

# 模型的力量

全国统计建模大赛获奖论文选


全国统计建模大赛执行委员会  
国家统计局统计教育中心 编

 中国统计出版社  
China Statistics Press

# 模型的力量

全国统计建模大赛获奖论文选

全国统计建模大赛执行委员会  
国家统计局统计教育中心 编

 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

模型的力量:全国统计建模大赛论文选/全国统计建模大赛  
执行委员会,国家统计局统计教育中心编.

—北京:中国统计出版社,2009.1

ISBN 978-7-5037-5337-4

I. 模…

II. ①全… ②国…

III. 统计模型—文集

IV. C8-53

中国版本图书馆 CIP 数据核字(2009)第 002206 号

### 模型的力量:全国统计建模大赛论文选

---

作 者/全国统计建模大赛执行委员会

国家统计局统计教育中心

责任编辑/吕 军 孙 慧

封面设计/艺编广告

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 邮政编码/100073

电 话/邮购(010)63376907 书店(010)68783172

印 刷/河北天普润印刷厂

开 本/787×1092mm 1/18

字 数/300 千字

印 张/23.5

版 别/2009 年 1 月第 1 版

版 次/2009 年 1 月第 1 次印刷

书 号/ISBN 978-7-5037-5337-4/C·2172

定 价/70.00 元

---

版权所有。未经许可,本书的任何部分不准以任何方式在世界  
任何地区以任何文字翻印、拷贝、仿制或转载。

中国统计版图书,如有印装错误,本社发行部负责调换。

# 在全国统计建模大赛颁奖仪式上的讲话

(代序)

马建堂

首先,我代表国家统计局对全国统计建模大赛的圆满成功表示热烈的祝贺!对取得优异成绩的代表队表示热烈的祝贺!

在全国统计系统深入学习实践科学发展观,加快统计事业改革和发展的新形势下,举办这次比赛对于提高统计干部的专业素质,推动统计工作更好地服务科学发展具有十分重要的意义。比赛中,你们以扎实的统计知识、过硬的业务素质、饱满的参赛热情,充分展示了新时代统计人锐意进取、改革创新的精神风貌。这次参赛的同志们都很年轻,有很多同志只有20多岁,我心里感到十分高兴。年轻人思维敏捷,敢于突破,善于创新,是我国统计事业的希望,从你们身上我看到了我国统计事业改革和发展的希望和未来。希望你们发扬成绩,戒骄戒躁,加倍努力,继续为我国的统计事业贡献自己的聪明才智。

充分发挥统计系统自身的优势,围绕经济社会发展中的重大问题进行深入研究,是时代赋予我们的重要职责。这次大赛的题目选得很好,是《区域发展差距:成因与走向》。应当说,改革开放30年来,我国经济社会发展取得了巨大的成就,尤其是在小平同志“两个大局”思想的指导下,全国各地都有了很大的发展变化,但由于我国地域辽阔,区域经济发展不平衡、区域差距不断扩大的问题成为我国经济社会发展的一个突出问题,需要我们认真研究。选择这个题目,不仅是提高我们应用数据、分析数据、处理数据能力的需要,更是提高我们分析研究能力、服务科学发展能力的需要。统计调查工作中,科学准确的统计数据很重要,深度的分析研究同样重要,我们要坚持两手抓,抓两手,通过一定的数据分析模型,为各级党和政府提出有深度的分析报告、有见地的政策建议,进而更好地为党和政府发挥好参谋助手作用。

深化统计改革、提高数据质量、服务科学发展的历史使命落在了我们的肩上,时代需要更多高素质的统计人才。举办这样的活动,不仅为统计人才相互交流学

习提供了一个平台,也是我们发现和培养统计专业人才的重要途径。这次建模大赛为推动统计系统掀起学习业务、提升技能、增强本领的新高潮开了一个好头。我相信,随着我们一些行之有效的教育活动和培养方法实施,统计人才一定会辈出不穷,统计人也一定能够谱写出我国统计事业改革和发展的新篇章。

明天就是国庆节了,在这里向大家致以节日的问候和良好的祝愿!

谢谢大家!

2008年9月

# 目 录

## 国家统计局局长马建堂在全国统计建模大赛颁奖仪式上的讲话

### 一、全国统计建模大赛决赛论文

- 第一届全国统计建模大赛赛题 ( 1 )
1. 区域发展差距:成因与走向  
北京市统计局、国家统计局北京调查总队第二联队 杜婷、郑晓光、刘狄菲 ( 4 )
  2. 区域发展差距:成因与走向  
天津市统计局 郑礼、冯志刚、李娜 ( 26 )
  3. 区域发展差距:成因与走向  
国家统计局广东调查总队 许鹏、杨际昌、戴伟星 ( 43 )
  4. 中国区域发展差距:成因与走向  
山东省统计局 山军勇、杨镇、张毅 ( 56 )
  5. 我国地区经济增长收敛性的实证研究  
国家统计局综合司队 付凌晖、戴昱乐、俞炳彬 ( 73 )
  6. 我国区域发展差距的现状、成因及走向实证研究  
国家统计局重庆调查总队 汤雁冰、孙卓霖、邓云松 ( 90 )
  7. 我国区域经济发展差异成因与走向研究  
国家统计局新疆调查总队 张敬东、杨磊、秦子啸 ( 108 )
  8. 我国区域发展差距的成因分析及走向预测  
河南省统计局 徐良、马召、张亚珂 ( 129 )

### 二、全国统计建模大赛优秀专题论文

1. 区域政府投资的经济增长效应  
浙江省统计局 张荣飞、杨士鹏、周维松 ( 144 )
2. 我国 CPI 变化规律实证分析  
天津市统计局 冯志刚、郑礼、李娜 ( 159 )

3. 对我国当前物价形势的基本判断  
上海市局队联队 阮大成、杨建忠、朱国众 (172)
4. 资本与劳动力要素对陕西经济增长的贡献研究  
陕西省统计局 赵翔、李建平、宋鑫 (185)
5. 山西农村居民消费行为研究  
国家统计局山西调查总队 刘建业、张春雷、韩建华 (194)
6. 山东宏观经济需求模型实证研究  
山东省统计局 山军勇、杨镇、张毅 (210)
7. 中国农户土地承包经营权流转动因实证分析  
国家统计局河南调查总队 李德洗、郑凯、王亚钊 (226)
8. 经济增长、收入分配与农村减贫的实证分析  
国家统计局农村司 余新华、阎芳、柏先红 (254)
9. 基于 VAR 方法的宏观经济预警系统构建  
北京市统计局、国家统计局北京调查总队  
第一联队 丁文斌、陈倩、丁海峰 (270)

### 三、第二届青联杯国家统计局机关统计建模比赛一等奖论文

- 第二届青联杯国家统计局机关统计建模比赛赛题 (284)
1. 我国财政分权与地方经济增长的关系研究  
国家统计局核算司 施发启、郑学工、吕峰 (287)
  2. 财政分权与地方经济增长的关系研究  
国家统计局综合司 付凌晖、戴旻乐、俞炳彬 (302)
  3. 财政分权与地方经济增长的关系研究  
国家统计局调查中心 王利、侯守国、江永宏 (316)
  4. 财政分权与地方经济增长的关系研究  
国家统计局农村司 柏先红、王海龙、马治鑫 (330)
  5. 财政分权与地方经济增长  
国家统计局贸经司 卢山、叶盛、刘晓燕 (347)

国家统计局总统计师李强关于大赛论文的评价和建议 (366)

# 第一届全国统计建模大赛赛题

## 区域发展差距：成因与走向

### 背景

改革开放 30 年来，中国经济持续快速增长，综合国力显著提高，人民生活水平明显改善，但在增长、提高、改善的同时，各地区经济增长速度快慢不一、区域发展不平衡的现象日益突出。

区域发展差距已经成为中国经济社会发展的重大挑战。

区域发展差距已经成为中国政府和学术界关注的焦点。

近年来，政府高度重视区域发展差距，提出了各种应对措施；学术界也围绕区域发展差距给出了很多研究，提出了不少有价值的观点。

请根据赛会提供的数据，分析改革开放以来我国各省级行政区之间或全国各经济区域之间的发展差距，探讨其成因，并对我国区域发展差距的可能走向进行统计分析。

### 要求

1. 对我国各省级行政区或全国各经济区域之间发展差距的历史演进情况进行分析。

2. 探讨自然地理区位、政策制度、地区经济结构、人力资本等因素对我国区域发展差距的影响。

3. 建立我国区域发展差距统计模型并讨论其适用条件、合理性及可靠性。

4. 根据所建立的模型对我国区域发展差距情况进行分析，并预测我国区域发展差距的可能走向。

5. 在上述工作的基础上，提出相应的政策主张。

在以上分析中，请集中考虑以经济发展水平测量的区域发展差距。

### 数据

赛会提供了下述数据资源：

1. 中国各省级行政区的人均国内生产总值



2. 中国国内生产总值平减指数
  3. 中国各省级行政区的基础设施水平
  4. 中国各省级行政区的全社会固定资产投资
  5. 中国各省级行政区的中国进、出口总额
  6. 中国各省级行政区的实际利用的外商直接投资
  7. 中国的人民币与美元汇率
  8. 中国各省级行政区的产业结构
  9. 中国各省级行政区的教育情况
  10. 中国各省级行政区的人口规模
  11. 中国各省级行政区的农业和非农人口比例
- 可以使用其他的数据资源,但请说明数据来源。

### 参考文献

为帮助理解区域发展差距的内涵,赛会提供了一部分参考文献供参考。  
引用他人的文字、观点、模型必须说明。

### 对答卷的要求

答卷应该是一篇包括问题描述、数据和指标描述、模型假设、建立和求解、结果分析和检验、政策主张的论文。

论文的篇幅以 10000 字为限。

论文请参考下述格式:

1. 论文题目
2. 作者:请署“第 X 代表队”,这里的 X 为赛会提供的参赛队编码
3. 摘要
4. 关键词
5. 正文
6. 附录
7. 参考文献

论文正文请使用四号仿宋字。

### 参考信息:关于论文评审和答辩

论文评审主要考虑论述的科学性、模型的适用性、结果的正确性和表述的清晰性。

论述的科学性占评审总分的 20%，主要考察：1. 理论的运用、阐述是否科学、恰当，是否抓住了赛题的本质和核心内容；2. 论证过程是否正确、严谨，逻辑是否严密，结构是否合理。

模型的适用性占评审总分的 40%，主要考察：1. 模型是否和要讨论的问题匹配，模型的假设是否合理；2. 模型的适用条件及合理性和可靠性是否清晰，模型的建构过程是否有创新；3. 模型的求解、结果分析和检验是否适当；4. 数据的选择、清理和转换是否适当。

结果的正确性占评审总分的 20%，主要考察：1. 得出的结论是否正确、严谨；2. 政策主张是否适当，是否有创新。

表述的清晰性占评审总分的 20%，主要考察：1. 行文是否通畅，精练，有否可读性；2 术语是否准确，行文及图表是否符合学术规范。

论文答辩，主要考虑论文报告表现出来的论文水平（50%）、对提问的理解（15%）、回答的针对性（15%）以及回答的准确性（20%）。

# 区域发展差距:成因与走向

北京市统计局、国家统计局北京调查总队第二联队  
杜婷、郑晓光、刘狄菲

## 摘要

区域发展差异是一国经济社会发展进程中的必经阶段,多年来各国政府一直致力于在保证经济总体向前发展的前提下,促进各区域协调、共同发展,缩小不同地区间的差距。本文在总结国内外众多学者研究成果的基础上,运用我国 1997 年至 2006 年间的统计数据,从国内投资和消费、教育发展水平、区域开放程度以及产业结构等多个方面出发,对区域经济发展差异的成因进行了探讨。

文章首先使用偏离—份额分析法对影响东中西部各省市国内生产总值增长的因素进行了分析,结果显示东部地区具有明显的发展优势,而中西部地区,虽然取得了一定的发展,但与东部地区仍有较大差距。随后,运用主成份方法将多个指标降维成四大综合因子,作为后续模型的自变量。在此基础上,以各省市(除西藏自治区)为横截面,构建个体变系数的 Panel Data 模型,并结合实际情况进行了分析。第三,建立灰色预测模型对 2007 年全国人均 GDP 及东中西部地区人均 GDP 进行了预测,并对东中西部地区的变异系数进行了测算,得出未来几年东中西部地区发展差距将会进一步缩小的结论。最后,在理论及模型分析的基础上,针对我国区域发展现状提出了政策建议。

**关键词:**区域经济差距 偏离—份额分析 面板数据模型 灰色模型 R 软件

## 一、问题的提出

改革开放以来,中国经济实现了近 30 年的持续高速发展,不断增长的经济实力与竞争力,引来了世人的关注与瞩目。但是,由于我国各地区自然资源禀赋、地理位置、改革开放程度以及政策制度等方面具有较大差别,导致我国各省市及特定经济区域间的发展很不均衡,区域差异问题日趋凸显,且某些方面,差距还有进

一步扩大的趋势。

如何解决区域之间经济发展不平衡的问题、有效遏制区域经济发展差距扩大的势头、实现全国人民的共同富裕,是摆在我们面前的严峻课题。研究区域差距及其成因,无论是对未来区域经济发展规划还是对制定区域政策都具有重大意义。

## 二、研究现状及存在的问题

近年来,国内外学者采用多种方法对区域经济发展水平及其差异进行了大量探索,但由于选取的指标、研究的时点及采取的方法不同,得出的结论也不尽相同。

从研究方向看,目前已有的研究主要集中在以下几个方面:一是采用描述统计的方法对区域经济发展差距进行研究;二是利用经济学中反映区域差异变化的各种指标对经济增长和地区经济差距进行测度和分解;三是从某一角度,如政策制定角度,或是产业结构角度对区域差距进行比较;四是采用计量经济学方法对反映区域经济发展差异的各类指标建立模型、进行分析。

从众多学者的研究内容可以看出,对区域发展差距的探索或集中在对某几个因素的研究上、缺乏综合考虑,或以定性分析为主、缺少统计模型的有力支撑,或仅对差距的成因进行分析、没有对这种差距的未来走向进行定量判断。有鉴于此,本文从国内投资和消费、教育发展水平、区域开放程度以及产业结构等多个方面出发,对区域经济发展差异的成因进行探讨。

## 三、区域经济发展差距的历史演进及其影响因素分析

### (一) 区域经济发展差距的历史演进

在区域研究中,一般把我国划分为东部、中部和西部三大区域(林毅夫,1998;万广华,1998;宋德勇,2001;邱宜干,2001)。从目前情况看,东部地区属较发达地区,中部和西部地区相对落后。三大区域之间经济发展差距的形成主要受历史背景和国家政策两方面因素的影响。

#### 1. 历史背景对区域经济发展差距的影响

中、西部地区位于黄河上中游,是中华民族的重要发祥地。由于朝代变迁及近代以来外国殖民主义入侵等原因,中国经济社会的重心开始逐渐南移,东部地区由于沿海的区位优势逐渐成为外国资本对华贸易和资本输出的重点地区。

#### 2. 国家经济政策的影响

1978年改革开放以来,我国开始实行先发展东部拉动中、西部的梯次发展战略,对东部地区给予了一系列政策优惠,并在东部沿海省市建立了5个经济特区。同时在财政、税收、金融、贸易、引进外资、人事制度等方面也赋予了诸多优惠政

策。这些都在客观上对区域经济差异的形成起到了推动作用。

本文将采用传统的东中西部划分方法对区域差异问题进行研究,西部包括重庆、四川等 12 个省份;中部包括黑龙江、吉林等 8 个省份;东部包括上海、北京等 11 个省份。

## (二) 1997—2006 年我国东中西部地区经济发展影响因素的偏离—份额分析

历史已经成为过去,目前阶段,究竟哪些因素对我国区域经济发展产生着重要影响? 本文使用偏离—份额分析法,从总量、份额、产业结构、竞争力四个方面,对 1997 至 2006 年影响各省市国内生产总值的增长因素进行了分析,并按照东中西部三大区域的划分,对区域特点进行研究。

### 1. 偏离—份额分析法

偏离—份额分析法(Shift-Share Analysis,简称 SS 分析法)的基本思路是将被研究区域的增长与标准区域(通常是一个国家或一个省)的增长联系起来比较,从而判定区域经济增长的差异。

### 2. 利用偏离—份额分析法对东、中、西部产业结构进行分析

#### (1) 数据的来源和处理

以 1997 年为基年,2006 年为目标年,利用《中国统计年鉴》中全国各地区的国内生产总值和第一、二、三产业产值,通过 SS 模型分析计算份额分量、产业结构偏离分量和竞争力分量三个方面对经济增长的影响。

#### (2) 经济增长的总体差异

对全国 31 个省区进行 SS 分析,并对各区域的总增长量、份额分量、产业结构偏离分量以及竞争力分量进行排序,结果见以下各表。

表 1 东部地区各省市经济增长的 SS 分析排序

省市	G	省市	N	省市	P	省市	D
广东	18429.95	广东	13230.19	上海	703.86	广东	4691.07
山东	15540.29	江苏	11368.17	广东	508.69	山东	4550.10
江苏	14964.74	山东	11124.36	北京	492.95	江苏	3536.82
浙江	11056.41	浙江	7974.52	天津	186.16	浙江	2944.27
河北	7706.65	河北	6728.29	辽宁	185.18	北京	1769.53
上海	6927.58	辽宁	6096.47	浙江	137.62	河北	1129.06
北京	5794.65	上海	5851.91	江苏	59.75	天津	756.30
辽宁	5668.65	福建	4885.51	福建	-69.81	上海	371.81
福建	4743.65	北京	3532.17	海南	-71.67	海南	13.68
天津	3094.52	天津	2152.07	山东	-134.17	福建	-72.05
海南	641.69	海南	699.69	河北	-150.69	辽宁	-612.99
合计	94568.78	合计	73643.35	合计	1847.87	合计	19077.60

表2 中部地区各省市经济增长的SS分析排序

省市	G	省市	N	省市	P	省市	D	省市	S
河南	8454.88	河南	6876.87	山西	85.88	河南	2025.83	河南	1578.01
湖北	4724.85	湖北	4860.96	黑龙江	-72.46	山西	678.91	山西	764.79
湖南	4719.62	湖南	4848.70	吉林	-124.18	江西	545.19	江西	332.16
安徽	3801.41	黑龙江	4539.38	江西	-213.03	吉林	443.04	吉林	318.86
黑龙江	3521.40	安徽	3994.52	湖北	-302.26	湖南	294.07	湖南	-129.08
山西	3276.55	江西	2732.60	安徽	-383.20	安徽	190.09	湖北	-136.11
江西	3064.76	山西	2511.76	湖南	-423.15	湖北	166.15	安徽	-193.11
吉林	2810.78	吉林	2491.92	河南	-447.81	黑龙江	-945.52	黑龙江	-1017.98
合计	34374.25	合计	32856.71	合计	-1880.21	合计	3397.76	合计	1517.54

表3 西部地区各省市经济增长的SS分析排序

省市	G	省市	N	省市	P	省市	D	省市	S
四川	5396.34	四川	5516.12	陕西	-2.11	内蒙古	1811.28	内蒙古	1675.00
内蒙古	3637.97	广西	3092.48	宁夏	-2.62	陕西	841.77	陕西	839.65
陕西	3160.14	云南	2852.40	青海	-2.90	新疆	344.27	新疆	235.86
广西	3011.25	陕西	2320.49	西藏	-15.92	贵州	266.60	甘肃	132.67
云南	2330.55	重庆	2314.77	甘肃	-65.50	四川	255.45	贵州	104.97
重庆	2131.33	内蒙古	1962.97	重庆	-69.32	广西	218.11	宁夏	103.98
新疆	2005.41	新疆	1769.55	新疆	-108.41	甘肃	198.18	青海	93.69
甘肃	1483.12	贵州	1371.24	云南	-128.54	宁夏	106.59	西藏	82.33
贵州	1476.21	甘肃	1350.45	内蒙古	-136.28	西藏	98.25	广西	-81.23
宁夏	486.17	宁夏	382.19	贵州	-161.63	青海	96.58	四川	-119.78
青海	438.78	青海	345.09	广西	-299.34	重庆	-114.12	重庆	-183.44
西藏	213.77	西藏	131.44	四川	-375.23	云南	-393.31	云南	-521.85
合计	25771.04	合计	23409.19	合计	-1367.80	合计	3729.65	合计	2361.85

从SS分析的结果可以看出,十年间,东部地区表现出明显的发展优势,无论是总增长量,还是其各分量,东部各省份均位居全国前列,反观广大的中西部地区,虽然在十年的时间里,经济取得了一定的发展,但与东部地区的差距仍较大,特别是在产业结构方面,远远低于全国平均水平。

从总增长量看,总增长量排名全国前十位的省份中,除河南、四川两省外,均为东部省市,如广东、山东和江苏等省。而总增长量较低的地区主要集中在西部

地区,如西藏、青海、宁夏、贵州和甘肃等省份,特别是西藏、青海、宁夏三省,总增长量低于500,处于极低水平。

从份额分量看,排名全国前十位的省份与总增长量的排名基本一致,但北京市退出了全国前十位。另外,除广东、江苏、山东、浙江等东部沿海省份外,河南、河北、四川等人口大省也包含在内,而份额分量较低的省份依然主要集中在广大的中西部地区。

推进产业结构调整和优化升级,是转变经济增长方式、提高经济增长质量的重要途径和迫切任务。从产业结构偏离分量看,东部地区总体产业结构水平较高、产业优势明显,但就其内部而言,河北、山东、海南和福建四省产业结构仍需优化调整,而上海、广东和北京三省市产业优势较为明显。产业结构贡献小、劣势明显的省份主要有河南、湖南、安徽、四川、湖北和广西,值得关注的是排名后十位的省份中,中部省份占据了五席,其中,排名最后三位的河南、湖南和安徽三省均为中部省份,这表明虽然中部地区的总体发展水平好于西部地区,但产业结构不合理的问题却较为突出,制约着该区域今后的发展。

从竞争力分量看,竞争优势突出的省份仍然主要集中在东部地区,而对比中部地区和西部地区各省,西部地区的竞争优势明显高于中部地区,我国西部地区幅员辽阔、各类资源丰富,受开发程度低,劳动力价格低廉,特别是在国家西部大开发战略部署实施后,诸多因素使其较之东部地区具有更大的上升空间,更容易吸引外来投资、形成强劲的竞争力。在西部各省市中,内蒙古和陕西两个省份表现最好,在竞争力排名中位居前十。而竞争劣势明显的省份则主要为福建、重庆、云南、辽宁以及黑龙江等地区。

#### 四、模型建立前的准备

##### (一)若干假设

1. 本文通过人均国内生产总值来度量全国各区域的经济水平。
2. 造成全国各区域经济差异的因素很多,本文仅考虑国内投资和消费、区域开放程度、教育水平以及产业结构四大因素,并认为它们是影响区域经济差异最重要的因素。
3. 在模型建立的过程中,不考虑经济危机、政策变化等特殊因素或重大事件对经济产生的影响。
4. 本文在数据处理时不考虑价格因素对各经济指标的影响。
5. 在提取主成份的过程中,不考虑信息量损失对模型及预测结果的影响。

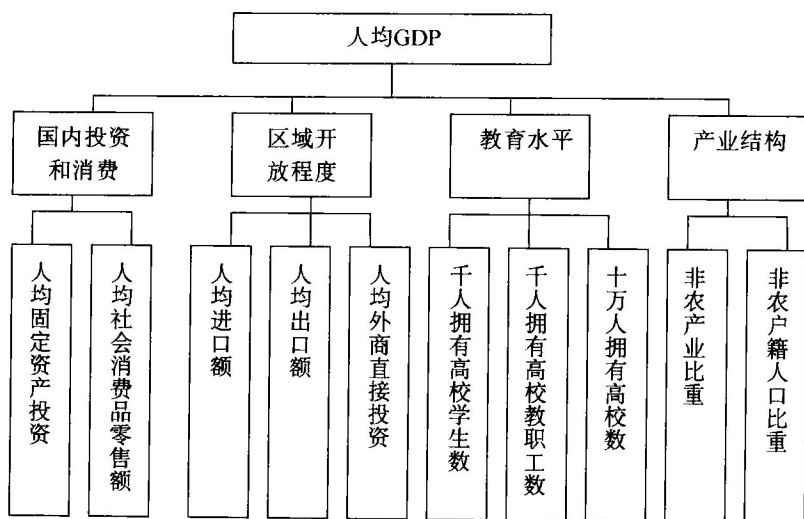


图1 衡量区域发展差异的指标体系图

## (二) 指标的选取

### 1. 人均 GDP

人均国内生产总值,也称“人均 GDP”。作为经济学中衡量经济发展状况的重要指标,它是人们了解和把握一个国家或地区宏观经济运行状况的有效工具。

### 2. 国内投资和消费

需求是拉动经济增长的重要因素。因此我们选择了反映投资需求的人均全社会固定资产投资和反映消费需求的人均社会消费品零售额两个指标。考虑到不同地区人口规模差异较大,单纯使用总量指标容易造成偏差,因此本文使用了平均指标而非总量指标。

### 3. 区域开放程度

根据历史经验,一个地区的开放程度越高,其经济发展就越快。进出口总额、外商直接投资(FDI)以及外贸依存度等指标都是反映对外开放程度的重要指标。为了便于计算并考虑到数据可得,本文选取了人均进口总额,人均出口总额以及人均外商直接投资这三个指标来衡量区域对外开放程度。

### 4. 教育水平

教育的发展水平特别是高等教育的发达程度对一个地区的经济发展至关重要,它不仅为区域经济的发展提供了重要的人才储备,同时也为区域层次的提升贡献了智力支持。因此,本文选择了千人拥有高校学生数、千人拥有高校教职工数和十万人拥有高校数等指标来反映教育水平的发展程度。



## 5. 产业结构

产业结构是推动经济增长的重要因素,高效率的产业结构是推动经济增长的根本动力。为了衡量产业结构对区域经济发展的作用,本文选取了非农产业比重和非农户籍人口比重两个指标。其中,非农产业比重是第二、三产业增加值占地区生产总值的比例,而非农户籍人口是非农业户口人数占户籍人口总数的比例。

### (三)数据的来源与预处理

#### 1. 数据来源的说明

在 20 世纪 90 年代,中国的经济体制发生了重大变革,市场经济的主导地位逐步确立。特别是从第九个五年规划开始,国家及时调整了经济的发展思路,着眼未来,构建了 2010 年远景目标,开始了全面完成现代化建设的第二步战略部署。从这个时期开始,我国的产业结构逐步优化,人民生活水平日益提高,科教兴国战略的实施使得科技和教育与经济发展更紧密的结合。可以说从九五开始我国经济发展步入了一个新的发展阶段。基于这些实际情况,结合数据的可得性,我们在建立模型时,选取了除西藏自治区外的 30 个省市自治区 1997 年—2006 年的数据进行研究,所用数据除“人均社会消费品零售额”通过国家统计局数据库数据计算得到外均来源于大赛提供的数据。

#### 2. 汇率因素影响的剔除

考虑到进出口额、FDI 等以美元计算的指标受汇率浮动影响很大,为了保证各时期数据可比,本文在构建模型时决定剔除汇率因素的影响,对以美元为单位计算的价值量指标使用历年(1997—2006 年)平均汇率换算成人民币。

#### 3. 数据的标准化处理

由于各指标的量纲、经济意义以及表现形式各不相同,在消除汇率变动因素影响后,必须对其进行标准化处理。经过标准化处理的数据消除了量纲影响,为接下来进行多种统计分析打下了基础。

#### 4. 数据的堆栈(Stacked)处理

由于本次大赛提供的数据均为非堆栈(Unstacked)格式,为了进行统计分析并建立 Pannal Data 统计模型,必须对其进行处理。考虑到数据从非堆栈(Unstacked)到堆栈(Stacked)的转换较为繁琐,我们使用 R 软件编写程序对数据进行转换,同时进行标准化处理。

## 五、区域经济差距分析模型的构建与检验

本部分使用主成份方法,将上文选取的指标综合为四大综合因子,作为后续模型的自变量。在此基础上,分别对全国 30 个省市(除西藏自治区)建立了个体