财政部规划教材
普通高等学校“十三五”规划教材

统计分析方法与结果解读

——基于CSSCI期刊14篇论文精讲

邰秀军 郭泽英 等\编著

中国财经出版传媒集团
中国财政经济出版社

山西省研究生教育改革课题（2019JG121）和山西师范大学研究生双语课程建设项目（YJSSY201805）和山西师范大学优质课程（2017YZKC -19）的联合资助

统计分析方法与结果解读：基于 CSSCI 期刊 14 篇论文精讲

郜秀军　郭泽英　等\编著

中国财经出版传媒集团
中国财政经济出版社

图书在版编目（CIP）数据

统计分析方法与结果解读：基于 CSSCI 期刊 14 篇论文精讲 / 邰秀军等编著. —北京：中国财政经济出版社，2019. 11

ISBN 978 - 7 - 5095 - 9312 - 7

Ⅰ. ①统… Ⅱ. ①邰… Ⅲ. ①统计分析 - 分析方法 - 应用 - 论文 - 写作 - 研究 Ⅳ. ①H152. 3

中国版本图书馆 CIP 数据核字（2019）第 232455 号

责任编辑：彭 波　　责任印制：史大鹏

封面设计：卜建辰　　责任校对：李 丽

中国财政经济出版社 出版

URL：http：//www. cfeph. cn

E - mail：cfeph @ cfemg. cn

社址：北京市海淀区阜成路甲 28 号 邮政编码：100142

营销中心电话：010 - 88191537

北京财经印刷厂印装 各地新华书店经销

787 × 1092 毫米 16 开 13. 25 印张 264 000 字

2020 年 7 月第 1 版 2020 年 7 月北京第 1 次印刷

定价：58. 00 元

ISBN 978 - 7 - 5095 - 9312 - 7

序言
PREFACE

经过近二十多年的发展，使用统计实证方法进行数据处理已成为国内社会科学学术领域开展学术研究的重要工具和方法之一。因其便于操作和模仿，许多研究生在论文写作中也将统计实证方法作为论文研究的首选方法。然而，在笔者十数年辅导硕士研究生论文写作中仍发现许多研究生的论文中在选择统计方法、统计结果解读以及论文撰写过程中存在着许多突出的问题，如统计方法选择不当、计算方法错误、对结果理解不正确等。由于许多论文并不需要答辩，或者是由于答辩教师将主要精力放在理论模型建构或是创新点等高层次的论文问题上，并没有特别关注这些看起来是“细枝末节”的统计问题，很多问题便被掩盖起来，致使一些论文结论建立在不恰当甚至是错误的统计方法之上。

近几年来，部分学者已经关注到这类统计学方法滥用并将其称为“伪科学研究”。受学术研究中统计实证潮流的影响，许多研究生在论文写作中，也喜欢模仿学者已发表的论文进行自己的创作，然而，单纯机械模仿，在统计方法上食古不化、食洋不化的例子很多。笔者在近几年研究生论文评审、学术会议点评上见到过很多这样的例子，当我问学生为什么在论文中选用这样的方法或者同他们探讨一些细节的统计问题时，大部分学生都茫然无措，或者简单回答说“已有文献都是这样处理的等等”搪塞之语。究其原因是因为目前部分期刊论文篇幅已经较长了，但在交代论文中的统计过程时，仍然略去了很多细节，学生在模仿学者们发表的论文时，有些被省略的步骤就不能重复，又不能见到本人去问，只能自己想当然的理解，久而久之，一些细节问题在学术界长期积累了下来，反而成为大家认为正确的东西。另外，一些期刊注重观点和理论，在统计方法上审稿不严，论文中的统计方法虽有问

题，但被视为无伤大雅，或者由于统计方法复杂，审稿人或编辑也没有能完全搞明白等原因，致使错误的统计方法以讹传讹。

“中文社会科学引文索引”（CSSCI）是国家、教育部重点课题攻关项目。CSSCI 遵循文献计量学规律，采取定量与定性评价相结合的方法从全国 2700 余种中文人文社会科学学术性期刊中精选出学术性强、编辑规范的期刊作为来源期刊。近十年来，由于期刊竞争激烈，各编辑部加大了匿名评审的力度，论文质量越来越多获得国内学术界认可，在很多学校成为评判论文级别的重要标准。

若用 CSSCI 期刊论文为例，分析其中涉及的统计方法，在场景学习里使学生更易理解其内涵、原理及使用方法。同时也能促使学生思维不断深化，并使学生对统计方法的运用上培养创造性思维能力。这种方式还能使硕士研究生较快了解科研思路、范式及论文的写作规范和格式，同时，避免学生将错误的东西当作真理来理解。

本书就是试图建立“以 CSSCI 论文案例理解统计学方法”的教学方法、思维模式，并最终提高硕士研究生科研能力的教学体系。包括形成使硕士研究生能容易理解的统计方法参考书、教案、讲授方法、实践操作方法等，让学生通过此参考书的学习后，以案说法，掌握基本的统计学思维和理论知识，会应用统计方法解决实际问题的能力；学会如何提出问题，如何进行数据操作，如何判断数据质量，如何评价模型和方法的有效性，以及如何准确清晰地呈现分析结论。培养硕士研究生定量的分析问题能力和解决实际问题的能力，在硕士研究生一年级就打牢研究生的科研能力。最后的汇编报告给硕士研究生提供可供借鉴参考的书目。

本书共十四讲，包括 14 种常用的统计方法，是笔者在长期的研究生统计教学中积累下的资料。其中，第一讲由山西师范大学经济与管理学院研究生杨鑫撰写；第二讲由山西师范大学经济与管理学院研究生杨瑞霞撰写；第三讲由山西师范大学经济与管理学院研究生张维亮撰写；第四讲由山西师范大学经济与管理学院研究生郭金生撰写；第五讲由山西师范大学经济与管理学院研究生邢玉凤撰写；第六讲由山西师范大学经济与管理学院研究生王彦君撰写；第七讲由山西师范大学经济与管理学院研究生储玲林撰写；第八讲由山西师范大学经济与管理学院研究生王彬撰写；第九讲由山西师范大学经济与管理学院研究生马睿撰写；第十讲由山西师范大学经济与管理学院研究生

蒲则文撰写；第十一讲由山西师范大学经济与管理学院研究生朱桦燕撰写；第十二讲由山西师范大学经济与管理学院研究生张晋旗撰写；第十三讲由山西师范大学经济与管理学院研究生冯玉琪撰写；第十四讲由山西师范大学经济与管理学院研究生侯亚静撰写。笔者负责了全书的审阅与定稿。

由于是初次尝试，可能存在许多错误和不足，恳请读者给予指出，我将在以后的教学实践中进行修正。

郜秀军

2019 年 7 月

目录
CONTENTS

第一讲 描述性统计方法

第一节 方法介绍

描述性统计，是指运用制表和分类、图形以及计算概括性数据来描述数据特征的各项活动。描述性统计分析要对调查总体所有变量的有关数据进行统计性描述，主要包括数据的频数分析、集中趋势分析、离散程度分析、分布以及一些基本的统计图形。数据的频数分析是指在数据的预处理部分测量数据出现的次数，利用频数分析和交叉频数分析可以检验数据中的异常值；数据的集中趋势分析，用来反映数据的一般水平，常用的指标有平均值、中位数和众数等；数据的离散程度分析主要是用来反映数据之间的差异程度，常用的指标有方差和标准差；数据的分布是指在统计分析中，通常要假设样本所属总体的分布属于正态分布，因此需要用偏度和峰度两个指标来检查样本数据是否符合正态分布；用图形的形式来表达数据，比用文字表达更清晰、更简明。在统计学软件里，可以很容易地绘制各个变量的统计图形，包括条形图、饼图和折线图等以及表现不同变量关系的图形。

一、频数分析

频数，又称“次数”。指变量值中代表某种特征的数（标志值）出现的次数。按分组依次排列的频数构成频数数列，用来说明各组标志值对全体标志值所起作用的强度。各组频数的总和等于总体的全部单位数。频数的表示方法，既可以用表的形式，也可以用图形的形式。

二、集中趋势分析

平均数，统计学术语，是表示一组数据集中趋势的量数，是指在一组数据中所有数

据之和再除以这组数据的个数。在统计工作中，平均数（均值）和标准差是描述数据资料集中趋势和离散程度的两个最重要的测度值。平均数是统计学中最常用的统计量，用来表明资料中各观测值相对集中较多的中心位置，在生产实践和科学研究中，平均数被广泛用来描述或比较各种技术措施的效果、某些数量性状的指标等。

中位数，统计学中的专有名词，代表一个样本、种群或概率分布中的一个数值，其可将数值集合划分为相等的上下两部分。对于有限的数集，可以通过把所有观察值高低排序后找出正中间的一个作为中位数。如果观察值有偶数个，通常取最中间的两个数的平均数作为中位数。

众数是统计学名词，在统计分布上具有明显集中趋势点的数值，代表数据的一般水平。众数是一组数据中出现次数最多的数值，有时众数在一组数中有好几个，用 M 表示。简单地说，就是一组数据中占比例最多的那个数。众数是样本观测值在频数分布表中频数最多的那一组的组中值，主要应用于大面积普查研究之中。众数是在一组数据中出现次数最多的数据，是一组数据中的原数据，而不是相应的次数。

三、离散程度分析

方差是衡量随机变量或一组数据时离散程度的度量。概率论中方差用来度量随机变量和其数学期望（即均值）之间的偏离程度。统计学中的方差（样本方差）是每个样本值与全体样本值的平均数之差的平方值的平均数。在统计描述中，方差用来计算每一个变量（观察值）与总体均数之间的差异。为避免出现离均差总和为零，离均差平方和受样本量的影响，统计学采用平均离均差平方和来描述变量的变异程度。在许多实际问题中，研究方差即偏离程度有着重要意义。

标准差，又常称均方差，是离均差平方的算术平均数的平方根，用 σ 表示。标准差是方差的算术平方根，在概率统计中最常用作统计分布程度上的测量，能反映一个数据集的离散程度。平均数相同的两组数据，标准差未必相同。

四、偏度与峰度分析

偏度也称为偏态、偏态系数，是统计数据分布偏斜方向和程度的度量，是统计数据分布非对称程度的数字特征。正态分布的偏度为 0，两侧尾部长度对称。以 bs 表示偏度，$bs < 0$ 称分布具有负偏离，也称左偏态，此时数据位于均值左边的比位于右边的少，直观表现为左边的尾部相对于右边的尾部要长，因为有少数变量值很小，使曲线左侧尾部拖得很长；$bs > 0$ 称分布具有正偏离，也称右偏态，此时数据位于均值右边的比位于左边的少，直观表现为右边的尾部相对于左边的尾部要长，因为有少数变量值很大，使

曲线右侧尾部拖得很长；而 bs 接近 0 则可认为分布是对称的。若知道分布有可能在偏度上偏离正态分布时，可用偏离来检验分布的正态性。右偏时一般算术平均数 > 中位数 > 众数，左偏时相反，即众数 > 中位数 > 平均数。正态分布三者相等。

峰度又称峰态系数。表征概率密度分布曲线在平均值处峰值高低的特征数。直观看来，峰度反映了峰部的尖度。在统计学中，峰度衡量实数随机变量概率分布的峰态。峰度高就意味着方差增大是由低频度的大于或小于平均值的极端差值引起的。峰度以 bk 表示，正态分布的峰度为 3。一般而言，以正态分布为参照，峰度可以描述分布形态的陡缓程度，若 bk < 3，则称分布具有不足的峰度；若 bk > 3，则称分布具有过度的峰度；若知道分布有可能在峰度上偏离正态分布时，可用峰度来检验分布的正态性。

第二节　方法的应用

一、论文主题及主要解决的问题

本章的描述性统计方法应用将以《农户家庭成员职业选择及影响因素分析》论文为例进行讲解。该论文发表于《管理世界》2007 年第 7 期，是基于 2003 ~ 2005 年苏浙沪两省一市 15 村固定跟踪观察点办公室调查的农户资料，对长三角农户家庭成员职业选择及外出活动等进行分析，了解长三角地区农村劳动力就业的现实选择，进而深入剖析非农化进程中影响农民职业的选择和流动的主要因素，为加速我国非农化进程提供相关对策和建议。

论文的背景从现实情况的分析中提出来。改革开放以来，以上海为中心的长三角地区经济迅速崛起，越来越引起世人的关注。2005 年，长三角地区就以占全国 1% 的土地和 6% 的人口，创造了占全国 18.52% 的 GDP、21.43% 的地方财政收入和 41.27% 的国外直接投资。长三角地区的经济发展，不仅体现在拥有集聚效应的城市与企业上，也体现在农村与农户的变化上。20 世纪 90 年代中后期，当全国农村经济普遍处于徘徊或下降状态之际，该地区的农村与农户经济则率先走出徘徊，迅速实现家庭经济结构转换，走上了农村经济增长的“快车道”。这对于由于农村工业化的发展以及农村人口的增长，人地矛盾日益趋紧的长三角地区简直是奇迹。

二、思路与步骤

目前，在有关农村就业问题的研究中，国外学者的研究视角较为多元，国内学者对

如何实现农村劳动力有效和有序地转移给予了广泛关注，而有关农村劳动力市场发展，特别是有关其内部职业选择及其影响因素的研究文献尚不多见。对影响农户家庭从事非农就业的因素进行实证研究的文献也相对少见。结合已有研究和现状分析，发现活跃的劳动力市场和相对较合理的农村就业结构是一个非常重要的原因。

论文先介绍了文中使用的资料来源及样本信息，再利用调查到的资料，通过使用描述性统计方法进行比较分析，从农户家庭劳动力成员的性别、年龄、所在区域、受教育时间以及调查年份的职业分布、行业分布、就业方向的选择和就业时间的分布进行不同侧面的比较分析，再基于上述分析结果从农户家庭劳动力成员的外出就业收入水平、外出就业活动圈两个因素进行分析，最后得出结论，并对此提出政策性建议。

三、结果解读

（一）农户家庭有劳动能力成员的性别、年龄、所在区域等的分布状况

1. 职业分布。

论文将农户家庭有劳动能力的成员从事的职业分为 8 种，分别是家庭经营农业劳动者、家庭经营非农业劳动者、受雇劳动者、个体合伙工商劳动经营者、私营企业经营者、乡村及国家干部、教育科技医疗卫生和文化艺术工作者、其他职业者。具体如表 1－1所示。

表 1－1　　长三角 15 村农户家庭劳动者从事职业的分布　　单位：%

	性别		区域			年份			总体
	男性	女性	上海	江苏	浙江	2003	2004	2005	
1	20.51	33.24	18.25	38.17	24.24	28.57	25.76	25.39	26.61
2	9.09	6.48	2.64	11.19	11.84	6.65	7.75	9.20	7.84
3	49.12	41.92	62.17	31.85	37.52	44.70	46.72	45.63	45.67
4	2.75	2.90	2.08	1.40	6.11	3.21	2.98	2.26	2.82
5	1.35	0.93	0.72	0.70	2.50	1.21	1.18	1.05	1.15
6	2.80	0.87	2.21	1.91	1.28	1.84	1.71	2.09	1.88
7	1.90	1.07	1.68	1.75	0.83	1.49	1.55	1.46	1.50
8	12.49	12.57	10.25	13.02	15.68	12.33	12.35	12.92	12.53

续表

	性别		区域			年份			总体
	男性	女性	上海	江苏	浙江	2003	2004	2005	
年龄	17－18 岁	19－25 岁	26－35 岁	36－45 岁	46－55 岁	56－65 岁	66－75 岁	76 岁＋	
1	8.33	5.94	7.07	17.55	32.50	44.33	49.13	24.66	
2	8.33	11.42	9.21	11.16	7.39	4.18	4.78	2.05	
3	50.00	64.84	65.69	56.39	43.18	30.15	8.91	4.11	
4	0.00	1.37	4.13	4.63	2.86	0.97	1.09	0.00	
5	0.00	0.46	1.11	1.36	1.63	0.82	0.22	0.00	
6	0.00	0.23	0.95	1.84	3.30	1.19	0.87	2.74	
7	0.00	2.05	1.67	0.75	1.45	2.69	0.22	0.00	
8	33.33	13.70	10.17	6.33	7.70	15.67	34.78	66.44	

注：表中职业编号分别为："1" 家庭经营农业劳动者；"2" 家庭经营非农业劳动者；"3" 受雇劳动者；"4" 个体合伙工商劳动经营者；"5" 私营企业经营者；"6" 乡村及国家干部；"7" 教育科技医疗卫生和文化艺术工作者；"8" 是其他。

从表 1－1 可以看出，长三角 15 村农户家庭有劳动能力的成员中，从事最多的职业为"受雇劳动者"，比例高达 45.67%；第二大职业是"家庭经营劳动者"，比例为 34.45%，其中农业劳动者为 26.61%，非农业劳动者为 7.84%；第三大职业是"企业经营者"，比例为 3.97%，其中个体合伙工商劳动经营者为 2.82%；第四大职业是"乡村及国家干部"，比例为 1.88%；第五大职业为"教育科技医疗卫生和文化艺术工作者"，比例为 1.50%，从事其他职业的比例为 12.53%。

从区域差异比较看，在两大职业选择上，上海 5 村劳动者从事的职业主体是受雇劳动者，比例高达 62.17%，而从事家庭经营的比例只有 20.89%；江苏 5 村劳动者从事的职业主体则是家庭经营，比例为 49.36%；在浙江 5 村受雇劳动者的就业比例为 37.52%，从事家庭经营就业的比例也高达 36.08%。

从职业选择与年龄的关系看，青年人选择的职业以"受雇劳动者"和"家庭非农业经营"为主，中老年人选择的职业则以"家庭农业经营"为主。中年人在"企业经营者"方面具有很高的就业机会。

从职业选择与受教育情况看（见图 1－1），随着受教育时间的加长，家庭经营特别是农业经营是放弃选择的职业，而教育科学文化卫生等是期盼选择的职业，对于"受雇劳动者"的选择与教育的关系表现为一种倒"U"形上升关系，这一趋势表明，从事受雇劳动者也需要一定的文化程度做基础，但当受教育达到一定程度后，"受雇劳动者"这一职业由受欢迎转为弃选职业。

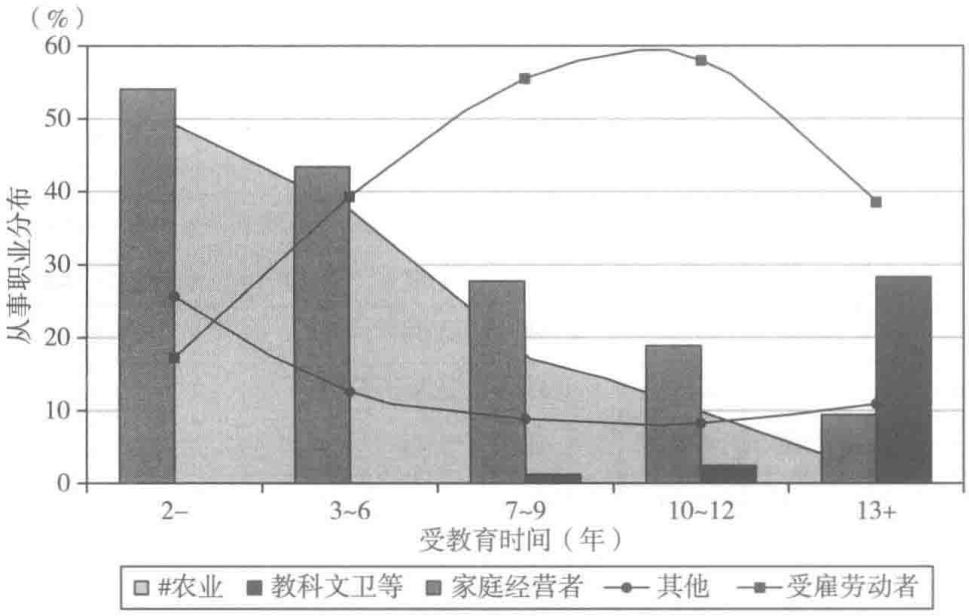

图 1-1　长三角 15 村农户家庭成员职业选择与教育的关系

2. 行业分布。

论文分析了长三角地区 15 村农户家庭劳动者从事的行业分布，具体如表 1-2 所示。从表 1-2 中可以看出，第一大行业为工业占 39.80%，第二大行业为农业占 26.67%。从从事行业的性别比较看，农业和服务业女性优于男性，工业、建筑业和运输业则是男性优于女性。从区域分布看，上海 5 村农户家庭成员从事的主体行业为工业，其比例占 49.73%，从事农业的比例只有 19.77%；江苏 5 村农户家庭成员从事的行业是农业与工业并重，其比例分别为 35.96% 和 34.21%；浙江 5 村农户家庭成员从事的行业主体是工业，只有 31.20%，从事农业的比例为 24.96%。

表 1-2　　长三角 15 村农户家庭劳动者从事行业的分布　　单位：%

行业	性别		区域		年份				总体
	男性	女性	上海	江苏	浙江	2003	2004	2005	
农业	20.53	33.34	19.77	35.96	24.96	28.31	25.88	25.72	26.67
工业	41.37	38.11	49.73	34.21	31.20	38.51	41.00	39.97	39.80
建筑	4.34	0.74	2.07	3.75	1.89	2.78	2.38	2.66	2.61
运输	4.34	0.91	1.30	2.85	4.79	2.67	2.79	2.61	2.69
服务	7.18	8.00	4.17	5.70	15.93	7.69	7.89	7.12	7.57
其他	22.25	18.91	22.97	17.53	21.23	20.04	20.05	21.92	20.65

从农户家庭成员从事行业与其年龄及教育程度的关系看（见图 1-2），随着农户家庭成员年龄的上升，从事农业的比例呈明显上升趋势，从事工业的比例呈显著倒“U”形下降趋势，从事服务业的比例呈下降趋势；而随着农户家庭成员受教育时间的加长，从事农业

的比例呈显著下降趋势。与此相对，从事工业、建筑业、服务业的比例呈明显上升趋势。

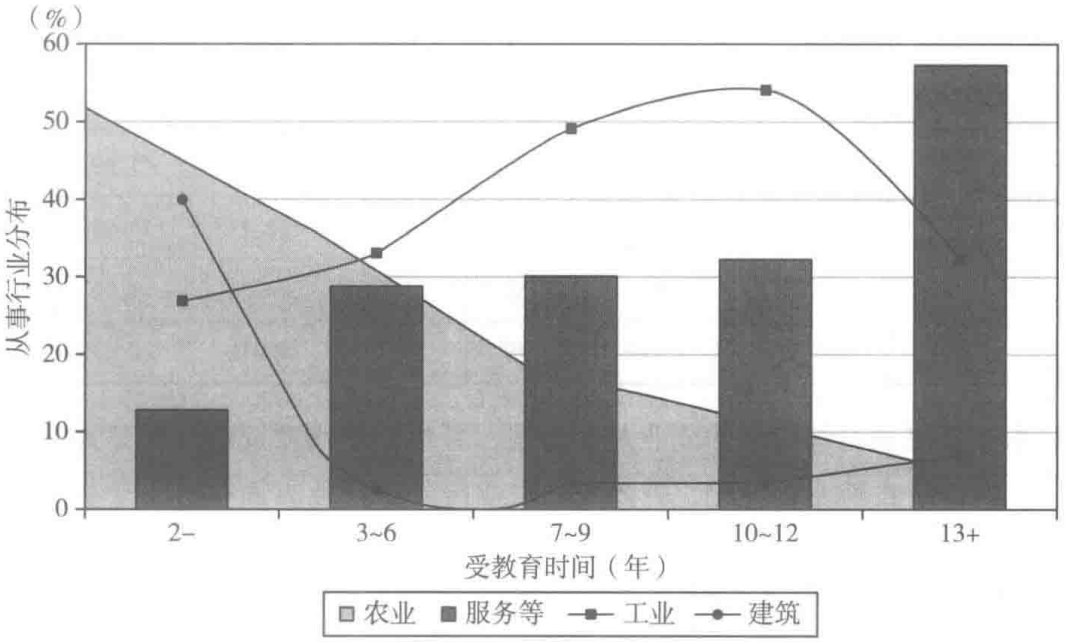

图 1－2　长三角 15 村农户家庭成员从事行业与教育的关系

3. 就业方向的选择。

对长三角 15 村农户家庭劳动成员在就业方向选择上的分析，具体如表 1－3 所示。

表 1－3　　长三角 15 村农户家庭成年成员在村及外出就业

		全年工作（日）	其中：在村内与外出（%）			外出就业（元/日）		
			农业	非农业	外出	收入	支出	净收入
性别	男性	256.70	14.73	30.74	54.52	71.79	13.89	57.91
	女性	223.45	21.20	27.44	51.36	43.55	6.49	37.06
区域	上海 5 村	200.92	11.81	28.92	59.28	48.17	2.01	46.17
	江苏 5 村	254.80	17.37	32.53	50.10	27.00	1.85	25.15
	浙江 5 村	285.46	24.53	25.15	50.32	122.47	35.41	87.05
总体		241.02	17.56	29.30	53.14	59.86	10.76	49.10
年龄	17～18 岁	214.38	2.33	0.00	97.67	22.36	3	19.35
	19～25 岁	269.92	3.67	16.88	79.45	37.19	4.12	33.07
	26～35 岁	275.91	4.77	17.54	77.69	79.52	17.61	61.92
	36～45 岁	273.77	13.02	27.71	59.28	66.70	13.36	53.34
	46～55 岁	234.24	22.79	40.16	37.06	49.61	5.11	44.51
	56～65 岁	198.44	30.37	31.14	38.49	33.60	3.63	29.98
	66～75 岁	156.83	53.23	28.95	17.82	20.40	1.56	18.84
	76 岁＋	65.66	44.28	34.55	21.17	40.35	1.23	39.12