

# 软件工程统计方法

## 统计学导论

陈振宇

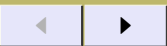
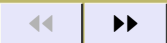
南京大学软件学院

Email: [zychen@software.nju.edu.cn](mailto:zychen@software.nju.edu.cn)

Homepage: [software.nju.edu.cn/zychen](http://software.nju.edu.cn/zychen)



日常生活中的统计  
统计学的产生  
统计学的发展  
统计学的应用  
基本概念  
调查与实验  
数据来源  
案例：扬子早报



第 1 页 共 100 页

返 回

全 屏

关 闭

# 内容提纲

- ❑ 统计学导论
- ❑ 描述统计
- ❑ 概率论基础
- ❑ 随机变量及其分布
- ❑ 统计量及其抽样分布
- ❑ 参数估计
- ❑ 参数假设检验
- ❑ 非参数假设检验
- ❑ 方差分析
- ❑ 相关性与回归分析



日常生活中的统计  
统计学的产生  
统计学的发展  
统计学的应用  
基本概念  
调查与实验  
数据来源  
案例：扬子早报



第 2 页 共 100 页

返回

全屏

关闭

# 本节内容

- 什么是统计学?
- 统计学的产生
- 统计学的发展
- 统计学的应用
- 基本概念
- 调查与实验
- 数据来源
- 案例分析



日常生活中的统计

统计学的产生

统计学的发展

统计学的应用

基本概念

调查与实验

数据来源

案例：扬子早报



第 3 页 共 100 页

返 回

全 屏

关 闭

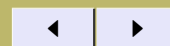


日常生活中的统计  
统计学的产生  
统计学的发展  
统计学的应用  
基本概念  
调查与实验  
数据来源  
案例：扬子早报

# 1 什么是统计学

Wiki：统计学是研究如何测定、收集、整理、归纳和分析反映客观现象总体数量的数据，以便给出正确认识的方法论科学，被广泛的应用在各门学科之上，从自然科学和社会科学到人文科学，甚至被用来工商业及政府的情报决策之上。

- ❑ 描述统计：给定一组数据，统计学可以摘要并且描述这份数据。
- ❑ 推断统计：观察者以数据的形态建立一个用以解释其随机性和不确定性的数学模型，通过样本和模型推断总体信息。



第 4 页 共 100 页

返 回

全 屏

关 闭



## 统计学的故事

在一家超市中，人们发现了一个特别有趣的现象：尿布与啤酒这两种风马牛不相及的商品居然摆在一起。但这一奇怪的举措居然使尿布和啤酒的销量大幅增加了。这可不是一个笑话，而是一直被商家所津津乐道的发生在美国沃尔玛连锁超市的真实案例。原来，美国的妇女通常在家照顾孩子，所以她们经常会嘱咐丈夫在下班回家的路上为孩子买尿布，而丈夫在买尿布的同时又会顺手购买自己爱喝的啤酒。这个发现为商家带来了大量的利润。

日常生活中的统计

统计学的产生

统计学的发展

统计学的应用

基本概念

调查与实验

数据来源

案例：扬子早报



第 5 页 共 100 页

返回

全屏

关闭



## 统计学的故事

统计学最早是在棒球大联盟中运用取得巨大成功, 比如教练场上的排兵布阵, 临场调度都参考统计数据. 如今统计专业人才也开始步入NBA俱乐部的决策层, 比如火箭刚聘请的经理就是个统计学专家, 他上任后的第一个任务就是利用他统计学的专业知识帮助火箭在选秀中选得优秀人才. 在比赛中的运用也越来越多. 比如我看到的最新的关于小牛的文章, 小牛场上的阵容经常根据对对方阵容的统计分析来决定的. 比如对太阳用跑跳更好的迪奥普而不用丹皮尔.

日常生活中的统计

统计学的产生

统计学的发展

统计学的应用

基本概念

调查与实验

数据来源

案例: 扬子早报



第 6 页 共 100 页

返回

全屏

关闭



## 2 日常生活中的统计

在日常生活中我们经常直接或者间接使用一些统计数据和统计结论。不论我们是否认识到这一点, 我们总是有意或者无意地使用某些统计推断方法去预测未来。

- 2009年全国的GDP为8.7%;
- 2009年南京居民收入超过了8000美金;
- 2009年全国房价没有泡沫;
- 刷牙能有效较少蛀牙;
- 第二个出生的小孩比第一个出生的小孩聪明;
- ...



### 3 统计学的产生

- ❑ 古希腊时期：“城邦政情”阶段始于古希腊的亚里斯多德撰写“城邦纪要”。
- ❑ 资本主义社会：开始出现专业的统计机构和研究组织，统计学正式产生。
- ❑ 统计学最初是作为政府以及管理阶层的工具。
  - ◆ 统计学的英文Statistics最早是源于现代拉丁文Statisticum Collegium (国会)以及意大利文Statista (政治家)。
  - ◆ 德文Statistik最早是代表对国家的资料进行分析的学问，也就是“研究国家的科学”。

目前统计学已被广泛应用在自然科学、医学、社会科学、人文科学、工业和商业决策。



日常生活中的统计  
统计学的产生  
统计学的发展  
统计学的应用  
基本概念  
调查与实验  
数据来源  
案例：扬子早报



第 8 页 共 100 页

返回

全屏

关闭



# 4 统计学的发展

## 1. 古典统计学

17世纪中叶至19世纪初叶；

代表学派: 政治算术学派, 1690年英国威廉·配第出版(政治算数)一书作为这个阶段的起始标志。

## 2. 近代统计学

18世纪末到19世纪末；

大数法则、最小平方法、相关与回归分析、以及正态分布等理论；

代表学派: 数理统计学派、社会统计学派。当时的统计学内容仍然是分析研究社会经济问题。

## 3. 现代统计学

1900年, 英国统计学家Karl Pearson发表了 $\chi^2$ 分布；

同时期, Francis Galton发展了回归分析；

1908年, William Sealey Gosset发表 $t$ 分布的论文；

二十世纪初, Ronald Aylmer Fisher发展了方差分析；

1930s, Jerry Neyman与Egon Pearson对假设检验理论作了系统的研究；

1934, Jerry Neyman发展了区间估计方法；

1940s, Frank Wilcoxon发展了非参数假设检验；

二战时期、二战后、...



日常生活中的统计

统计学的产生

统计学的发展

统计学的应用

基本概念

调查与实验

数据来源

案例: 扬子早报



第 9 页 共 100 页

返 回

全 屏

关 闭

# 5 统计学的应用

## 1. 企业发展战略

企业发展战略是一个企业发展的长远方向。统计学能够对一组准确的观察数据进行概括、整理得出结论用以支持企业营业经营管理决策的原理和方法。

## 2. 市场研究

对市场信息进行科学的统计分析, 得出有用的结论, 为生产和营销提供依据。

## 3. 财务分析

财务数据是公司发展决策的基础, 上市公司的财务数据是股民投资选择重要参考依据。

## 4. 产品质量管理

质量控制已经成为统计学在生产管理的一个重要应用。各种统计质量控制图被广泛应用于监测生产过程。

计算技术和一系列新技术、新方法在统计领域不断得到开发和应用。近几十年间, 计算机技术不断发展, 使统计数据的搜集、处理、分析、存贮、传递、印制等过程日益现代化, 提高了统计工作的效能。计算机技术的发展, 日益扩大了传统的和先进的统计技术的应用领域, 促使统计科学和统计工作发生了革命性的变化。如今, 计算机科学已经成为统计科学不可分割组成部分。随着科学技术的发展, 统计理论和实践深度和广度方面也不断发展。



日常生活中的统计

统计学的产生

统计学的发展

统计学的应用

基本概念

调查与实验

数据来源

案例: 扬子早报



第 10 页 共 100 页

返 回

全 屏

关 闭

## 6 基本概念

### □ 个体(Individual)

每一个研究对象称为个体。这里个体可以是一个人、一个企业、一个产品；也可以是一组人，一批产品等。

### □ 总体(Population)

总体是包含研究的全部个体的集合。总体范围有时比较容易确定，但有时较难确定。例如新推出一款3G手机，调查消费者是否喜欢。这个消费者群体通常比较难确定。总体分为有限总体和无限总体。

### □ 样本(Sample)

从总体中抽取部分个体构成的集合。经常说，来自(或取自)某总体的样本。

### □ 样本容量(Sample Size)

样本中所含个体的数目。

注意：样本具有二重性。样本是一组独立同总体分布的随机变量。



日常生活中的统计  
统计学的产生  
统计学的发展  
统计学的应用  
基本概念  
调查与实验  
数据来源  
案例：扬子晨报



第 11 页 共 100 页

返回

全屏

关闭

# 例子

为检验一批手机的质量,从中选择100部,则

- 总体: 这批手机(有限总体)
- 个体: 这批手机中的每一只
- 样本: 抽取的100部手机
- 样本容量: 100
- 样本值:  $x_1, x_2, \dots, x_{100}$

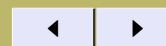
显然,可以选择“样本的函数”

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (1)$$

作为手机质量的一个评价指标。



日常生活中的统计  
统计学的产生  
统计学的发展  
统计学的应用  
基本概念  
调查与实验  
数据来源  
案例: 扬子早报



第 12 页 共 100 页

返 回

全 屏

关 闭

# 基本概念

## □ 变量(Variable)

变量是说明某种特征的概念,例如:销售额、考试成绩、产品寿命等都是变量。变量的具体取值称为变量值。例如考试成绩为80分,80就是变量值。统计数据就是统计变量的某些取值。

## □ 参数(Parameter)

参数是用来描述总体特征的概括性度量,用来表达总体的某种特征值。总体参数通常使用希腊字母表示:例如总体平均数 $\mu$ ,总体标准差 $\sigma$ 等。

## □ 统计量(Statistic)

统计量是用来描述样本特征的概括性度量。为了集中描述样本所带来的总体信息,采用特定的样本函数(不含任何未知参数)。这样的“不含未知参数的样本的函数”称为统计量。例如,关于样本平均值的统计量,记为 $\bar{X}$ ;关于样本标准差的统计量,记为 $s$ 。



日常生活中的统计  
统计学的产生  
统计学的发展  
统计学的应用  
基本概念  
调查与实验  
数据来源  
案例: 扬子早报



第 13 页 共 100 页

返回

全屏

关闭

# 7 调查与实验

## □ 调查

- ◆ 调查是取得直接统计数据的重要手段。为研究一些特定的社会问题或经济问题, 需要进行一些专门的调查。
- ◆ 常见的如市场调查机构进行的市场调查等。调查分为问卷调查、电话调查、网上调查等多种形式。

## □ 实验

- ◆ 实验设计是统计学的一个分支。实验设计研究如何制定实验方案, 以提高实验效率, 缩小随机误差的影响, 并使实验结果能够有效地进行统计分析的理论与方法。
- ◆ 实验设计的基本思想是减少偶然性因素的影响, 使实验数据有一个合适的数学模型, 以便使用方差分析或者其它统计方法对数据进行分析。
- ◆ 费希尔提出实验设计应遵循三个原则: 随机化, 局部控制和重复。



日常生活中的统计  
统计学的产生  
统计学的发展  
统计学的应用  
基本概念  
调查与实验  
数据来源  
案例: 扬子早报



第 14 页 共 100 页

返 回

全 屏

关 闭

## 8 数据来源

统计数据无外乎来自两种方式：调查或实验。从使用者的角度看, 统计数据来自于两种渠道：

### ❑ 直接获取

直接获取数据是指研究者通过调查方法和实验方法获取的第一手数据, 分别成为调查数据和实验数据。

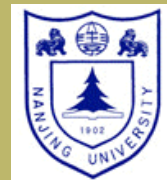
### ❑ 间接获取

如果研究内容相关的信息已经存在, 我们只是对这些信息进行加工整理, 成为我们进行统计分析的数据, 则称其为间接获取数据。

对于数据来源, 我们需要明确一下几个问题：

1. 数据获取的目的(Why)
2. 获取数据的形式(What)
3. 数据获取的时间(When)
4. 数据获取的方式(How)
5. 谁获取的数据(Who)

关于统计道德



日常生活中的统计  
统计学的产生  
统计学的发展  
统计学的应用  
基本概念  
调查与实验  
数据来源  
案例：扬子早报



第 15 页 共 100 页

返 回

全 屏

关 闭



## 9 案例：扬子早报

扬子早报是一份二十多年年发行历史的报纸，其主要新闻版块和栏目都深受读者喜爱，主要的发行区域是以江苏为中心的华东地区一带。扬子早报版面由最初的4开4版增至4开48版，并不定期加张免费附送，最多时日出120版。目前星期一至星期五的发行量为250 000份，而在星期六和星期天则为300 000份。扬子晚报在江苏直辖市和周边地区的上海、安徽等地共设有17个分印点，绝大多数读者能看到当天的报纸。通过对该地区一些出版物的收购和兼并，在过去的十年中这份报纸迅速成长起来。现在，报纸的管理层想通过增加收入和增加效率降低成本的方式继续扩展他们的事业。

本案例由《经理人Excel统计学》中的“春墅报”案例改编而成。



日常生活中的统计  
统计学的产生  
统计学的发展  
统计学的应用  
基本概念  
调查与实验  
数据来源  
案例：扬子早报

« »

◀ ▶

第 16 页 共 100 页

返回

全屏

关闭

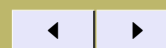




日常生活中的统计  
统计学的产生  
统计学的发展  
统计学的应用  
基本概念  
调查与实验  
数据来源  
案例：扬子早报

## 案例分析任务

章节	分析任务	相关技术
第二章	广告错误分析	统计图表, 数据汇总
第四章	投递改进与印刷改进分析	二项分布, 正态分布
第六章	订单估计	参数估计
第七章	电话营销分析	参数假设检验
第八章	电话营销分析	非参数假设检验
第九章	电话营销分析	方差分析
第十章	电话预约分析	回归分析



第 17 页 共 100 页

返 回

全 屏

关 闭



日常生活中的统计  
统计学的产生  
统计学的发展  
统计学的应用  
基本概念  
调查与实验  
数据来源  
案例：扬子早报

## 内容回顾

- 统计学的产生和发展
- 统计学的应用
- 基本概念：  
个体、总体、样本、变量、参数、统计量。
- 调查与实验
- 数据来源：  
直接获取、间接获取。
- 案例分析



第 18 页 共 100 页

返回

全屏

关闭