

海洋经济分析评估 理论、方法与实践

何广顺 丁黎黎 宋维玲 编著



HAIYANG JINGJI FENXI PINGGU LILUN
FANGFA YU SHIJIAN



海洋出版社

海洋经济分析评估

理论、方法与实践

何广顺 丁黎黎 宋维玲 编著

海洋出版社

2014 年 · 北京

图书在版编目(CIP)数据

海洋经济分析评估理论、方法与实践 / 何广顺, 丁黎黎, 宋维玲编著. —北京: 海洋出版社, 2014. 5

ISBN 978 - 7 - 5027 - 8867 - 4

I. ①海… II. ①何… ②丁… ③宋… III. ①海洋经济 - 经济分析 - 研究 IV. ①P74

中国版本图书馆 CIP 数据核字(2014)第 095279 号

责任编辑: 高朝君

责任印制: 赵麟苏

海洋出版社 出版发行

<http://www.oceanpress.com.cn>

北京市海淀区大慧寺路 8 号 邮编: 100081

北京旺都印刷有限公司印刷

2014 年 5 月第 1 版 2014 年 5 月北京第 1 次印刷

开本: 787 mm × 1092 mm 1/16 印张: 27.25

字数: 548 千字 定价: 68.00 元

发行部: 62132549 邮购部: 68038093

编辑室: 62100038 总编室: 62114335

海洋版图书印、装错误可随时退换

前 言

海洋经济分析评估是反映海洋经济发展状况的一项基础性工作，也是开展海洋经济管理工作的的重要依据。随着海洋经济的快速发展，社会各界对海洋经济的关注度不断提高，各级海洋管理部门和科研院所对海洋经济进行分析评估的需求也在日益增长。特别是在国家和省级海洋经济运行监测与评估系统建设过程中，需要分析评估哪些内容？用什么方法来进行分析评估？这些现实问题一直困扰着大多数海洋经济工作者。本书从拓展海洋经济研究的深度和广度出发，以宏观经济理论和统计学理论方法为指导，以海洋经济统计资料为基础，对我国海洋经济运行过程、规律以及海洋经济总量、产业结构和布局等进行了较为系统的分析评估，其目的是为全面深刻认识海洋经济发展规律及其特征提供一把开启思路的钥匙。

本书的核心思想是通过实证来考察和检验宏观经济理论和方法在海洋经济领域中的适用性。因此，在每章结构和内容的编排上力求简明易懂，对涉及的各种理论只做简要梳理和阐述，对涉及的分析评估方法只介绍怎样使用，尽量避免数学推导，希望能使读者易学易用。全书共分为十章，除第一章概述和第十章统计分析报告以外，每章节都力求能解决某一类专门问题，因此全书不仅具有一个完整的框架体系，而且各个章节均具有较强的独立性，可以根据实际需要单独使用。

本书是集体智慧与力量的结晶，也是在师友和家人的鼓励、支持和帮助下才得以完成的。在编写过程中，从确定框架结构和写作提纲，到明确编写体例和实际撰写，都经过了大家的充分讨论和交流。其中丁黎黎参与撰写了第三、五、八章，宋维玲参与撰写了第三、六、八章，郑慧参与撰写了第四章，王涛参与撰写了第二、三、八章，李琳琳参与撰写了第四、五、七章，何佳霖参与撰写了第六、七章；全书的实证基本由李琳琳和王涛完成；全书的统稿工作由笔者和宋维玲完成。在本书的写作过程中，我们参考、引用了大量的相关文献，在此向各位作者表示衷心的感谢！所引用的文献虽尽量给予了注明，但恐仍有疏漏，不周之处敬请谅解。

由于我们的学识、能力与水平有限，加之海洋经济分析评估本身又是比较新的研究领域，在某些问题上我们仅仅做了一些有益的探索和尝试，在分析评估的深度和广度上还远远不够，因此我们期待同行专家批评指正，并将致力于进一步的完善。

何广顺

2014年2月于天津

目 录

第一章 绪 论	1
第一节 我国海洋经济发展历程	1
第二节 海洋经济分析内容与作用	2
第三节 海洋经济分析指标	4
第四节 海洋经济分析数据预处理	8
第五节 海洋经济分析方法	17
第二章 海洋经济总量分析	28
第一节 海洋经济总量变动分析	28
第二节 海洋经济总量构成分析	31
第三节 海洋经济对国民经济贡献分析	40
第四节 海洋经济总量预测分析	43
第三章 海洋产业分析评估	56
第一节 海洋产业结构分析	56
第二节 海洋产业关联分析	82
第三节 海洋主导产业的选择	97
第四节 海洋产业布局分析	107
第五节 海洋产业集聚分析	116
第四章 区域海洋经济分析评估	127
第一节 区域海洋经济分析的一般问题	127
第二节 区域海洋经济特征分析	130
第三节 区域海洋产业结构分析	151
第四节 区域海洋经济发展模式分析	159
第五节 区域海洋经济协调发展分析	173
第五章 海岛经济分析	190
第一节 海岛经济发展历程及特征分析	190
第二节 海岛经济发展水平分析	196
第三节 海岛经济发展模式分析	213
第六章 海洋经济增长分析	223
第一节 经济增长分析的一般问题	223

第二节 海洋经济增长因素分析	225
第三节 海洋经济增长质量分析	234
第四节 海洋经济增长方式转变分析	247
第五节 海洋经济可持续发展分析	253
第七章 海洋经济周期分析	264
第一节 经济周期相关理论和方法概述	264
第二节 海洋经济周期的测定与特征分析	272
第三节 海洋经济周期的影响机制分析	280
第四节 海洋经济与国民经济波动的协调性分析	291
第八章 海洋经济监测预警分析	305
第一节 经济监测预警分析的一般问题	305
第二节 基准循环的确定和监测指标的分类方法	310
第三节 海洋经济监测预警实证分析	320
第九章 国际海洋经济发展分析	335
第一节 世界海洋产业发展分析	335
第二节 主要国家海洋经济发展分析	347
第三节 主要国家海洋经济发展政策分析	386
第十章 海洋经济统计分析报告	397
第一节 统计分析报告的特点和分类	397
第二节 统计分析报告的写作步骤	405
第三节 海洋经济统计分析报告编写实例	412

海洋经济分析评估是在海洋经济统计调查的基础上,运用各种分析评估方法对海洋经济运行的数量关系进行研究的实践活动,是经济统计工作的统计设计、统计调查、统计整理和统计分析四个阶段中最后一个阶段。海洋经济分析的实质是以宏观经济理论为指导,以海洋经济统计资料为基础,以统计分析方法为手段,对海洋总量、产业结构、区域布局、经济增长和周期波动等所进行的分析评估。

第一节 我国海洋经济发展历程

全面了解我国海洋经济发展的历程,对开展海洋经济分析会有极大的帮助。纵观我国海洋经济发展的历程,大体经历了初级发展、快速发展和全面发展三个阶段。

一、海洋经济初级发展阶段

从新中国成立到改革开放前,是海洋经济初级发展阶段。这一阶段由于受到整个国家经济环境的影响和生产力发展水平的制约,加之对海洋的认知程度较低,海洋经济发展主要以资源依赖型、劳动密集型、自给自足型的产业为主;海洋产业结构比较单一,以海洋捕捞为主的海洋渔业占有绝对优势,海洋经济规模很小。由于该阶段我国开发海洋资源的能力不足,所以经济活动对海洋资源和环境产生的压力较小,人与自然的矛盾尚不突出,海洋资源环境承载能力尚未被削弱,海洋环境基本处于“原生态”水平。

二、海洋经济快速发展阶段

从改革开放以后到20世纪末,是海洋经济高速发展阶段。这一阶段随着对外开放水平的不断提高,沿海的区位优势逐步显现,吸引了大量的资金、技术、劳动力向沿海一带聚集。海洋经济总量快速增长,特别是“八五”和“九五”期间,

海洋经济的增长速度都高于国民经济。该阶段海洋开发科技水平显著提高,开发强度日趋增大,但海洋三次产业结构仍不尽合理,海洋渔业仍然占据半壁江山。由于海洋管理的法律法规还不健全,管理经验不足,海洋开发方式较为粗放,出现了海洋资源开发“无序”“无度”“无偿”的现象,在追求海洋经济高速发展的同时,给近海资源和环境带来巨大压力,海洋环境质量整体下降。

三、海洋经济全面发展阶段

进入 21 世纪以来,随着科学发展观的确立与实践的深入,人们逐渐认识到海洋资源过度开发带来的后果,越来越关注海洋经济发展与资源环境的关系,关注沿海地区的可持续发展。海洋经济开始向又好又快的方向转变。“十五”期间,海洋经济年均增长 16.7%,比同期国民经济高出 3 个百分点。“十一五”期间,海洋经济年均增长 13.5%,持续高于同期国民经济增速。在此期间,海洋三次产业结构不断调整优化,由“十五”期初的 7:44:49 调整为 2012 年的 5:46:49,呈现出“三二一”的发展格局。随着海洋技术创新的不断突破,海洋传统产业得到改造提升,海水利用业、海洋可再生能源业等以海洋高技术为支撑的战略性新兴产业快速发展,“十一五”期间年均增速超过 20%。同时,邮轮、游艇、休闲渔业、海洋文化、涉海金融及航运服务业等一批新型服务业初露端倪,成为“十一五”期间海洋经济发展的新亮点。同时,海洋经济增长方式从注重数量向注重质量过渡,海洋环境保护工作得到重视和加强,但近岸海域环境尚未根本好转,海洋环境污染形势依然严峻。进入“十二五”以来,随着世界经济的持续低迷和国内经济的增速放缓,我国海洋经济发展的外部环境正在发生深刻变化,海洋经济虽然仍保持着平稳的发展态势,但已难以保持前些年的高速增长。据初步核算,2012 年海洋生产总值为 50 087 亿元,比上年增长 7.9%。可以说,海洋经济已经由高速增长期进入结构调整期。

第二节 海洋经济分析内容与作用

一、海洋经济分析的内容

海洋经济分析的基础是海洋经济统计数据,离开统计数据,统计工作只能是纸上谈兵。因此,海洋经济分析的内容和深度是随着统计数据的可获得性而不断发展和完善的。在现有条件下,目前能够开展的海洋经济分析评估的内容主要包括:海洋经济总量、海洋产业、区域海洋经济、海岛经济、海洋经济增长、

海洋经济周期、海洋经济监测预警和国际海洋经济发展等。

根据海洋经济工作实践,海洋经济分析主要包括制度化统计分析和专题统计分析。海洋经济制度化统计分析主要依托《海洋统计报表制度》和《海洋生产总值核算制度》,是在海洋经济统计指标体系和统计资料的基础上,对海洋经济发展总量、产业结构、增长速度、取得的主要成绩、存在的主要问题、相关对策建议以及海洋经济规划实施效果评估等所进行的全面系统分析。如:国家海洋局每年年初发布的《中国海洋经济统计公报》,每年8月发布的《上半年海洋经济统计分析报告》以及随着统计频次的增加而增加的季度和月度海洋经济统计分析等,其主要任务是为海洋经济管理决策提供系统、客观和科学的依据。

海洋经济专题统计分析是根据海洋经济管理工作的需要,针对具体领域或某一具体问题而不定期开展的专项统计分析。如:国家海洋局2003年开展的涉海就业人员统计分析,2007年开展的海岛经济统计分析以及2009年开展的金融危机对海洋产业发展的影响分析等。专题统计分析是制度化统计分析的一种补充,其主要目的是对所研究的专题或问题进行更加深入的分析 and 认识,以解决制度化统计分析无法或难以完成的工作。因此,单纯依靠制度化统计资料是不足以完成专题统计分析工作的,在具体实践中,要结合补充调查、抽样调查、典型调查和问卷调查等手段,去采集更全面、更详细的基础数据和资料。

二、海洋经济分析评估的作用

随着海洋经济统计工作的不断完善,海洋经济分析也在不断完善,在总量和结构上不断深化系统分析的理论和方法,并从静态的、简单的分析逐步向动态的、复杂的分析发展,同时也在不断加强对海洋经济运行机制方面的系统分析。因此,海洋经济分析在海洋经济宏观指导和调控中发挥着越来越重要的作用,主要体现在以下方面^①:

(1)把握海洋经济运行的基本状况,对海洋经济运行做出正确的判断。海洋经济是一个复杂的运行系统,不同行业、不同地区以及不同环节之间存在着复杂的经济联系,只有对海洋经济发展各个方面和各个层次进行全面分析,才能把握海洋经济运行的全貌,并在此基础上对海洋经济运行的状态做出正确的判断。

(2)揭示海洋经济运行中的主要矛盾和问题,为制定海洋经济政策、进行宏观指导和调控服务。通过开展海洋经济分析,可以揭示海洋经济运行中存在的主要矛盾和问题,及时发现新出现的重要矛盾和问题,分析问题的性质、形成的原因、对海洋经济发展和运行的影响,揭示这些矛盾和问题发展变化的趋势和动

^① 参考陈瑾玫.宏观经济统计分析与实践[M].北京:经济科学出版社,2005.

向，从而为政府决策部门有针对性地采取一些经济政策措施提供依据。

(3) 分析海洋经济政策的效应和有效性。海洋经济运行质量如何，很大程度上反映了海洋经济政策实施的效应，海洋经济分析能够反映经济政策实施的效果与预期效果的差异、经济政策执行情况、经济政策对各行业的影响情况以及政策实施后所产生的新情况和新问题，从而检验海洋经济政策的正确性和适应性，为政府决策部门重新制定或修正现行政策提供全面、可信的参考。

(4) 预测海洋经济发展趋势和运行态势，提出今后海洋经济管理的对策建议。通过开展海洋经济分析，可以对海洋经济未来发展走向做出基本判断，包括经济增长的适合速度、经济运行的变化和平稳程度、经济运行质量前景以及经济运行中可能出现的一些问题，从而为决策者提供具有重要参考价值的政策咨询建议。

第三节 海洋经济分析指标

统计指标是经济分析的基础和工具。海洋经济分析是从海洋经济现象的数量方面来认识海洋经济活动的，因此就要借助于海洋经济统计指标。根据统计指标的作用和特点，海洋经济分析所应用的指标包括绝对指标（亦称总量指标）、相对指标和平均指标三类。这三类指标在统计学中都是用于概括和分析现象总体的数量特征和数量关系的，统称为综合指标。海洋经济分析评估离不开这三种基本的综合指标，其他统计指标都是在这些基本指标的基础评估上进一步加工、演化和具体运用的结果。

一、绝对指标

绝对指标是反映经济现象在一定时间、地点、条件下的总规模或总水平的统计指标。绝对指标也称为总量指标或绝对数。绝对指标是计算相对指标、平均指标和各种分析指标的基础。相对指标和平均指标一般都是由两个相关的绝对指标对比得到的，它们都是绝对指标的派生指标。

绝对指标是统计分析过程中所使用的最基本的综合指标，在实际工作中应用十分广泛。在经济分析中，绝对指标主要用于经济总体或各产业的价值量、实物量的差额分析和反映某一经济现象在两个不同时期数量增减的变化。如：2011 年我国海洋生产总值为 4.55 万亿元，海水养殖产量为 1551.3 万吨，沿海地区海盐产量比上年增长 35.79 万吨等都是绝对指标。

按绝对指标反映的时间状况不同，绝对指标又分为时期指标和时点指标。时期指标又称时期数，它反映的是经济现象在一段时期内的总量。如海水养殖产

量、盐产量、海洋服务业收入等。时期数通常可以累积,从而得到更长时期内的总量。时点指标又称时点数,它反映的是经济现象在某一时刻上的总量,如年末涉海就业人数、涉海企业设备台数、海洋专业在校学生数等。时点数通常不能累积,各时点数累积后没有实际意义。

在经济分析中,绝对指标具有一定的经济内容,一般都有计量单位。按绝对指标所反映的客观事物性质的不同,计量单位可分为实物单位、货币单位和劳动单位。如万标准箱、万元、天等。

使用绝对指标时应注意:①一般情况下,只有同类现象才能加总。例如,海洋水产品产量和海洋化工产品产量,两者的性质不同,所以不能将两者加总。但在某些特殊情况下,对于具体形式不同但使用价值相同的产品,可以折算为标准品再行加总,如原煤、原油、天然气、水电等,可以折算成标准煤后加总。②必须明确绝对指标的含义。只有明确绝对指标的含义,才能科学地确定指标的计算范围和计算方法,进而准确地计算绝对指标。例如,海洋电力业增加值是指电力企业和单位一定时期内在沿海地区利用海洋能、海洋风能进行电力生产活动的最终成果的货币表现形式。沿海火力发电、核能发电都不是海洋电力业应包括的内容。因此,一定要根据研究目的,统一规定指标的含义。③统计汇总时必须要有统一的计量单位。同类现象的绝对指标的数值,其计量单位只有统一时,才能加总,否则,在统计汇总时,首先要换算成统一的计量单位。如:海水水产品的计量单位可用吨或千克表示,计算时要统一。

二、相对指标

相对指标是说明经济现象之间数量对比关系的统计指标,是两个有联系的绝对指标数值之比,又称相对数。相对指标在经济分析中有着较为广泛的应用,其表现形式通常为比例和比率两种。如:2012年,我国海洋经济增长速度为7.9%;海洋生产总值占国内生产总值的比重为9.6%等都是相对指标。

根据对比基础和研究目的的不同,相对指标可以分为结构相对指标、比例相对指标、动态相对指标、比较相对指标、强度相对指标和计划相对指标六种。在经济分析实践中,比较常用的是前四种。

结构相对指标是总体某部分数值与总体全部数值之比,又称比重或比率,一般用百分数(%)表示。如海洋三次产业比重。利用结构相对指标可以研究总体内各组成部分的分配比重及其变化情况,从而深刻认识事物各个部分的特殊性及其在总体中所占的地位。

比例相对指标是同一总体各组成部分数量之间的对比而得到的相对指标,又称比例相对数。利用比例相对指标,可以分析总体内各组成部分或各局部之间的

数量关系是否协调一致。按比例发展是事物发展的客观要求,如各产业之间的比例、人口性别比例等都可以运用比例相对指标进行分析研究。在比例相对数的计算过程中,分子和分母数值可以互换颠倒。

动态相对指标是指将同类指标在不同时期数量对比关系的相对指标,一般用百分数(%)或倍数表示。其中的基期可以是上期、上年同期、历史水平最好时期,也可以是特殊意义的时期。动态相对指标实际上就是发展速度,在统计分析中应用广泛,利用该指标可以了解经济活动的发展动态和增长速度(即发展速度减1)。当动态相对指标值大于0小于1时,表明报告期水平低于基期水平;当动态相对指标值等于1或大于1时,表明报告期水平达到或超过基期水平。

比较相对指标是将两个不同地区、部门、单位的同类指标做静态对比得到的相对指标,一般用百分数(%)或倍数表示。使用时,相比较的两个指标所属的指标含义、口径、计算方法和计量单位必须要一致。比较相对指标可以用绝对指标进行对比,也可以用相对指标或平均指标进行对比。但由于总量指标易受总体范围大小的影响,因而,计算比较相对指标时,更多地采用相对指标或平均指标。例如我国人均GDP与英国人均GDP相比。

强度相对指标又称强度相对数,是有一定联系的两种性质不同的总量指标相比较形成的相对指标。通常以复名数和百分数(%)、千分数(‰)等无名数表示。例如人口密度。

计划相对指标又称计划完成百分数。将实际完成数与计划任务数相比较,用以表明计划完成情况的相对指标,通常用百分数(%)表示。计划完成程度指标的分子是实际完成数,分母是计划任务数,分子和分母在指标含义、计算方法、计量单位以及时间长度等方面应该完全一致。同时,分子、分母不允许互换。

相对指标的使用原则是:①要与绝对指标相结合。相对指标虽可以反映现象之间的差异程度,但把现象的绝对水平抽象化了,说明不了现象之间在绝对数量上的差异。因此,应用相对指标进行统计分析时,必须与其背后的绝对水平以及两个对比指标的绝对额结合起来,以全面、正确地认识客观事物。②要注意分子与分母的可比性。主要是指对比的分子和分母两个指标之间在经济内容、计算范围、计算方法和计量单位等方面要保持一致或相互适应的状态。例如,由于不同时期商品和劳务价格水平的不同,不能简单地将2011年海洋生产总值同2005年海洋生产总值进行对比,为了保证两者的可比性,应剔除价格因素影响,统一使用不变价。③要综合运用多种相对指标。一种相对指标只能说明一个方面的问题,在分析研究复杂现象时,应该将多种相对指标结合起来运用,这样才能把从不同侧面反映的情况结合起来观察分析,从而较全面地说明客观事物的情况及其发展规律。

三、平均指标

平均指标又称平均数，是海洋经济统计中十分重要的指标。平均指标是指同质总体的某一标志值在一定时间、地点条件下达到的一般水平。它在一定意义上反映了总体分布的集中趋势。利用平均指标还可以比较不同空间同一事物一般水平的差异，比较总体各种标志的一般水平在时间上的变动过程和趋势，分析现象之间的依存关系。

平均指标是把各个变量之间的差异抽象化，从而说明总体的一般水平。平均指标只能在同质总体中计算，这是计算平均指标的前提。常用平均指标的计算方法主要有五种：简单算术平均数、加权算术平均数、几何平均数、众数和中位数。

简单算术平均数是计算平均指标最基本、最常用的方法，是将总体各个单位的指标数值加总除以总体单位个数。在计算算术平均数时，分子与分母必须同属一个总体，在经济内容上有着从属关系，即分子数值是分母各单位标志值的总和。也就是说，分子与分母具有一一对应的关系，有一个总体单位必有一个标志值与之对应。

加权算术平均数是具有不同比重的数据（或平均数）的算术平均数。比重也称为权重，反映了该变量在总体中的相对重要性，每种变量权重的确定与一定的理论经验或变量在总体中的比重有关。依据各个数据的重要性系数（即权重）进行相乘后再相加求和，就是加权。加权与所有权重之和的比就是加权算术平均数。加权算术平均数主要用于原始资料已经分组，并得出次数分布的条件。当各个标志值的权数都完全相等时，权数就失去了权衡轻重的作用，这时候，加权算术平均数就成为简单算术平均数。

几何平均数是 n 项变量值连乘积的 n 次方根。在经济分析中，几何平均数常用来计算多年平均发展速度。

众数是总体中出现次数最多的标志值。它能直观地说明客观现象分配中的集中趋势。如果总体中出现次数最多的标志值不是一个，而是两个，那么，合起来就是复（双）众数。只有在总体单位数较多，各标志值的次数分配又有明显的集中趋势时才存在众数；如果总体单位数很少，尽管次数分配较集中，那么计算出来的众数意义也不大；或尽管总体单位数较多，但次数分配不集中，即各单位的标志值在总体分布中出现的次数较均匀，那么也无所谓众数。由于众数是由标志值出现次数多少决定的，不受资料中极端数值的影响，因此使用众数可以增强对总体一般水平的代表性。

中位数是将各单位标志值按大小排列，居于中间位置的那个标志值就是中位数。对于未分组资料先将数据按从小到大顺序排列，如项数为奇数，居于中间的

那个单位标志值就是中位数；如项数为偶数，那么中位数就是中间两个数据的平均数。使用中位数可以不受数列中极大或极小数据的影响，同样也可以增强对总体一般水平的代表性。

第四节 海洋经济分析数据预处理

高质量的数据是统计分析结论可靠性的根本保障。作为统计整理阶段的重点工作，统计数据预处理是对原始数据质量进行审查、诊断、评估及提升的一个过程，它直接决定着分析数据的质量，影响到统计产品的可信度以及以此所做决策的科学性。本节重点就统计数据预处理的必要性、处理过程和处理方法进行论述。

一、统计数据预处理的必要性

在统计工作中，人们普遍重视对数据收集和统计分析的研究，却相对忽视对数据收集之后、正式分析之前这一中间阶段的研究，而这一阶段的主要工作就是统计数据的预处理。在数据收集阶段，无论如何仔细认真，不管是一手数据还是二手数据，总是不可避免地会存在一些质量问题。统计调查数据由于调查过程中的工作失误、被调查者不配合、抽样方法选取不当、问卷设计不合理等因素而存在误差；利用信息采集系统收集到的数据，由于数据录入、转换及数据库链接等过程中的失误，可能会出现错误字段、记录重复或缺失等问题；政府统计部门生产的宏观统计数据，也会因人为干扰、体制缺陷等原因而存在数据质量问题；一些上市公司在财务数据上弄虚作假、发布虚假信息；一些商业性调查由于样本选择不规范、调查偷工减料、弄虚作假，甚至人为编造数据，让人对数据质量产生质疑。正是由于这些问题的客观存在，降低了统计结果的可信度，同时也给后续的研究工作带来严重影响。

统计数据的质量需要贯穿统计工作始终，数据质量是计量经济模型赖以建立和成功应用的基础条件，保障统计数据的质量是统计分析的关键，为了满足统计分析的实际需要，提高数据质量，保证统计分析结果的客观、有效，在正式开展统计分析之前，对统计数据进行预处理是十分必要的。

二、统计数据预处理的步骤

统计数据预处理包括数据审查、数据清理、数据转换和数据验证四大步骤。^①

^① 程开明，统计数据预处理的理论与方法述评，统计与信息论坛[J]，2007，2(6)：98-103。

（一）数据审查

主要检查数据的数量（记录数）是否满足分析的最低要求，字段值的内容是否与调查要求一致，是否全面；还包括利用描述性统计分析，检查各个字段的字段类型，字段值的最大值、最小值、平均数、中位数等，记录个数，缺失值或空值个数等。

（二）数据清理

主要针对数据审查过程中发现的明显错误值、缺失值、异常值、可疑数据，选用适当的方法进行“清理”，使“脏”数据变为“干净”数据，有利于后续的统计分析得出可靠的结论。当然，数据清理还包括对重复记录进行删除。

（三）数据转换

数据分析强调分析对象的可比性，但不同字段值由于计量单位等不同，往往造成数据不可比；对一些统计指标进行综合评价时，如果统计指标的性质、计量单位不同，也容易引起评价结果出现较大误差，再加上分析过程中的其他一些要求，需要在分析前对数据进行变换，包括无量纲化处理、线性变换、汇总和聚集、适度概化、规范化以及属性构造等。

（四）数据验证

数据验证的主要目的是初步评估和判断数据是否满足统计分析的需要，决定是否增加或减少数据量。利用简单的线性模型以及散点图、直方图、折线图等图形进行探索性分析，利用相关分析、一致性检验等方法对数据的准确性进行验证，确保不把错误和偏差的数据带入到数据分析中去。

上述四个步骤是逐步深入、由表及里的过程。先是从表面上查找容易发现的问题（如数据记录个数、最大值、最小值、缺失值或空值个数等），接着对发现的问题进行处理，即数据清理，再就是提高数据的可比性，对数据进行一些变换，使数据形式上满足分析的需要。最后则是进一步检测数据内容是否满足分析需要，诊断数据的真实性及数据之间的协调性等，确保优质的数据进入分析阶段。

三、统计数据预处理的方法

选用恰当方法开展统计数据预处理，有利于保证数据分析结论真实、有效。根据处理对象的特点和各步骤的不同任务，统计数据预处理可采用的方法包括描述和探索性分析、缺失值处理、异常值处理、数据变换技术、信度与效度检验、宏观数据诊断六类。^①

^① 程开明. 统计数据预处理的理论与方法述评. 统计与信息论坛[J]. 2007, 2(6): 98-103.

对应于统计数据预处理的四个步骤，各有不同的处理方法。数据审查阶段主要是对调查数据进行信度、效度检验，利用描述及探索性分析手段对数据进行基本的统计考察，初步认识数据特征；数据清理阶段主要是利用多种插值方法对缺失值进行插补，采用平滑技术进行异常值纠正；数据转换阶段则根据不同的需要可供选择的方法较多，针对计量单位不同可采用无量纲化和归一化，针对数据层级不同可采用数据汇总、泛化等方法，结合分析模型的要求可对数据进行线性或其他形式的变换、构造和添加新的属性以及加权处理等；数据验证阶段包括确认上述步骤的正确性与有效性，检查数据的逻辑转换是否造成数据扭曲或偏差，并再次利用描述及探索性分析检查数据的基本特征，对数据之间的平衡关系及协调性进行检验。

（一）描述和探索性分析

描述性统计技术主要是对数据开展频数、描述统计量及列联表分析。频数分析是利用非连续变量的频数表，报告出变量个数、记录数以及缺失值等；描述统计量分析主要是计算连续变量的均值、标准差、最小值、最大值、偏度、峰度等统计量，以便检查出超出范围的数据或极端值；列联表分析主要起到交叉分类的作用，从中可以很容易地发现逻辑上不一致的数据。

探索性分析利用图形直观地考察数据所具有的特征，反映数据的分布特征、发展趋势、集中和离散状况等，主要包括茎叶图、箱形图、散点图、直方图、折线图、条形图等。茎叶图把观测数据分为茎和叶两部分，使我们认识到数据接近对称的程度、是否有数据远离其他数据、数据是否集中、数据是否有间隙等特征。箱形图有助于直观地描述分布与离散状况，利用最大值、最小值、中位数、上四分位数和下四分位数等反映出数据的实际分布。散点图用于直观地表现两个或多个变量之间有无相关关系，并反映数据的分布、集中、离散状况；P-P图和Q-Q图则可用于展示数据是否符合正态分布；直方图、折线图、饼图、雷达图等，都可从不同侧面直观反映出数据的特征和趋势。

（二）缺失值处理

对缺失数据的处理方法可以分为四类。

1. 忽略

若一条记录中有属性值缺失，则将该条记录排除在数据分析之外。该方法简单易行，但是容易导致严重的偏差，仅适用于含有少量缺失数据的情况。

2. 插补(替代)

可采用以下几种方法：①使用一个固定的值代替缺失值。所有缺失值用一个常量代替，譬如用字母“N”代替缺失值。当某一属性的缺失值较多，使用此方法