015521 号

人口和就业统计分析技术

作 者/国家统计局人口和就业统计司, 中国人民大学社会与人口学院编

责任编辑/徐 涛

装帧设计/杨 超 李雪燕

出版发行/中国统计出版社

通信地址/北京市西城区月坛南街 57 号 邮政编码/100826

办公地址/北京市丰台区西三环南路甲 6 号

网 址/[www. stats. gov. cn/tjshujia](http://www.stats.gov.cn/tjshujia)

电 话/邮购(010)63376907 书店(010)68783172

印 刷/北京联兴盛业印刷股份有限公司

经 销/新华书店

开 本/710×1000mm 1/16

字 数/504 千字

印 张/27.25

版 别/2012 年 4 月第 1 版

版 次/2012 年 4 月第 1 次印刷

书 号/ISBN 978-7-5037-6475-2/C. 2602

定 价/58.00 元

中国统计版图书, 版权所有。侵权必究。

中国统计版图书, 如有印装错误, 本社发行部负责调换。

人口和就业统计分析技术

编 委 会

主 任：张为民

副 主 任：冯乃林 赵云城 李希如 彭永涛

胡 英 翟振武

委 员：孟灿文 肖 宁 吴 珊 段成荣

杜 鹏 陈 卫 宋 健 刘 爽

杨菊华 宋月萍 巫锡炜

责任编辑：徐 涛

前 言

2010年,我国进行了第六次全国人口普查,获得了人口总量、结构、分布等大量的基础数据,这是一笔极为宝贵的信息财富。为了做好人口普查资料的开发利用工作,提高人口和就业统计分析人员的业务能力,更好地帮助基层统计工作者掌握人口和就业统计分析技术,国家统计局人口和就业统计司与中国人民大学社会与人口学院联合编写了《人口和就业统计分析技术》一书。

本书围绕人口普查内容,较全面地提供了人口和就业统计的分析和相关技术。全书分为17章,第1—5章介绍主要概念和基本方法,包括数据质量评估与调整、人口和就业统计的基本概念和方法、人口预测方法、人口性别年龄结构分析等;第6—17章分不同专题介绍人口和就业统计分析技术,包括人口生育、死亡、婚姻家庭、迁移流动、城镇化、老龄化、受教育状况、民族状况、劳动力供给和需求、就业、失业、住房状况等。

本书主要面向从事人口和就业统计的实际工作者,在语言上力求通俗易懂;在内容上力求全面系统;在方法上力求操作性强。为方便读者理解,本书通过分析实例对一些主要指标的计算和应用进行了说明。本书章节中提到的一些观点和方法系作者本人研究成果,供读者参考。

由于编写时间仓促,本书难免存在一些不足和疏误,请读者予以指正。

编 者

二〇一二年二月

目 录

第一章 人口数据质量评估与调整	(1)
第一节 数据和数据质量	(1)
第二节 人口数据评估	(7)
第三节 数据修正与调整	(20)
第二章 人口与就业统计中的基本概念	(27)
第一节 人口统计基本概念	(27)
第二节 就业统计基本概念	(39)
第三节 人类发展指数	(47)
第三章 人口统计中的基本方法	(53)
第一节 人口统计基本方法	(53)
第二节 指标的标准化与分解	(59)
第三节 生命表技术基础	(67)
第四章 人口预测基本原理	(82)
第一节 人口预测的目的与意义	(82)
第二节 人口预测的基本方法与模型	(86)
第三节 队列要素预测法	(90)
第五章 人口性别与年龄结构	(105)
第一节 人口性别结构	(105)
第二节 人口年龄结构	(113)
第六章 生育分析	(128)
第一节 出生和生育统计基本指标	(128)
第二节 分孩次生育统计指标与方法	(136)

第三节	生育率分析方法	(141)
第四节	间接估计方法与模型	(146)
第七章	死亡分析	(154)
第一节	死亡水平的度量	(154)
第二节	死亡率的标准化解与分解	(164)
第三节	死因统计与分析	(171)
第四节	死亡率间接估计方法	(178)
第八章	婚姻与家庭分析	(194)
第一节	婚姻、家庭的基本概念及主要数据来源	(194)
第二节	婚姻基本指标与婚姻状况分析	(201)
第三节	家庭(户)及户主分析	(210)
第九章	人口迁移流动分析	(222)
第一节	基本概念	(222)
第二节	分析迁移流动人口的基本方法	(226)
第三节	人口迁移流动指标的计算	(234)
第十章	人口城镇化与人口分布	(240)
第一节	相关概念	(240)
第二节	城镇人口统计口径	(242)
第三节	城镇化分析	(244)
第四节	人口分布	(250)
第十一章	人口老龄化分析	(256)
第一节	人口老龄化分析	(256)
第二节	人口老龄化与老龄问题分析	(260)
第十二章	人口受教育状况分析	(268)
第一节	人口普查中的受教育状况项目	(268)
第二节	受教育状况的基本指标	(270)
第三节	受教育状况的分析	(277)

第十三章 民族人口状况分析	(288)
第一节 少数民族人口总量及其变动趋势	(288)
第二节 少数民族人口的年龄和性别结构	(292)
第三节 少数民族人口的分布情况	(297)
第四节 少数民族人口素质状况	(299)
第五节 少数民族人口的就业状况	(307)
第六节 少数民族人口婚育状况	(312)
第七节 少数民族人口流动状况	(316)
第十四章 劳动力供给分析	(322)
第一节 劳动力供给假说及理论	(322)
第二节 队列要素法预测劳动力年龄人口	(330)
第三节 劳动力生命表编制方法	(339)
第十五章 就业状况分析	(346)
第一节 人口的就业水平	(346)
第二节 就业人口的性别和年龄构成	(351)
第三节 就业人口的行业结构	(362)
第四节 就业人口的职业结构	(373)
第十六章 失业分析	(384)
第一节 失业概念及其国际比较	(384)
第二节 人口普查的失业信息	(395)
第三节 人口失业的描述统计分析	(398)
第十七章 住房状况分析	(410)
第一节 基本分析	(411)
第二节 户主特征与住房状况	(417)
第三节 家庭户特征与住房状况	(420)

第一章 人口数据质量评估与调整

人口数据是进行统计分析的基础,数据质量高低直接决定分析结果的正确与否。我们在进行人口统计分析之前,必须对数据质量进行检查评估,合格的数据可以直接使用,不合格的数据需经过调整,才能应用到分析过程。在人口学体系内,数据评估是一个重要的研究领域,有特定的概念、理论和方法,并随着人口研究实践的发展,不断地丰富着研究技术和方法。本章主要从数据、数据质量评估和调整等三方面展开,主要包括:数据误差的来源与主要类型、数据评价的基本方法、年龄数据评估方法、出生死亡数据评估,以及数据修正方法等。

第一节 数据和数据质量

数据(data)是统计研究的基础,数据分析就是对数据进行加工的过程,二者的关系如同石料和建房,建房质量决定于石料的好坏和蓝图设计的科学与否。我国人口普查工作自1953年第一次开展以来,到2010年已进行了六次,积累了大量的人口数据,并在数据开发过程中产生了许多有价值的理论和方法,提出了适合我国国情的人口发展战略。这些都是基于大量数据调查和高质量数据开发的产物。所以,在数据分析过程中,必须重视数据质量的评估,从根本上减少结论失误的可能性。

1 数据与数据来源

(1)数据

数据是一切统计分析的基础,在人口学研究领域,有广义和狭义之分。狭义的人口数据是指未经任何处理的人或人口事件的属性信息,如年龄、性别、职业、婚育状况以及各种人口事件的发生时间等。广义的人口数据还包括由这些基础数据所推导出的各种人口学指标,如人口数、出生数、死亡数、出生率、死亡率、生育率等。

人口调查过程中有时又把数据分为事实数据和意向数据。事实数据反映

人的特征或其经历的人口事件,如年龄、文化程度、婚育状况、职业等。意向数据反映人的价值取向及意愿,如意愿生育子女数、对实施计划生育扶助奖励措施的满意度评价等。

本章所介绍的数据评估和调整方法针对的是广义数据,无论绝对数、相对数,事实数据、意向数据,还是数字、文字等,都在质量评估范围内,需对其质量进行严格评估,为后续数据开发奠定基础。

(2)数据来源

人口数据的来源有直接和间接之分。直接来源是指专门收集有关人口数据的各类调查,主要包括人口普查、人口抽样调查、人口典型调查和经常性人口统计(又称生命统计或民事统计)。如人口普查(census),是在国家统一规定的时间内,按照统一的方法、统一的项目、统一的调查表和统一的标准时点,对全国人口所进行的一次性全面调查登记。它是当今世界各国广泛采用的搜集人口资料的一种最基本的科学方法,是提供全国基本人口数据的主要来源。在我国,按照《全国人口普查条例》规定,人口普查每10年进行一次,尾数逢0的年份为普查年度,第六次人口普查在2010年组织实施,标准调查时点为2010年11月1日零时。人口抽样调查周期较短,如在两次人口普查之间进行全国1%人口抽样调查,最近的一次是2005年组织实施的。除人口普查和1%人口抽样调查的年份外,还有每年进行的1‰人口变动抽样调查,以及各种专项调查,如全国人口和计划生育抽样调查,这些抽样调查能对我国人口的变动情况提供丰富信息。

间接来源是指不以专门提供人口数据为目的,但包括有关人口信息的来源。如国民经济、社会各部门的统计报表、调查报告,人力资源和社会保障部门的劳动力统计资料、公安部门的户籍和迁移登记资料、卫生部门的生育手术记录、民政部门对残疾人口和老年人口的保障金、补助金的发放统计,教育部门对入学人数、毕业生数的统计等等。这些统计资料都可作为人口数据的来源,也为评价直接来源的人口数据提供了可以参考的资料。

2 数据质量与误差

(1)数据质量

数据是建设科学结论的基石,但如果数据存在质量问题,未事先对数据本身的质量进行评估、核查就直接用于各种计算,可能会存在风险,影响结论的科学性,即所谓“假数真算”。因此拿到数据后,首要的是检查数据质量。

高质量数据是指所收集的数据达到以下要求,数据完整、真实和一致。所以我们在进行数据评估时,就是从数据的完整性、真实性和一致性等三方面进

行考察。

数据真实性是指数据反映真实情况,不能错报人口信息,如普查中的年龄,确实是被调查者的真实年龄,并且还要注意虚岁年龄、完全年龄(或称实足年龄、周岁年龄)和确切年龄的差别。数据真实性可通过变量的有效取值范围以及与相关变量的交叉汇总比较进行判断。

数据完整性是指调查个体记录完整、变量完整、值完整。记录完整是指调查对象按照要求全部统计到位,不能遗漏个体;变量完整是指数据中包含了研究主题相关的全部重要变量,这是研究设计阶段的问题,而不是得到数据后再考虑的问题;值完整是指调查对象的全部属性信息按照填表要求如实填答,不能存在变量遗漏,不符合的项目应使用规定数字或符号填涂。

数据一致性是指数据之间符合逻辑,不矛盾。如在人口数据中,6周岁的儿童其受教育状况不可能是大学,这种数据间的矛盾是我们进行数据评估的一个重要方面,我们可以通过项目之间的逻辑审查发现错误并改正。

(2)影响数据质量的因素

首先是数据收集方式本身的缺陷,这种缺陷又称“结构性缺陷”,是由调查或方案设计不科学引起的,其所产生的误差称为“结构性误差”。如生育率调查,作为一种回顾性调查,在调查设计上存在着两个先天性限制,一是调查对象只能是在调查时点仍存活的人口,如果这部分人口的婚育史与调查时点前死亡人口的情况不同,按照当前存活人口状况推断的人口参数会存在较大误差,这称为选择性误差;二是调查对象只限于一定年龄以下的人口,调查时超过调查年龄的人口不在调查范围内,所以不能完整反映各个时期人口生育状况,这也在不同程度上影响了对时期生育率、婴儿死亡率估计的准确性,这称为截断性误差。

普查的结构性缺陷是由于普查范围广,工作量大,涉及人员多,调查时间长等因素所致。如果在操作过程中监督不力、弄虚作假,结构性缺陷是很难及时发现的。这一问题的解决方法是在普查过程中同时进行部分人口的普查长表调查,将这部分抽样调查数据与普查总体数据进行比对,判断数据可靠性。这需要各级政府、统计部门、普查员和普查对象共同努力,增强统计数据的质量意识,做到依法普查、科学普查;同时做好普查宣传,提高个人如实填报人口信息的意识。

经常性人口统计在理论上不存在结构性误差,但其数据质量受到人口统计网络的完整程度、相关组织和人员的工作效率等影响,使得经常性人口统计数据残缺不全,甚至无法正常使用。

其次,收集数据过程中的每一环节都可能产生影响。一个调查过程可以划分为几个环节,调查设计、人员培训和后勤准备、实地调查、数据回收并编码录入、汇总打印结果等,任何一个环节的疏忽大意都可能产生致命错误。如:

调查设计——调查设计过程中,对调查对象认识不清,抽样框不完整,遗漏了部分总体单位,致使汇总数据有偏;有关调查项目的概念不清或调查概念的具体化较差,违背了调查设计过程中问题设计的唯一性、客观性、非侵犯性和避免歧义等要求,使得调查较难展开。

人员培训和后勤准备——给调查人员的说明中没有详细刻画被调查人口的定义,使调查人员不能正确选择调查对象;调查区域的地图或住址名单不完整;缺乏足够的交通工具,增大了调查人员的工作难度,造成部分调查对象被遗漏。

实地调查——调查人员责任心不强,疏忽大意,或缺乏经验,不能正确提问;调查中出现意外问题得不到及时指导,缺乏有效监督措施;调查环境不利于真实回答,被调查者不肯合作等都会影响数据质量。所以,调查人员的经验、责任心和公正的态度以及调查督导人员的工作水平和态度,是保证数据质量的一个重要条件。

数据回收并编码录入——资料在转移过程中遗失,编码不统一,录入过程操作错误等,会影响数据质量。所以随着现代计算机的普及提高,应该统一编码、联网录入,及时比对,减少错误的发生。

数据汇总——汇总过程中,方法采用不当,计算的人口统计指标错误。如计算结构相对数,一定要注意研究目的以及比较的总体是什么,不要错误地计算。

因此,在进行人口统计分析之前,必须对数据质量进行评价和调整,以使数据达到质量要求。

(3) 误差模型

数据误差可以分为抽样误差和非抽样误差两类。其中抽样误差是由于样本的随机性造成的,使得样本统计量与总体参数之间存在着一定程度的差异,这种差异随着样本规模和抽样方式、方法的不同而变化。非抽样误差起源于数据收集过程中的其他因素,如调查员、数据录入员和数据分析人员的选择及被调查者回答的变异等,具体分为结构性误差、覆盖面误差、申报内容误差和数据录入误差。具体内容请参见相关参考文献,在此不再赘述。现仅对误差的模型分解进行讲解。

在抽样调查中,经常利用样本统计量作为总体参数的估计量,用样本值作

为参数的估计值。样本统计量作为一个随机变量,与总体待估参数之间存在离差,这种离差或大或小,或正或负,体现为一种随机分布,我们把描述这种离差分布状况的统计量称为估计量方差。估计量方差是从平均的意义上说明估计值与待估参数的差异程度,它随样本量的增大而减少。

估计量方差: $V(\hat{\theta}) = E[\hat{\theta} - E(\hat{\theta})]^2$

偏差 $B(\hat{\theta})$ 是指按照某一抽样方案多次抽样,估计值的数学期望 $E(\hat{\theta})$ 与待估参数 θ 之间的离差,其表达式为: $B(\hat{\theta}) = E(\hat{\theta}) - \theta$ 。

均方误差 MSE 是指所有可能的估计值与待估参数之间离差平方的均值,其与估计量方差 $V(\hat{\theta})$ 和偏差平方间存在如下关系: $MSE = E[(\hat{\theta} - \theta)^2] = V(\hat{\theta}) + [B(\hat{\theta})]^2$, 其中 $\hat{\theta}$ 为总体参数 θ 的估计。

估计量方差实际测度的是样本估计值与同一条件下所有可能样本平均值之间的差距,这种方差由抽样方差和非抽样方差构成,其中抽样方差源于抽样调查过程中样本的随机性所致,即用抽选的一个样本来代表总体使得样本统计量与总体参数间存在差异,抽样方差通常会随着样本量的增加而缩小。非抽样方差的特性恰好与抽样方差相反,随样本增加而增大,其产生主要可归为:①由单位的重复、遗漏和不回答所引起的收敛误差(coverage error)。②由访查者和受访者自身的粗心、失误或不当的抽样所引起的内容误差(content error)。在抽样调查过程中,主要是通过提高样本的代表性,降低抽样方差;增强调查过程的控制,减少非抽样方差,最终提高样本统计量估计参数的信度。

从以上均方误差、偏差平方、估计量方差及其分解的关系式中,我们可以发现降低均方误差的方法,可以从降低估计量方差和偏差等方面入手,提高数据估计的准确度。

(4) 漏报和错报

人口数据错误的来源,有两种途径:漏报和错报。

漏报又称总量误差(completeness errors),是指部分调查统计对象的信息遗失或被重复。错报又称内容误差(content errors),是指统计调查对象的信息被歪曲,主要是由于被调查者记忆不清或时间混淆所致。在人口统计数据的质量评估中,区别数据错误形式较难,因为同一现象既可能存在漏报,又可能存在错报,或两者结合。但区分这两种错误形式的意义重大:

①不同的人口分析技术对不同形式错误的敏感程度不同,为了改进人口统计的数据质量,有必要知道哪种形式的错误为主,以选择正确的数据调整方法。

②调整数据时需要对数据的错误形式、错误大小及影响做出推断,以确定数据调整的方向和力度。

在普查数据汇总中,应深入分析漏报的产生原因。随着经济社会的发展,漏报的原因是多方面的:①人口流动性增大。流动人口的居住地具有临时性、分散性,相当一部分人的居住地还具有隐蔽性。2010年普查结果显示我国的流动人口(居住地与户口登记地所在的乡镇街道不一致且离开户口登记地半年以上的,不包括市辖区内人户分离的人口)是2.21亿人,这样大规模的流动人口是人口漏报的重点人群。②人户分离现象增多。在同一地区或城市内拥有两套及以上住房,人员实际居住地与户口所在地不一致,结果两个小区的普查员都以为户主在对方的小区登记,造成人员的漏登。③各种人员或单位为了自身利益故意瞒报。如业主逃避所雇职工登记后缴纳各种社会保险以及相关税费,计划外生育人员担心经济处罚等。④个人隐私保护意识增强。很多人不愿把个人情况,如房屋状况、婚姻状况、生育状况、迁移状况等告之外人,不愿意参加人口普查登记。人口漏报为普查组织者提出了新的问题,针对不同的漏报原因,应采取相应的应对措施减少漏报所造成的影响,同时也应从普查设计、实际组织实施和加大普查宣传等方面入手,从普查事前、事中的各个环节全面保证普查的登记率,减少漏报所造成的数据质量下降。

错报主要是由于被调查者记忆不清或时间混淆所致,这就要求在设计调查项目时应明确时间界限,缩短时间间隔,避免采用模糊的词汇设计题目或选项。在分析错报时,还应注意毛误差和净误差的问题。毛误差是所有错误的总和,是以被调查者个人为单位计算的,对指标内部关联的分析影响较大,如分年龄生育率或死亡率的估计等等。净误差是某些错误相互抵消(漏报和重复)后的余差,是以被调查人口的某种集合计算的,将对总量分析影响较大。

(5)统计数据质量控制

对人口普查数据质量作出科学评估是人口普查整体工作的重要组成部分。科学严格的事后质量抽查只是一种事后补救措施,我们应树立从源头、全程控制数据质量的意识。

统计数据质量控制的全程性,要求数据质量控制应从统计调查设计、数据收集、统计汇总与报告、统计监督等各个环节入手,落实专人负责制,全面提高数据质量;要求普查组织者、普查员应该提高人口普查的质量意识,媒体及宣传部门等应加大普查的宣传力度,提高公众积极申报人口信息的意识,要“如实申报普查资料,真实反映基本国情”。

①统计设计阶段的质量控制。统计设计是统计工作的首要环节,其设计的好坏直接决定数据质量的高低。设计阶段应从调查项目的内涵、外延及其具体操作性问题的分解、统计指标体系的涵义解释与计算方法等方面准确阐述,不

能存在歧义,要便于理解和收集信息,并规定普查表的填写方法和数据审查方法。

②资料整理鉴别阶段的质量控制。统计资料整理鉴别阶段出现的差错是统计数据质量问题的重要方面,要特别注意审查资料的可靠性和适用性,要弄清楚统计指标的统计范围、计算方法等。对于统计口径不一致或不完整的资料,需要进行调整、补充;对于相互比较的资料,必须注意其可比性;一旦发现数据有严重的质量问题,应及时核实,避免有质量问题的资料进入汇总处理阶段。

③人为错误的质量控制。通过统计业务培训,全面提高统计人员的业务素质,要严格杜绝因业务不熟悉而造成的数据质量问题。加大统计信息化建设的支持力度,通过先进的统计技术手段,达到统计数据“快、精、准”的高质量标准,并做到数据反馈的“准确性和及时性”。

第二节 人口数据评估

本节所关注的是数据在整理汇总之后,对公布数据的质量评估问题。在讨论中,对前期准备阶段和调查阶段所发生的影响数据质量的问题不再追究考察,假设其在工作过程中都是严格按照普查条例和具体指导进行的。

1 数据评估的一般原则

数据分析应该从数据评估开始,因为准确的数据是获得科学结论的必要条件。如果不对数据资料进行检验评估,很可能会得出与实际情况相差很大的结论。人口数据是提供人口决策的重要基础,在对人口数据进行分析 and 整理时,要对人口数据抱有审慎的态度加以探索,同时需要把握数据评估的一般原则。

(1)效度(Validity)。效度指测量工具或手段能够准确测出所需要测量事物的程度。效度能够反映测量某一现象或观念的标识是否得当,体现客观现实的反映程度,因而成为衡量数据质量的首要标准。

(2)信度(Reliability)。信度指在同样条件下,重复测度某现象或观念而获得的结果的一致性。信度是衡量数据质量的指标,它可以通过重复测度,在结果的一致性上加以判断。从统计学的角度而言,信度反映的是数据的稳定性,即排除随机误差干扰的数据质量。

(3)准确度(Accuracy)。准确度是指调查值与真实值之间的差距。信度只能反映观察值中随机扰动的成分,准确度不但包括随机扰动,而且包括非随机的偏差成分。因此准确度是效度的必要条件,不是充分条件。一个测量的效度高,其准确度必然高;信度高时,准确度并不一定高。

2 数据评估的一般方法

在获得人口普查资料后,需要对数据进行评估和检验,以确保普查资料的可靠性和准确性。数据的准确性将直接影响到后期科学研究分析的结论。一般而言,数据评估的方法主要包括四种:

(1)利用人口本身的逻辑关系进行判断

人口变动受自然规律和社会规律的共同制约,人口数量也在两大规律的共同影响下表现一定的特征,即在一定的稳定水平下呈现一定范围的波动,或者沿着一个方向缓慢变动。人口变动的规律性给予我们判断人口数据的逻辑关系,反映人口自然特征的指标可以通过该指标的合理性范围进行检验。例如出生婴儿性别比数据一般比较稳定,平均每出生 100 名女性婴儿,便相应应有 103—107 名男性婴儿。如果一个地区的出生婴儿性别比大于 107,或低于 103,那么就要进一步分析,是由于登记过程中存在误差,还是实际情况如此。另一种情况是出生性别比一直沿着一个方向缓慢变化,但某个年份突然大幅度升高或降低,如果没有合理的解释,那很可能是数据质量出现问题。

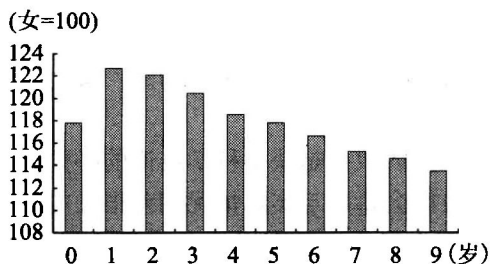


图 1-1 2000 年中国 0-9 岁分年龄性别比

资料来源:中国 2000 年全国人口普查资料。

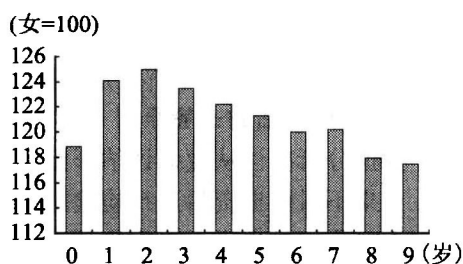


图 1-2 2005 年中国 0-9 岁分年龄性别比

资料来源:2005 年全国 1% 人口抽样调查资料。

如对 2000 年第五次人口普查资料统计的全国 0-9 岁人口分年龄的性别比(见图 1-1)与 2005 年全国 1% 人口抽样调查资料统计的全国 0-9 岁人口分年龄性别比(见图 1-2)的分析可以发现,两次调查都显示了同一个规律:0 岁人口性别比明显低于 1 至 4 岁人口各年龄组的性别比,2000 年全国 0 岁年龄组性别比与 1 岁年龄组性别比相比低 4.9,2005 年这个差距扩大到 5.2。

0 岁组性别比突然性的降低,显然不太可能。因此可以判断,在人口普查和人口抽样调查这样的调查系统中,相对于其它年龄组,0 岁人口的统计中存在明显孩子漏报,包括大量男孩漏报或瞒报的现象,使 0 岁人口组的性别比与

其他年龄组相比较显得异常的低。这可能是近些年来中国出生性别比持续偏高的态势引起了各方关注,0岁组的性别比成为一个备受瞩目的指标,所以存在0岁人口数据瞒报的情况。

(2) 人口数据相关指标的一致性检验

人口数据存在内在规律,这表现为人口指标之间存在着一定的数量关系。通过对调查资料的内部核查,可以评估人口数据的可靠性。例如,通过比较男女两性人口的年龄分布曲线,可以检验年龄申报的准确性。一般来说,当人口数据准确时,男性与女性人口的年龄分布曲线呈现吻合特征,曲线的形状十分相似。我国第五次人口普查得到的男性与女性人口的年龄分布曲线差异较小,说明人口年龄状况资料的质量较高(见图1-3)。对于在数据检验中发现的不合理处,如果不能给出合理解释就要对数据本身的质量提出质疑。

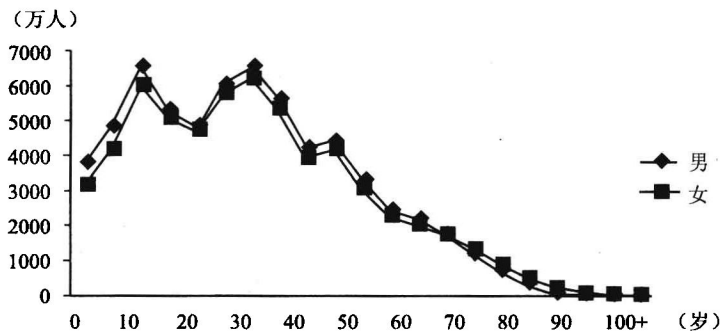


图 1-3 第五次人口普查中国男性和女性分年龄组人口数分布

资料来源:中国 2000 年全国人口普查资料。

(3) 根据两次或多次普查间的人口变动状况和规律检查数据的可靠性

我国曾于 1953、1964、1982、1990、2000、2010 年进行过六次人口普查。这种方法利用人口变动的规律对普查资料的数据进行检验。需要注意的是,这种对比检验需要依据一定的原则:第一,用来比较的数据的质量必须是可靠的,并且经过检验。如果用甲材料来验证乙材料的真实性,需要假定甲材料是真实准确的,如果用前次普查资料对本次普查资料进行检验,则需要假定两次普查间的一些指标的变化具有一定的规律。第二,从多种途径选择对照来源,并且尽量避免对比数据的相关性。第三,保证对照资料的统计口径和统计范围的一致性。具体而言,有两种方法进行检验。一种是利用人口变化规律和前次的普查资料推断本次调查的结果,将实际数据与模拟数据构造对比。另一种是根据最