

# 《统计学》

## 图书基本信息

书名：《统计学》

13位ISBN编号：9787503767890

10位ISBN编号：7503767898

出版时间：2013-3-1

作者：吴喜之

页数：219

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：[www.tushu000.com](http://www.tushu000.com)

# 《统计学》

## 内容概要

《统计学--从数据到结论（第4版）》由浅入深地把统计最基本和最有用的部分在这么一本不厚的教科书中完整地介绍给读者，而且让读者可以边学习，边着手用统计软件处理数据。很少有一本统计教材包括像本书那么多的统计内容，本书每一章的主要部分是用日常语言来引进和解释一些概念，如果可能就通过例子来说明。如果不涉及应用，这部分就足够了。涉及应用的各章后面的小结中，有一部分是说明如何通过统计软件来处理本章的数值例子，这会给多数想要自己动手分析数据的读者以方便。小结的最后还展示了与概念及计算有关的一些数学公式，使那些精力充沛的读者能更深刻地理解内容。这种安排使得本教材能够适用于各种不同水平、不同要求的读者群体。本教材不仅可供没有学过概率论和数理统计的非统计专业的本科生和研究生使用，也可以供统计专业的本科生作为理解统计本来含义的教材使用（以代替不能满足需要的“描述统计学”等类课程），它还可以为各领域的广大实际工作者作为应用各种统计方法的参考书。

## 书籍目录

### 第一章 一些基本概念

- 1.1 统计是什么？
- 1.2 现实中的随机性和规律性，概率和机会
- 1.3 变量和数据
- 1.4 变量之间的关系
  - 1.4.1 定量变量间的关系
  - 1.4.2 定性变量间的关系
  - 1.4.3 定性和定量变量间的混和关系
- 1.5 统计、计算机与统计软件
- 1.6 小结
- 1.7 习题

### 第二章 数据的收集

- 2.1 数据是怎样得到的？
- 2.2 个体、总体和样本
- 2.3 收集数据时的误差
- 2.4 抽样调查和一些常用的方法
- 2.5 计算机中常用的数据形式
- 2.6 小结
- 2.7 习题

### 第三章 数据的描述

- 3.1 如何用图来表示数据？
  - 3.1.1 定量变量的图表示：直方图、盒形图、茎叶图和散点图
  - 3.1.2 定性变量的图表示：饼图和条形图
  - 3.1.3 其他图描述法
- 3.2 如何用少量数字来概括数据？
  - 3.2.1 数据的“位置”
  - 3.2.2 数据的“尺度”
  - 3.2.3 数据的标准得分
- 3.3 小结
  - 3.3.1 本章的概括和公式
  - 3.3.2 R语句的说明
- 3.4 习题

### 第四章 机会的度量：概率和分布

- 4.1 得到概率的几种途径
- 4.2 概率的运算
- 4.3 变量的分布
  - 4.3.1 离散随机变量的分布
  - 4.3.2 连续随机变量的分布
  - 4.3.3 累积分布函数
- 4.4 抽样分布、中心极限定理
- 4.5 用小概率事件进行判断
- 4.6 小结
  - 4.6.1 本章的概括和公式
  - 4.6.2 本章例题和R语句说明
  - 4.6.3 生成本章图形的R代码
- 4.7 习题

### 第五章 简单统计推断：总体参数的估计

## 5.1 用估计量估计总体参数

### 5.2 点估计

### 5.3 区间估计

#### 5.3.1 一个正态总体均值 $\mu$ 的区间估计

#### 5.3.2 两个正态总体均值之差 $\mu_1 - \mu_2$ 的区间估计

#### 5.3.3 总体比例 (Bernoulli 试验成功概率) 的区间估计

#### 5.3.4 总体比例 (Bernoulli 试验成功概率) 之差 $\pi_1 - \pi_2$ 的区间估计

### 5.4 关于置信区间的注意点

### 5.5 小结

#### 5.5.1 本章的概括和公式

#### 5.5.2 R 语句的说明

### 5.6 习题

## 第六章 简单统计推断：总体参数的假设检验

### 6.1 假设检验的过程和逻辑

#### 6.1.1 假设检验的过程和逻辑

#### 6.1.2 假设检验在前计算机时代发展的一些概念和步骤

### 6.2 对于正态总体均值的检验

#### 6.2.1 根据一个样本对其总体均值大小进行检验

#### 6.2.2 根据来自两个总体的独立样本对其总体均值的检验

#### 6.2.3 成对样本的问题

#### 6.2.4 关于正态性检验的问题

### 6.3 对于比例的检验

#### 6.3.1 对于总体比例的检验

#### 6.3.2 对于连续变量比例的检验

### 6.4 非参数检验

#### 6.4.1 关于非参数检验的一些常识

#### 6.4.2 关于单样本位置的符号检验

#### 6.4.3 关于单样本位置的 Wilcoxon 符号秩检验

#### 6.4.4 关于随机性的游程检验 (runs test)

#### 6.4.5 比较两独立总体中位数的 Wilcoxon (Mann—Whitney) 秩和检验

### 6.5 从一个例子说明“接受零假设”的说法不妥

### 6.6 小结

#### 6.6.1 本章的概括和公式

#### 6.6.2 R 语句的说明

### 6.7 习题

## 第七章 变量之间的关系；回归和分类

### 7.1 问题的提出

### 7.2 定量变量的线性相关

### 7.3 经典回归和分类

#### 7.3.1 一个数量自变量的线性回归

#### 7.3.2 多个数量自变量的线性回归

#### 7.3.3 自变量中有定性变量的线性回归

#### 7.3.4 Logistic 回归

#### 7.3.5 自变量为数量变量时的分类：经典判别分析

### 7.4 现代分类和回归：机器学习方法

#### 7.4.1 决策树

#### 7.4.2 关于组合算法

#### 7.4.3 Boosting

#### 7.4.4 随机森林

- 7.4.5 支持向量机
- 7.4.6 交叉验证比较各个模型
- 7.5 频数或列联表数据
  - 7.5.1 列联表数据及二维列联表的独立性检验
  - 7.5.2 高维列联表和多项分布对数线性模型
  - 7.5.3 Poisson对数线性模型
- 7.6 小结
  - 7.6.1 本章的概括和公式
  - 7.6.2 R语句的说明
- 7.7 习题
- 第八章 多元分析
  - 8.1 寻找多个变量的代表：主成分分析和因子分析
    - 8.1.1 主成分分析
    - 8.1.2 因子分析
    - 8.1.3 因子分析和主成分分析的一些注意事项
  - 8.2 把对象分类：聚类分析
    - 8.2.1 如何度量距离远近
    - 8.2.2 事先要确定分多少类：k均值聚类
    - 8.2.3 事先不用确定分多少类：分层聚类
    - 8.2.4 聚类要注意的问题
  - 8.3 两组变量之间的相关：典型相关分析
    - 8.3.1 两组变量的相关问题
    - 8.3.2 典型相关分析
  - 8.4 列联表行变量和列变量的关系：对应分析
  - 8.5 小结
    - 8.5.1 本章的概括和公式
    - 8.5.2 R语句的说明
- .....
- 第九章 随时间变化的对象：时间序列分析
- 第十章 生存分析简介
- 第十一章 指数简介
- 附录A练习：熟练使用R软件

## 精彩短评

- 1、非常简洁的统计学入门书。前面章节非常简单，后面的章节对我来说确实有点吃力。先大概过了一遍，慢慢再啃后面几个章节。
- 2、语言朴实，读起来很舒服
- 3、深入浅出，厘清了一些统计概念的盲点。
- 4、非常好的教材，薄薄的一本书，将统计学的方方面面都介绍了，像主成分分析、机器学习、时间序列分析都有介绍，同时附有R语言代码。  
另外序言推荐阅读，写的很好，不亏大师手笔
- 5、不错的书，但后面讲的乱七八糟，R代码给一大堆，结论给一大堆，啊！费劲啊.....
- 6、对于我这种近乎统计小白来讲，这本书大致让我理解了统计学的基本概念和思想，解答了我一直以来关于数据处理的部分疑惑，统计学入门必选书籍！
- 7、统计学教师不把统计背后的哲学原理阐释清楚的话，那可以说是咸鱼了。
- 8、前面挺好，后面几章就开始囫囵吞枣的

## 章节试读

### 1、《统计学》的笔记-第107页

一个模型存在的首要条件是可以很方便地计算。在前计算机时代，人们必须对数据做出许多主观假定，才能够进行对数据做基于模型的计算和判断，而且也只能处理少量数据。

人类能力的局限性使得完全准确的模型是不存在的。

经典的统计模型是可以数学公式描述出来，但是，人们有理由怀疑这些有限的公式对于描述复杂的自然和社会现象的可靠程度。随着计算机的发展，就产生了用计算机算法来确定的基于数据本身而不是数学假定的模型，模型也就变得越来越复杂，可处理的数据量也越来越大，这些模型包括机器学习或数据挖掘领域所使用的众多的模型。

---

#这里啰嗦一小段#

大家会经常听到“假设检验”这个名词，同时也会听到“显著性”这个名词，到底是啥嘞？

我们拿到一组数据，看了个大概，有了个模模糊糊的感觉，接下来就是要检验下这个感觉是不是正确的。假设检验或者说显著性检验，就是在“待检验假设为真”的情况下用来计算以往观测数据发生的概率。当这个概率很低时，我们默认是无法得到这些数据的，也就是说我们的前提条件是有问题的——假设不成立。

在这个概念的早期，“significant”这个词只是用来指“概率低到可以被拒绝的程度”，如果数据能够用来拒绝某个假设，那么它就是显著的。所以“significant”的现代语义往往会使得我们感到迷惑。不得不说天下教科书大部分都是我想当然，然后大家任意抄。

这里注意，假设检验只是提供了一种“拒绝假设的途径”，不能因为这个假设和已有数据之间没有矛盾就接受这个假设。例如：当选定显著性水平为0.1时，即假设为真前提下出现已有数据的概率低于10%的话，我们就拒绝假设。

再啰嗦一点，何为p值呢？它也是一种概率——在假设为真的情况下你拒绝原假设的概率有多大？如果这个值非常小，也就是说你可以放心拒绝，因为犯错的概率很低嘛~

---

#线性相关#

1. Pearson's correlation coefficient

在数据来自正态总体的假设下，有相关系数的检验（r是否为0）

2. Kendall's T

不用假设总体的分布，Kendall的相关系数是一个非参数的度量，不依赖变量背后的总体分布。

3. Spearman rank correlation coefficient

---

来看组数据：我们看下这8个例子啊~ 第一幅图假设检验的结果告诉我们要拒绝“线性无关”的假设，也就是它们是

# 《统计学》

## 版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:[www.tushu000.com](http://www.tushu000.com)