生物统计学

Biostatistics

张爱莲
分子生物学教研室
8583259（lab)
zal@xju.edu.cn

- 绪 论：任务、内容、学习要求
- 第一章　数据的收集与整理
- 第二章 概率分布
- 第三章 几种常见概率分布
- 第四章 抽样分布
- 第五章 统计推断
- 第六章 参数估计
- 第七章 拟合优度检验
- 第八章 单因素方差分析
- 第十章 一元回归及简单相关分析

一、统计学概论

统计学是一个学科，研究变异性、不确定性。

---- 对所有科学技术都重要，注定是一门多学科的科学

（一）统计学的定义

统计学是收集、分析、表述和解释数据的科学。是一套处理数据的方法和技术。

统计的过程

收集数据

↓

整理数据

↓

分析数据

↓

解释数据

（二）、统计学的要素

1、总体

- 总体：统计研究所确定的客观对象。
  - 范围
    - 无限总体：含无限多个单位
    - 有限总体：含有限个单位。
  - 特点
    - 同质性：总体中各个单位具有某种共同的性质。
    - 大量性：总体总是包含大量的单位。

2、样本 从总体中随机抽取部分单位所构成的集合体。

特点
- 必须取自所要研究的总体；
- 从一个总体中可抽取许多个样本，这些样本的数值是不同的，也即存在着随机的差异；
- 样本必须具有代表性；
- 样本必须具有客观性，也即排除主观因素的影响

3、推断：就是以样本信息为依据对总体的某些特征作出预测和估计。

4、推断的可靠性：根据随机抽取的样本对总体推断所得出的结果的可靠程度，也称置信度。

（三）、统计学的内容

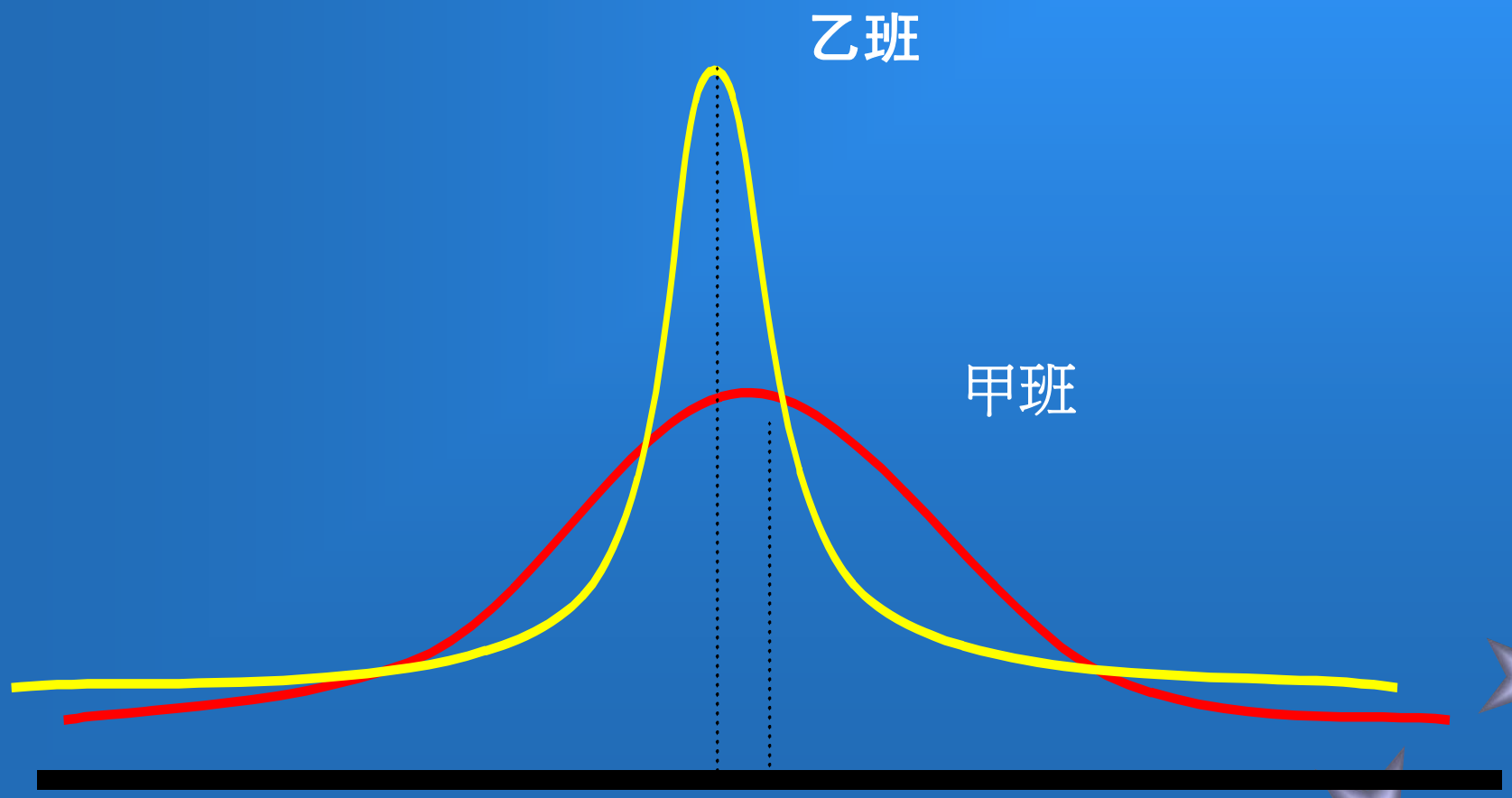
1、描述统计 ═ 指搜集、整理、分析并提供统计资料的理论和方法。

主要任务：使统计数据可以一目了然，条理清晰，使用方便，可以说明现象的数量特征和数量关系。

2、推断统计 ═ 是只依据样本资料推断总体特征的技术和方法，包括参数估计和假设检验的方法。

描述统计是推断统计的前提，推断统计是描述统计的发展。

- 某次生物统计学考试，甲班平均分为80分，标准差20，乙班平均分为75分，标准差5，哪一班成绩较优？



（四）、统计学科研的作用

- 统计学越来越不同于其它数学领域，计算机和信息科学工具至少和概率论一样重要。
- 统计学科研的作用是发展新的工具，应用于科学前沿。

统计的应用领域

Agriculture (农业）

Animal science(动物学）

Anthropology(人类学）

Archaeology(考古学）

Auditing(审计学）

Cristallography(晶体学）

Demograpohy(人口统计学 ）

Dentisity(牙医学）

Ecology(生态学）

Econometrics(经济计量学）

Education(教育学）

Engineering(工程）

Epidemiology(流行病学）

Finance(金融）

Fisheries research(水产渔业研究）

Genetics(遗传学）

Geography(地理学）

geology（地质学）

Management science(管理科学）

Marketing(市场营销）

Medical diagnosis(医学诊断）

Meteorology(气象学）

Military science(军事科学）

Physics(物理学）

Political science(政治学）

Psychology(心理学）

Psychophysics(心里物理学）

Socialogy(社会学）

二、 生物统计学的概念

生物学研究中通过观察、测量、实验可以得到许多数据，这些数据参差不齐，杂乱无章，必须对其进行统计分析，找出内在联系与规律。

生物统计学是关于生物学实验中所产生数据的处理和分析的一门科学。

- Biometrics，来自希腊文的两个词根，“Bio-”生命，“metron-”测量。它是用数理统计的原理和方法来分析和解释生物界的各种现象和规律的科学。
- 另一个常用英文词为Biostatistics.

三、生物统计学的任务

合理地进行调查或实验设计，科学地整理分析所收集得来的资料是生物统计的根本任务。

目的：通过对样本的研究来推断总体的性质。

核心问题：是总体与样本的关系问题。

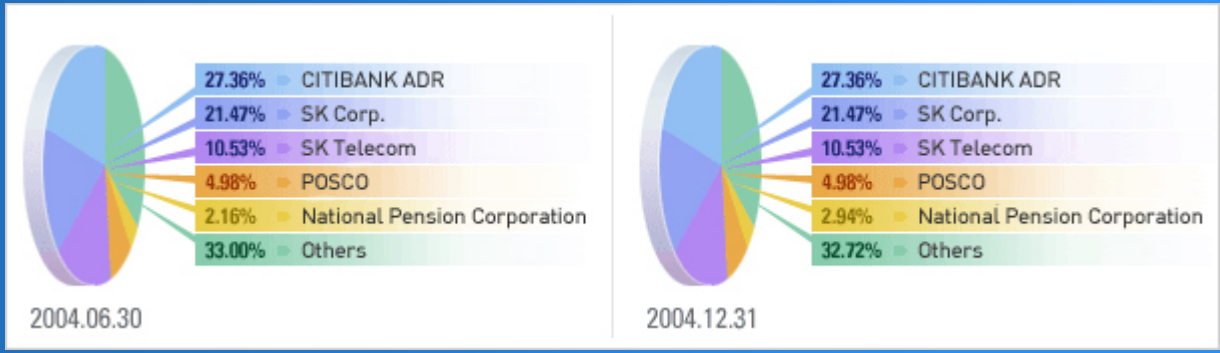
四、生物统计学的主要内容

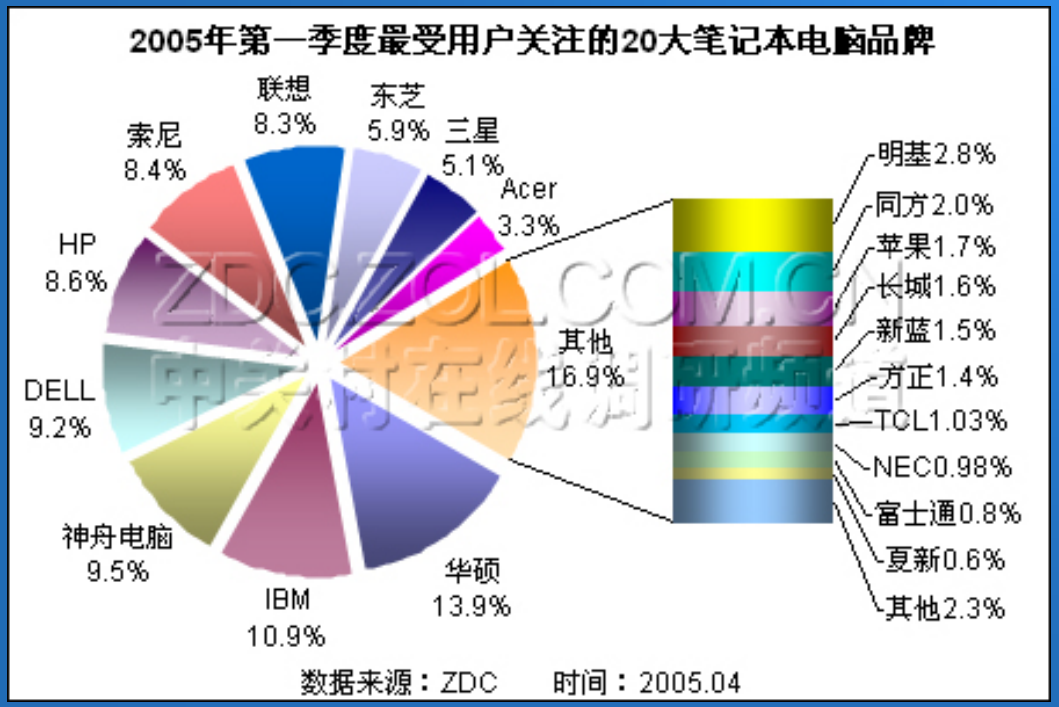
1、提供整理、分析资料的方法

整理资料的基本方法是根据资料的特性将其整理成统计表、绘制成统计图。

$\bar{X}$ ——— 集中性

S ——— 离散性


27.36% CITIBANK ADR
21.47% SK Corp.
10.53% SK Telecom
4.98% POSCO
2.16% National Pension Corporation
33.00% Others
2004.06.30
27.36% CITIBANK ADR
21.47% SK Corp.
10.53% SK Telecom
4.98% POSCO
2.94% National Pension Corporation
32.72% Others
2004.12.31


2005年第一季度最受用户关注的20大笔记本电脑品牌
联想 8.3%
东芝 5.9%
三星 5.1%
Acer 3.3%
索尼 8.4%
HP 8.6%
DELL 9.2%
神舟电脑 9.5%
IBM 10.9%
华硕 13.9%
其他 16.9%
明基2.8%
同方2.0%
苹果1.7%
长城1.6%
新蓝1.5%
方正1.4%
TCL1.03%
NEC0.98%
富士通0.8%
夏新0.6%
其他2.3%
数据来源：ZDC 时间：2005.04

Prefer unplanned travel activities
No answer 8%
Agree 45%
Not sure 26%
Disagree 21%

What restaurant diners want
Atmosphere 11%
Quick service 8%
Don't know 3%
Reasonable prices 22%
Great food 56%

2、统计分析 最重要的内容是 差异显著性检验。

显著性检验的目的就在于承认并尽量排除无法控制的偶然因素的干扰，将处理间是否存在本质差异揭示出来。

显著性检验的方法很多，常用的有：

t 检验——主要用于检验两个处理平均数差异是否显著；

方差分析——主要用于检验多个处理平均数间差异是否显著；

χ^2 检验 —— 主要用于由质量性状得来的次数资料的显著性检验等。

五、生物统计学的发展概况

1. 统计学的诞生与发展

- 数理统计最早的起源可追溯到十七世纪。

2. 生物统计学的起源

- 达尔文是最早应用生物统计方法的生物学家（1809-1882）。进化论的本质是生物统计。孟德尔（1822-1884）1866年的豌豆杂交试验结果也是一个生物统计问题。

3. 生物统计学的发展

重要发展在十九世纪以后：

- 1870年英国遗传学家Galton研究人类体高遗传，发现子女体高与父母体高有回归趋势，这就是“回归”术语的由来，后人推崇他为“生物统计和优生学之父”。
- Pearson在1899年提出了测量实际数与理论数之间偏离度的统计量。在属性统计分析中有着重要作用，同年他创建了《**Biometrika**》，并建立了一所数理统计学校。

4. 我国在统计学及生物学方面的工作

- 李悝（B.C.455-395） 《尽地力之教》中，运用平均数和分组法把收成分为丰年和饥年，丰年按上中下分组，饥年按大中小分组，计算平均产量。
- 沈秉诚1909年的《统计学纲领》，对我国早期统计学观点的形成和以后的发展都有影响。

5. 现代生物统计学

- 七十年代以来，由于电子计算机的日益普及，本来由于计算量太大而不得不放弃的统计问题获得了生命力，使其应用更为广泛。

六、国外统计学现状

- 在美国，现有86个统计学、生物统计学和生物计量学的博士点

- 很大一部分统计学家工作于统计系之外

  ----- 生物统计系、医学院、商学院、社会科学系

  ----- 政府机构、工业界（尤其是制药厂）

统计学的现状

- 科研合作的机会增多

  ---- 应用、多学科课题中统计学的错用严重

  ---- 由于大量数据的产生，人们主张统计人员进入研究团队

与未来统计学有关的生命科学的四个领域

-- 生物分子序列分析和功能基因组学

-- 遗传流行病学

-- 进化、群体遗传学和生态学

-- 计算神经科学

生物统计与生物信息

- 分子生物学已进化到信息科学，出现了新学科 Bioinformatics.

  -- 发现和实施算法，以了解生物学过程

  -- 主要工具：计算、统计

  这是统计科学家用武之地，尤其micro-array技术已经创造了新颖的统计学问题，激发了许多新的生物统计学研究。

- “统计学年轻人有足够的职业机会，学术界、工业部门、政府机构”。

  “在美国，很高比例的统计研究生出生于外国，毕业后留在美国”。

- “对生物统计学的需求从来没有象今天这样大，特别是美国。美国 National Research Council 研究结论： 在所有的卫生科研行业，最最缺乏生物统计学和流行病学人才”。（Zelen）

需求变了，训练必须变！

- 科学技术快速发展。可以预见，未来的生物统计学家需要生物信息学和遗传学方面的训练；更多注意数据分析，尤其是大型数据库的处理。(Zelen)

生物统计学的领导地位

- 可以预见，虽然生物统计学某些方面的领导地位目前在美国或英国，但将会逐渐被其它国家分享。新的生物统计学项目将在许多国家兴起，尤其是东方国家。（日本、中国大陆）。(Zelen)

七、生物统计学的学习方法与要求

1. 对自己要有一个积极的目标：

   一定要学好，一定能学好生物统计！

2. 数学知识要求不多，高中毕业水平就能学好。

3. 掌握基本概念原理，熟记公式方法的用途。大量作习题。生物统计是作为一门工具使用的，一定要会用。

4. 要会上课。课堂认真听讲，积极思考。

5. 课后按时完成作业。不允许抄袭。

平时成绩

每节课前以提问方式复习上节课内容

点名，考勤情况

每章后面作业情况

有不定时间的小测验，有期中小论文

平时成绩占30%，期末考试占70% 。