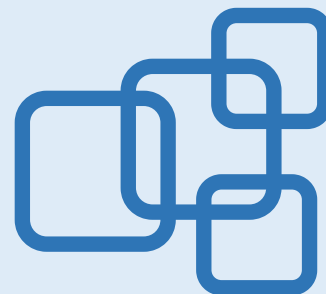


带你走入统计学

统计与管理学院 马俊玲



专题二 做一次抽样调查



专题二 做一次抽样调查

- 一、民意调查案例
- 二、抽样调查问题的思考

一、民意调查案例

（一）背景

美国总统选举的民意调查是调查数据的经典例子。

- 盖洛普民意调查

盖洛普是美国数学家，是抽样调查方法的创始人、民意调查的组织者，他几乎是民意调查活动的代名词。（现在的盖洛普公司——全球最著名的调查公司）

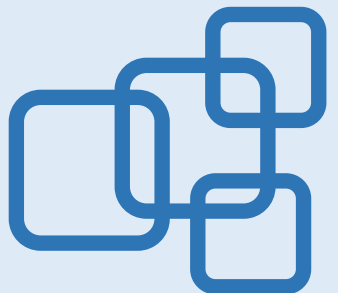


- 盖洛普民意调查

1932年，盖洛普的岳母作为民主党的候选人在爱荷华州竞选州务卿。在此之前，爱荷华州较高的公职一直都由共和党人保持，因而人们大都猜测他的岳母会落选。

但盖洛普没有凭猜测，而是运用他创立的方法进行科学的民意调查，结果发现爱荷华州选民对他的岳母的支持率超过对她的共和党对手的支持率。盖洛普于是预测说，他的岳母会获得选举的胜利。

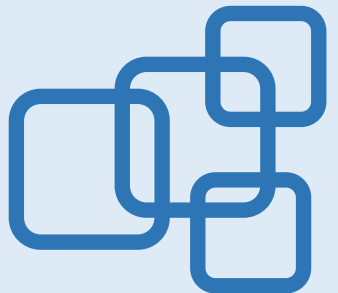
选举的结果证明了盖洛普应用科学的调查得出的预测是准确的。





- 盖洛普民意调查

这次民意调查就成为了政治史上第一次科学的民意调查。受这次成功的激励，盖洛普在**1935**年成立了盖洛普民意调查研究所，成为世界上第一个客观和科学的民意测验机构。





(二) 1936年美国大选

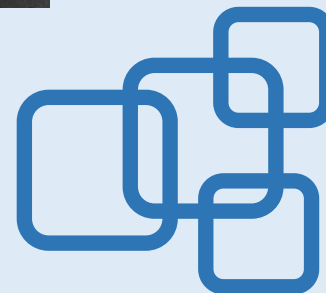
1936年民主党人罗斯福作为美国总统第一个任期满。共和党人兰登与他竞争总统。



富兰克林·罗斯福 (Franklin D. Roosevelt)
(1882.1—1945.4)



艾尔弗·兰登 (Alf Landon)
(1887.09-1987.10)

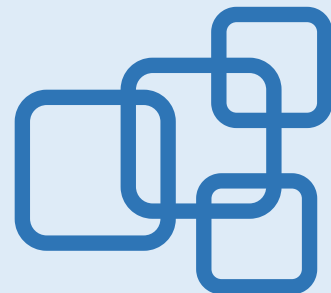




(二) 1936年美国大选

当时最为著名的民意调查机构，杂志《Literary Digest》曾在1936年以前的竞选中，进行民意调查5次准确预测了选举结果，因而具有一些声望。

《Literary Digest》的指导思想是样本越大，结果越准确。因此它发出去一千万份调查问卷，最后收到了237.6万份回复。结果显示，兰登将以压倒性的优势获胜。





(二) 1936年美国大选

《文学摘要》杂志，根据有约238万人参加的民意调查，预测：

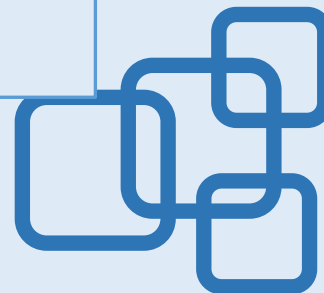
共和党人兰登的得票率为57%，
民主党人罗斯福的得票率为43%。

《文学摘要》杂志认为兰登
当选下一任总统。

1936年，盖洛普于一年前刚创立了他的民意调查机构。他根据一个约5万人参加的民意调查，预测

共和党人兰登的得票率为44%，
民主党人罗斯福的得票率为56%。

与《文学摘要》杂志针锋相对地，盖洛普认为罗斯福将连任下一届总统。

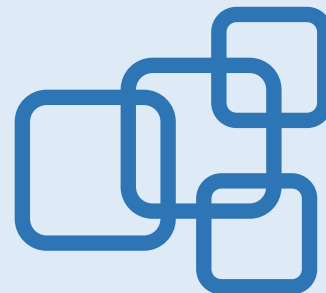




(二) 1936年美国大选

	罗斯福的得票率/%	误差/%
《文学摘要》杂志(样本238万)预测	43	-19
盖洛普(样本5万)预测	56	-6
实际选举结果	62	

《文学摘要》杂志依据238万人的样本所作出的预测居然有两位数误差，达到19%。如此之大的误差，让人惊异不已。





(三) 1948年美国大选

1948年，盖洛普联合全美约100家独立报纸，采用定额抽样的方法对5万人次进行寻访，预测杜威将战胜杜鲁门。

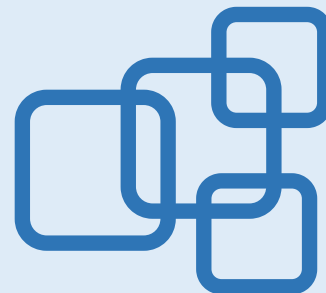
而结果是杜鲁门以52.8%的选票获胜，杜威的得票率为44.5%。





（四）案例思考

在这个美国总统选举的民意调查中，总体是什么？样本是什么？
为何《文学摘要》杂志的预测误差那么大？
定额抽样法为什么能成功，又为什么失败？



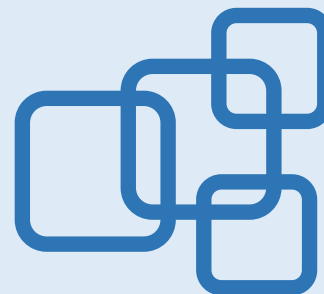


（四）案例思考

样本的代表性

《Literary Digest》调查：
1000万的调查对象大多数是从电话簿和俱乐部的花名册上选择。

盖洛普调查：
首次采用定额抽样（**quota sampling**）法，使用样本约**5万人**。



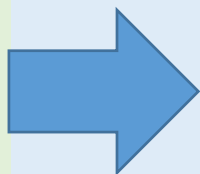


(四) 案例思考

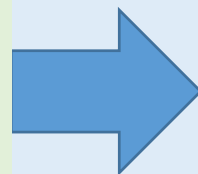
定额抽样法

盖洛普定额抽样法操作

1936年美国总统选举，盖洛普首先根据例如普查局的数据了解某地区的各项指标，例如性别、年龄、种族、住处和房屋租金等的分布情况。

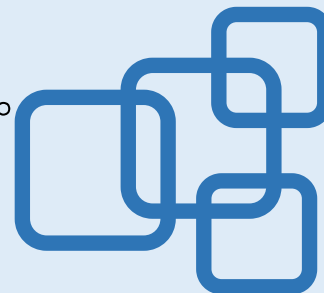


接下来根据该地区各项指标的分布情况按比例确定调查人数。



然后派调查员亲自去调查访问。

自1936年，定额抽样法成为早期美国总统选举民意调查采用的方法。





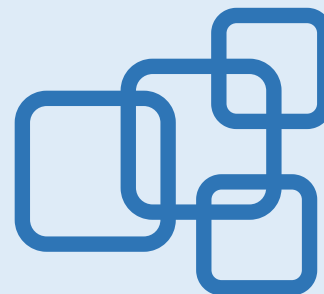
(四) 案例思考

定额抽样法

定额抽样法的目的是为了使得样本能代表总体，也就是使得被调查的5万人的意愿能代表美国全体选民的意愿。盖洛普面对全体选民，根据选民的类别按比例作调查。

由此看来，盖洛普的预测误差比《文学摘要》杂志的预测误差小得多是可想而知的。

然而1948年的调查预测失败，这是定额抽样法最著名的失败的例子。问题出在哪里？对方法如何改进？





(四) 案例思考

定额抽样法

例如盖洛普要求访问员在St. Louis地区访问13个对象，至于访问哪13个人不作规定，但要求必须满足下列条件：

6人住在近郊，7人住在市中心；

男性7人，女性6人；

7个男性中：

3人40岁以下，4人40岁及以上；

1个黑人，6个白人；

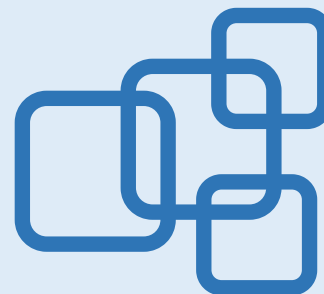
对女性有类似的定额规定；

对6名男性白人支付的房屋月租（这里指1948年的价格）有如下的规定：

1人支付的金额不少于44.01美元；

3人支付的金额在18.01美元与44美元之间；

2人支付的金额不多于18美元；





（四）案例思考

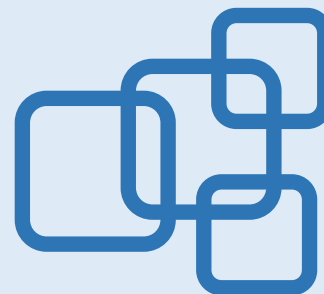
定额抽样法

这次预测失败的分析：

1. 定额抽样本身的不足；
2. 存在的主观性。

方法改进：

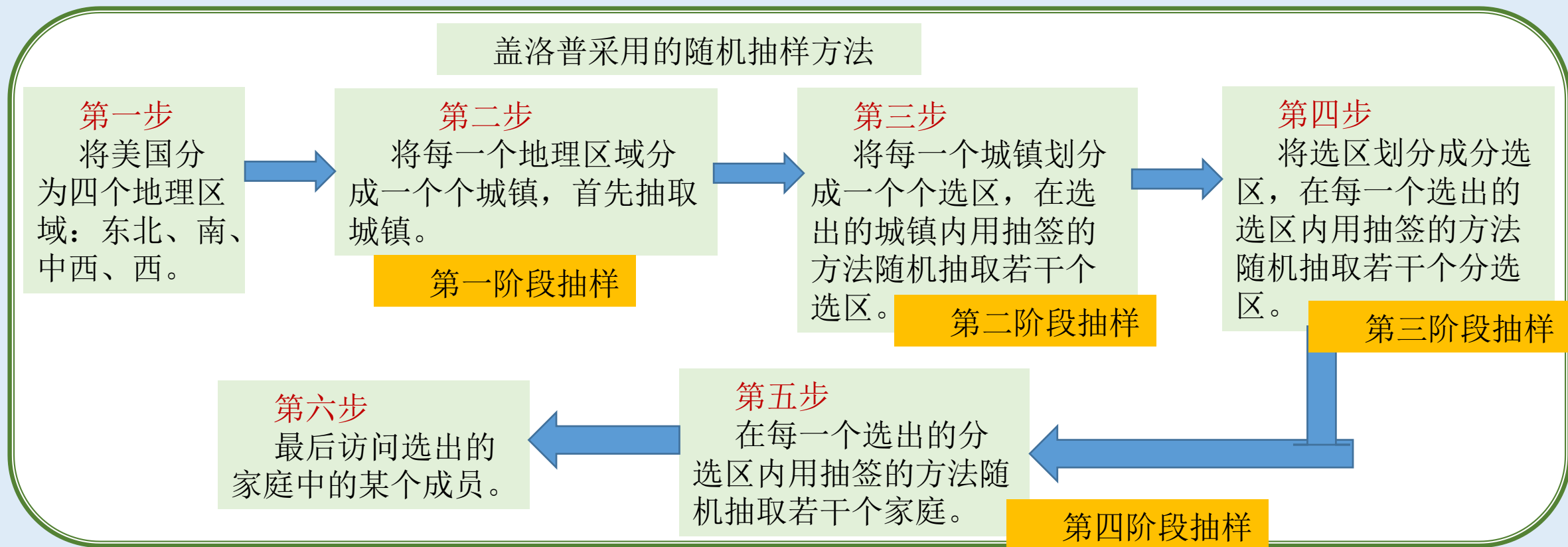
随机抽样调查。





(四) 案例思考

1952年之后的美国总统选举，盖洛普放弃定额抽样调查，改用**随机抽样调查**。





(四) 案例思考

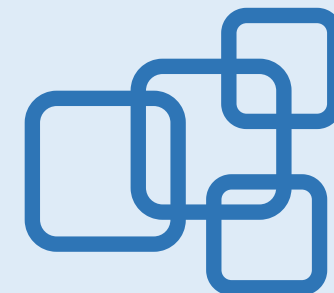
分层多阶段抽样方法

1948年以前，盖洛普民意测验采用的样本容量一般为5万。

之后对方法改进为分层多阶段抽样后，一般采用几千人的样本，而且精确度明显提高。

表1 1948年以后美国总统选举结果

年份	获胜总统	盖洛普预测得票率%	选举结果%	误差%
1952	艾森豪威尔（共和党）	51	55.4	-4.4
1956	艾森豪威尔（共和党）	59.5	57.8	1.7
1960	肯尼迪（民主党）	51	50.1	0.9
1964	约翰逊（民主党）	64	61.3	2.7
1968	尼克松（共和党）	43	43.5	-0.5
1972	尼克松（共和党）	62	61.8	0.2
1976	卡特（民主党）	48	50.1	-2.1
1980	里根（共和党）	47	50.8	-3.8
1984	里根（共和党）	59	59.2	-0.2
1988	布什（共和党）	56	53.9	2.1
1992	克林顿（民主党）	49	43.3	5.7
1996	克林顿（民主党）	52	50.1	1.9
2000	小布什（共和党）	48	47.9	0.1
2004	小布什（共和党）	49	51	-2
2008	奥巴马（民主党）	53	55	-2
2012	奥巴马（民主党）	58	61.7	-3.7
2016	特朗普（共和党）	48	55.7	-7.7



二、抽样调查问题的思考

（一）抽样调查研究涉及的问题

研究问题

研究对象

研究思路

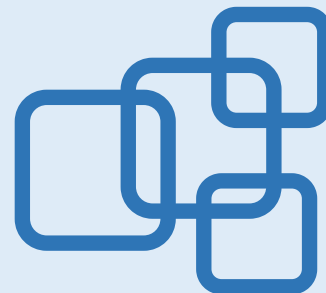


（二）抽样调查的案例

消费者信心指数调查

千村调查

中国综合社会调查（CGSS）



三、调查问卷设计

（一）设计问卷的注意事项：

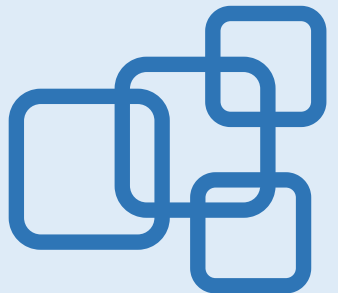
1. 明确问卷目的
2. 了解问卷对象
3. 选择问卷内容（相关性、完整性、简洁性）
4. 安排问卷结构（顺序性、分组性、平衡性）
5. 控制问卷格式（统一性、简洁性、适当性）
6. 选择问卷分析方法（有效性、科学性）



（二）问卷设计的步骤：

1. 确定问卷类型

确定问卷的类型。问卷的类型是指问卷的形式和方式，它可以分为纸质问卷、电子问卷、电话问卷、面对面问卷等。在确定问卷的类型时，要根据问卷的目的、对象、内容、成本、时间等因素进行综合考虑，选择最适合的问卷类型。例如，如果问卷的对象是分布广泛的消费者，可以选择电子问卷或电话问卷，以提高问卷的覆盖率和效率；如果问卷的内容是涉及敏感或复杂的问题，可以选择面对面问卷，以增加问卷的互动性和深入性。





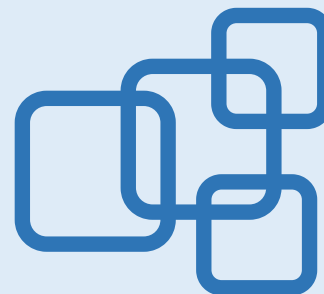
(三) 调查问卷设计

问卷设计的步骤：

2. 编写问卷题目

编写问卷的题目。问卷的题目是指问卷的具体问题，它可以分为单选题、多选题、填空题、判断题、评分题、排序题、开放题等。在编写问卷的题目时，要根据问卷的内容和目的，选择合适的题型和数量，避免使用过于简单或过于复杂的题目，以免影响问卷的有效性和可信度。同时，要注意以下几个方面：

- ◆ 避免引导性
- ◆ 避免双重否定
- ◆ 避免双重问题





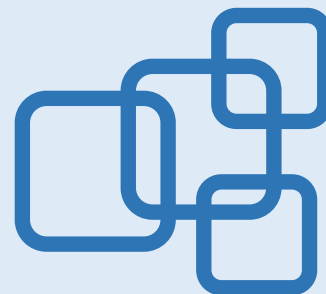
（三）调查问卷设计

问卷设计的步骤：

3. 设置问题选项

选项设置类型和数量都会影响问卷的敏感度和区分度。

- ◆ 一致性
- ◆ 完备性
- ◆ 互斥性





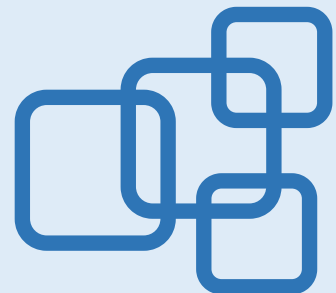
(三) 调查问卷设计

问卷设计的步骤:

4. 测试问卷效果

问卷的效果是指问卷的可行性和有效性，它可以通过预测试、正式测试、后测试等方式进行评估和改进。在测试问卷的效果时，要根据问卷的目的和对象，选择合适的测试方法和范围，避免使用过于简单或过于复杂的测试，以免影响问卷的可靠性和实用性。同时，要注意以下几个方面：

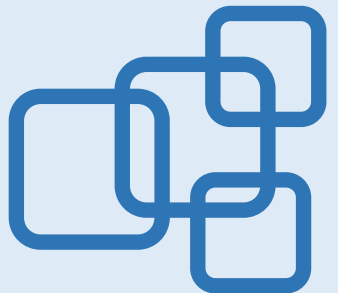
- ◆ 预测试（专家评审、小规模试验、认知访谈等）
- ◆ 正式调查
- ◆ 后测试（满意度、改进建议、效果评估等）





（四）设计问卷容易出现的问题

1. 问题所给答案不完整
2. 问题所给答案重叠
3. 问题所给答案与题干意义不符
4. 问题题干概念表述不明
5. 问题题干设置有诱导性
6. 问题设置不考虑被调查者背景
7. 问题设置不考虑数据的实际可得性
8. 问卷缺少控制性问题
9. 问卷缺少过滤性问题
10. 问卷结构不合理



四、讨论

(一) 千村调查问卷部分问题讨论

(二) 小组思考问题的讨论



思考与练习

1. 近几十年以来，我国每十年一次人口普查，五年一次1%人口抽查，请你设想下1%人口抽查如何进行抽样，可自己查找信息了解我国具体是怎么做的。

（可参考国家统计局网页

http://www.stats.gov.cn/tjzs/cjwtd/201308/t20130829_74325.html）

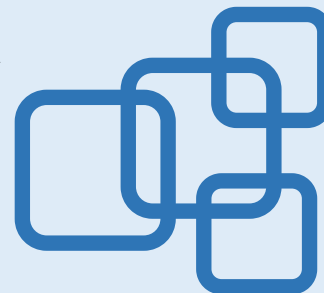
2. 国家统计局要定期公布相当多的数据，这些数据怎么获取的？国家统计局的抽样调查除了分层多阶段抽样方法，还有用别的什么方法吗？

（可参考国家统计局网页

如何获取统计数据<http://www.stats.gov.cn/tjzs/tjbk/201502/P020150508323847453391.pdf>）

统计词典<http://www.stats.gov.cn/tjzs/tjcd/>

3. 接着专题一思考与练习的问题，继续思考，确定一个你觉得值得研究的问题，比如你可以为了了解我校某方面的情况进行抽样调查研究，比如你觉得对全国人口、收入、消费等众多方面的某个问题有兴趣，等等。确定选题后可了解该问题相关研究现状，再思考如何获取数据并大致明确研究思路。





中华人民共和国国家统计局

National Bureau of Statistics of the People's Republic of China

走近统计

国家统计局
派驻纪检组
机构职能

统计数据

最新发布
数据查询
数据解读

统计工作

统计动态
通知公告
图片新闻

统计知识

统计百科
统计词典
常见问题解答

统计服务

网上办事
统计咨询
失信企业公示

信息公开

公开指南
公开规定
公开目录

统计数据

>> 返回首页

最新发布

| 更多

原油产量有所下降 天然气生产创新高

- 原煤生产恢复性增长 产业布局调整中优化 2018/03/19
- 电力生产稳步增长 电源结构调整优化 2018/03/19
- 解读：2018年2月份商品住宅销售价格延续总体稳定态势 2018/03/19
- 2018年2月份70个大中城市商品住宅销售价格变动情况 **解读** 2018/03/19
- 解读：制造业发展稳中向好 2018/03/16
- 国家统计局新闻发言人就2018年1-2月份国民经济运行情况... 2018/03/14
- 解读：消费品市场较快增长 2018/03/14

数据解读

| 更多

- 国家统计局城市司高级统计师刘建伟解读2月份商品住宅销售价格... 2018/03/19
- 解读：制造业发展稳中向好 2018/03/16

数据查询

- >> 月度数据
- >> 季度数据
- >> 年度数据
- >> 普查数据
- >> 国际数据
- >> 部门数据
- >> 可视化产品
- >> 中国统计年鉴

近期发布预告

发布日程表

- 3月26日 周一 9:30
流通领域重要生产资料市场价格变动情况
- 3月27日 周二 9:30
工业经济效益月度报告
- 3月31日 周六 9:00

统计公报

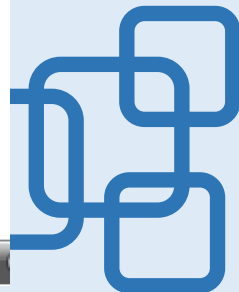
| 更多

- 年度统计公报
- 经济普查公报
- 人口普查公报
- 农业普查公报
- R&D普查公报
- 其他统计公报
- 基本单位普查公报
- 工业普查公报
- 三产普查公报

指标解释

| 更多

- 综合
- 国民经济核算
- 人口





THANK YOU!

