

# 概率统计(A)

Probability and Statistics(A)

主讲教师：郑旭玲（智能系）

E-mail：xlzheng@xmu.edu.cn

Tel：13306005915

QQ：2252291

助教：曹中伟（通信21级研究生）

QQ：838822657

**SPOC:** <https://www.icourse163.org/spoc/course/XMU-1450321504>

(选课密码：PTMS2023inf)



下载App



关注微信

概率统计A信院郑老师

22232PTMS信院zxl 8C4MZ/W



请确保与SPOC使用的是同一账号！



群名称:PTMS2023

群 号:720014776

# 课程简介



- ◆ **课程类别：**公共基本必修课程
- ◆ **先修课程：**高等数学或数学分析、线性代数或高等代数、集合论与图论
- ◆ **内容重点：**概率论与数理统计的基本理论、方法及实际应用（在计算机科学中的具体应用）。

# 教材与参考书

## 教材 (Text Book)

- ◇ 《概率论与数理统计》第五版/第四版，高等教育出版社，盛骤、谢式千、潘承毅

## 参考书目 (Reference)

- ◇ 《概率统计》第三版，同济大学出版社，同济大学概率统计教研组
- ◇ 《PROBABILITY AND RANDOM PROCESSES FOR ELECTRICAL AND COMPUTER ENGINEERS》Cambridge University, John A. Gubner.
- ◇ .....

# 第五版章节及主要内容

- ◆ 第一章 概率论的基本概念
- ◆ 第二章 随机变量及其分布
- ◆ 第三章 多维随机变量及其分布
- ◆ 第四章 随机变量的数字特征
- ◆ 第五章 大数定律和中心极限定理
- ◆ 第六章 数理统计的基本概念
- ◆ 第七章 参数估计
- ◆ 第八章 假设检验
- ◆ 第九章\* 方差分析及回归分析
- ◆ 第十章\* bootstrap方法（自助法）
- ◆ 第十一章\* 在数理统计中应用R软件
- ◆ 第十二章\* 随机过程
- ◆ 第十三章\* 马尔可夫链
- ◆ 第十四章\* 平稳随机过程
- ◆ 第十五章\* 时间序列分析

概率论部分

数理统计部分

随机过程部分





## 二、培养目标

本课程侧重于讲解概率论与数理统计的基本理论与方法，同时在教学中结合信息工程专业的特点介绍一些具体应用，培养目标在于要求学生初步掌握处理随机现象的基本理论和方法，培养学生解决相关实际问题的能力，并且为进一步深入学习打下坚实的基础。通过本课程的学习，要求学生能够理解随机事件、样本空间与随机变量的基本概念，掌握概率的运算公式，常见的各种随机变量（如 0-1 分布、二项分布、泊松（Poisson）分布、均匀分布、正态分布、指数分布等）的表述、性质、数字特征及其应用，一维随机变量函数的分布、二维随机变量的联合分布、边缘分布及条件分布及其函数的分布。理解数学期望、方差、协方差与相关系数的本质涵义及相关性质，熟练运用各种计算公式。了解大数定律和中心极限定理的内容及应用，熟悉数据处理、数据分析、数据推断的各种基本方法。

## 三、教学方法

本课程具有很强的应用性，在教学过程中要注意理论联系实际，从实际问题出发，通过抽象、概括，引出新的概念。由于本课程是研究随机现象的科学，学生之前从未接触过，学习起来会感到难度较大，授课时应突出重点，讲清难点。要使学生明白，本课程主要研究哪些方面的问题，从何角度、用何原理和方法进行研究的，是怎样研究的，得到哪些结论，如何用这些方法和结论处理今后遇到的社会经济问题。在教育中要坚持以人为本，全面体现学生的主体地位，教师应充分发挥引导作用，注意随时根据学生的理解状况调整教学进度。授课要体现两方面的作用：一是为学生自学准备必要的理论知识和方法，二是激发学生学习兴趣，引导学生自学。在教学中要体现计算机辅助教学的作用，采用多媒体技术，提高课堂教学的信息量。每次课后必须布置较大数量的思考题和作业，并加强课外辅导和答疑。

四、主要内容及学时安排		
章（或节）	主要内容	学时安排
第一章 概率论的基本概念	1、随机试验、随机事件与样本空间。 2、事件的关系与运算、完全事件组。 3、概率的概念、概率的基本性质、概率的基本公式。 4、等可能概型（古典概型）、几何型概率。 5、条件概率、全概率公式、贝叶斯公式。 6、事件的独立性、独立重复试验。	8
第二章 随机变量及其分布	1、随机变量及其分布函数的概念及其性质。 2、离散型随机变量及其分布律。 3、连续型随机变量及其概率密度。 4、常见随机变量的概率分布。 5、随机变量的函数分布。	6
第三章 多维随机变量及其分布	1、二维随机变量及其概率分布。 2、二维离散型随机变量的概率分布、边缘分布和条件分布。 3、二维连续型随机变量的概率密度、边缘密度和条件密度，常用二维随机变量的概率分布。 4、随机变量的独立性和相关性。 5、两个随机变量函数的分布。	10
第四章 随机变量的数字特征	1、随机变量的数学期望（均值）、随机变量函数的数学期望。 2、方差、标准差及其性质，切比雪夫（Chebyshev）不等式。 3、协方差、相关系数及其性质。 4、矩、协方差矩阵。	8
第五章 大数定律及中心极限定理	1、依概率收敛，伯努利大数定律、辛钦（Khinchine）大数定律。 2、中心极限定理，包括独立同分布的中心极限定理、李雅普诺夫（Lyapunov）定理、棣莫弗—拉普拉斯（De Moivre—Laplace）定理。	4

第六章 样本及抽样分布	1、总体、个体、简单随机样本、统计量、样本均值、样本方差和样本矩。 2、直方图和箱线图。 3、 $\chi^2$ 分布、t分布和F分布，分位数，正态总体的常用抽样分布。	6
第七章 参数估计	1、点估计的概念、估计量与估计值。 2、矩估计法、最大似然估计法。 3、估计量的评选标准。 4、区间估计的概念。 5、单个正态总体的均值和方差的区间估计。 6、两个正态总体的均值差和方差比的区间估计。 7、(0-1)分布参数的区间估计。 8、单侧置信区间。	8
第八章 假设检验	1、假设检验。 2、正态总体的均值的假设检验。 3、正态总体方差的假设检验。	6
习题讲解、复习及答疑		6
合计		64



# 教学进度表

周数 (按校历)	各章节教学内容纲要	课时安排(学时)					主讲人	备注
		授课	讨论	实验/ 上机	实践	其他		
1	第一章 概率论的基本概念	8						作业4次
2								
3	第二章 随机变量及其分布	8						作业4次
4								
5	第三章 多维随机变量及其分布	10						作业5次
6								
7	第四章 随机变量的数字特征	8						作业4次
8								
9	期中复习	4						期中考试
10	第五章 大数定律及中心极	4						作业2次
11	第六章 样本及抽样分布	6						作业3次
12								
13	第七章 参数估计	8						作业4次
14								
15	第八章 假设检验	6						作业2次
16	期末复习	2						
	期末考试							期末考试
	合计	64						

# 教学方法——线上线下混合式

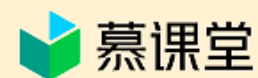
## ◆ 课前

- ◆ 基于SPOC平台的自主学习



## ◆ 课堂

- ◆ 线下课：基于传统课堂的主导学习、基于翻转课堂的互动学习
- ◆ 线上课：基于SPOC平台及厦大云盘的自主学习



## ◆ 课后

- ◆ 基于SPOC平台的自主练习



# 中国大学MOOC平台SPOC课程

◆ <https://www.icourse163.org/spoc/course/XMU-1450321504>

开课时间：2023年02月12日 ~ 2022年06月30日

选课密码：PTMS2023inf



**【注意】慕课堂小程序、中国大学MOOC电脑端、中国大学MOOC手机app端  
都需要用同一个账号登录。**

# 线上线下课时安排

2022-2023 学年第2学期 《概率统计A》 教学日历

2023年2月

日	一	二	三	四	五	六
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28				

2023年3月

日	一	二	三	四	五	六
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

2023年4月

日	一	二	三	四	五	六
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30						

2023年5月

日	一	二	三	四	五	六
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

2023年6月

日	一	二	三	四	五	六
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	

线下课（1号楼C204）

线上课（SPOC自主学习）

数字

法定假节日

# 线上课（SPOC自主学习）流程

1. 厦大智慧教务微信小程序签到（周三10:10~10:40；周五8:00~8:30）
2. 温故：复习上次课知识点及重难点，结合SPOC、厦大云盘等平台上发布的课件资源等查缺补漏、订正错题；
3. 知新：结合SPOC、厦大云盘等平台上发布的视频、课件等学习资源，自主完成本次课程的学习任务，并完成配套的慕课堂练习及SPOC作业；
4. 在线答疑

其中，2~4均不限于本次课时之内，但应在下次课前尽早完成






## 作业 (Homework Assignments)

### 要求:

1. 规范、简洁。

2. **完成要求:** 课后的所有习题，时间允许尽可能都做一遍，没有时间可以根据自己的情况选做。

 第一注重作业的质量，其次才是数量。

 **电子版:** 写在纸上，拍照SPOC提交，  
下次课前的24时截止

 **纸质版:** 及时订正，每章结束后提交一次

# 公告、作业、单元测试、讨论区

厦门大学

课程 ▾

学校

慕课堂



概率统计A信院郑老师  
郑旭玲

SPOC | 学校专有课程



公告



评分标准



课件



测验与作业



考试



讨论区

## 测验与作业

[查看帮助](#)

第0周 组合分析 (请在正式开课前完成)

▼ 第0周单元测试 截止时间: 2023/03/06 00:00

[前往测验](#)

共6个填空题, 试题由系统随机抽取。你有3次答题的机会, 取其中的最高分为单元测试的成绩。请注意截止提交时间。

截止时间 2023/03/06 00:00

请务必在截止时间之前提交, 截止时间后的提交不再计分

有效分数 0.00/24.00

你的每一次测验系统都将为你计分, 并提取最高得分作为你的有效分数

有效提交次数 0/3



## 课程评分方法 (Grading Policies)

- **去年**

平时成绩30%+期中成绩占40%+期末成绩占30%

- **前年**

平时成绩20%+期中成绩占40%+期末成绩占40%

- **大前年**

平时成绩20%+期中成绩占30%+期末成绩占50%

## 平时成绩统计设置

记分模块	占比 (%)
线上部分	<input checked="" type="checkbox"/> 线上课程成绩 70
	<input checked="" type="checkbox"/> 视频学习完成度 10
课堂活动	<input checked="" type="checkbox"/> 练习 10
	<input checked="" type="checkbox"/> 讨论 参与讨论的个数占总讨论数的百分比*100 10
课堂加分项	<input checked="" type="checkbox"/> 讨论优秀发言 获得优秀发言次数加分设置 <input checked="" type="radio"/> 加分后总分封顶100分

线上成绩	参与计分类型	总分占比(%)
<input checked="" type="checkbox"/> 单元测验	<a href="#">查看所有</a>	30
<input checked="" type="checkbox"/> 单元作业	<a href="#">查看所有</a>	60
<input checked="" type="checkbox"/> 课程讨论	设置积分维度 <input checked="" type="radio"/> 活跃度: 请输入获取满分用户需要在“课堂讨论”中回复的数量 20	10

# 概率论与数理统计概述





# 自然界与社会生活中的现象

1. **确定性现象**：在一定条件下必然发生的现象。
2. **不确定性现象**：在一定的条件下，可能出现这样的结果，也可能出现那样的结果，在试验或观察之前不能预知确切的结果。

例如：

- ◆ 太阳从东方升起；
- ◆ 明天的最高温度；
- ◆ 上抛物体一定下落；
- ◆ 新生婴儿的体重……

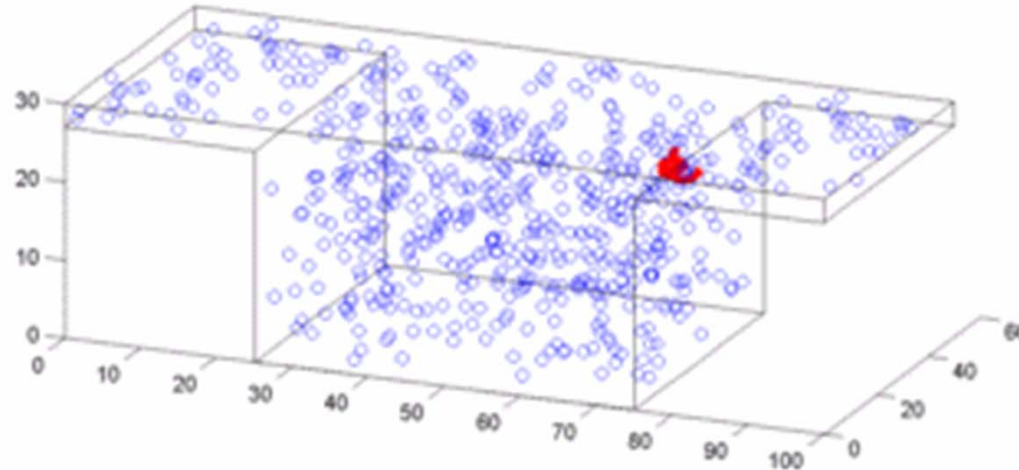


# 在我们所生活的世界上， 充满了不确定性

- ◆ 从扔硬币、掷骰子和玩扑克等简单的机会游戏，到复杂的社会现象；从婴儿的诞生，到世间万物的繁衍生息；从流星坠落，到大自然的千变万化……，我们无时无刻不面临着不确定性和随机性。



# 不确定性现象的特点



- ◆ 【特点 1】当人们在一定的条件下对不定性现象加以观察或进行试验时，观察或试验的结果是多个可能结果中的某一个。而且在每次试验或观察前都无法确知其结果。
- ◆ 【特点 2】不确定性现象在大量重复观察或试验下，它的结果却呈现出固有规律性。

统计规律性

# 随机现象

- ◆ 在个别试验中其结果呈现出不确定性，在大量重复观察或试验中其结果却具有统计规律性的现象，称为**随机现象**。





# 概率论 VS 数理统计

- ◆ **概率论**——以数量化方法(即以数学为研究工具)来研究随机现象及其规律性的一门学科。
- ◆ **数理统计**——为了研究某个问题，它（运用概率论和数学的方法）分析怎样**收集**与问题相关的数据，如何**整理分析**这些数据，最后对所研究的问题做出**推断**。
  - ◆ 概率论是数理统计学的理论基础，而数理统计则是概率论的重要应用。
  - ◆ 概率论从母群体中推导出样本的概率，而统计推论则正好相反——从小的样本中得出大的母群体的参数。



# 概率统计 VS 数学

- ◆ 概率论与数理统计是一门与数学密切相关的学科，但严格来说不是数学学科。
  - ◆ 概率统计的研究对象是随机现象，分析过程需要用到数学的证明与推导，结论形式是事件的概率；
  - ◆ 数学的研究对象是数量，分析过程是抽象的逻辑推理，结论形式是定理推论。
    - ◆ 好比国家与军队的关系，国家不能没有军队，但国家不等于军队，概率统计不能没有数学，但不是数学，数学只是作为一种描述刻画随机现象本质的工具被应用在概率统计中。
- ◆ 虽然概率统计本质不是数学，但它的任何计算与证明都是数学的东西，所以多年来学术界都把它归入数学类，这已经成为习惯。

# 概率论源于赌博？

- ◆ 15世纪末到16世纪初，赌博在欧洲盛行，于是有些数学家开始着手探讨赌博中出现各种情况的机遇或胜率，赌博中的这些计算方法，后来演变成了概率的古典定义。
- ◆ 1654年，巴黎一个名叫梅雷的赌徒要求当时著名的数学家帕斯卡解决“赌注分配问题”



# 赌注分配问题



- ◆ 例如：甲、乙二人赌博，各出赌注30元，共60元，每局甲、乙胜的机会均等，都是 $1/2$ 。  
约定：谁先胜满3局则他赢得全部赌注60元，现已赌完3局，甲2胜1负，而因故中断赌情，问这60元赌注该如何分给2人，才算公平？
  - ◆ 按2：1分配，即甲得40元，乙得20元？
  - ◆ .....
  - ◆ 正确的分法应考虑到如在这基础上继续赌下去，甲、乙最终获胜的机会如何
    - ◆ 至多再赌2局即可分出胜负，这2局有4种可能结果：甲甲、甲乙、乙甲、乙乙。前3种情况都是甲最后取胜，只有最后一种情况才是乙取胜，二者之比为3：1，故赌注的公平分配应按3：1的比例，即甲得45元，乙15元。
  - ◆ 引出了期望的概念，也就是现在常说的预期。



# 概率论源于赌博？

- ◆ 1713年，在J·伯努利(J·Bernouli)去世后的8年，他的著作《推测术》问世，书中提出了伯努利大数定律

- ◆ （关于算术平均值的定理）  
算术平均值是最常用的一种统计方法，人们经常使用并深信不疑。但其理论根据何在，并不易讲清楚。这实际上就是伯努利的大数定律要回答的问题，在某种程度上可以说，这个大数定律是整个概率论最基本的规律之一，也是数理统计学的理论基石。



- ◆ 首次出现了客观概率和主观概率的概念区分，“标志着概率概念漫长地形成过程的终结与数学概率论的开端”

# 概率论的发展

- ◆ 棣莫佛(DeMoivre)提出的**中心极限定理**成为后世自然科学和社会科学中处理各种观察值误差的理论基础；
- ◆ **贝叶斯理论**运用概率来解决从特殊推断到一般的问题，其理论中的先验概率的主观解释成为与经典的频率概率解释并行的两种理解方式，这种现象把概率问题提高到哲学讨论的范围；
- ◆ **拉普拉斯《概率分析理论》**问世，概率论从17世纪到19世纪初的经典时期才被划上了一个完整的句号，开始其近代征程。
- ◆ 1933年，苏联数学家**柯尔莫哥洛夫**完成了**概率论的公理体系**。
  - ◆ 在几条简洁的公理之下，发展出概率论整座的宏伟建筑，有如在欧几里得公理体系之下发展出整部几何。





# 新飞跃：与统计学的联姻

- ◆ 拉普拉斯发现：“概率的数理公式可以当作以大量观察而又易于有错误为基础的各項科学所要的辅助科学”。
- ◆ **高斯**基于对误差理论的分析，提出了**高斯分布曲线**让概率与统计的结合迈出了重要的一步。
- ◆ 麦克斯韦、玻尔兹曼将高斯由概率理论所导出的误差应用到速度分布中去，从而和吉布斯一起，把统计力学扩展到热理论和广义气体理论。
  - ◆ 将物质宏观的确定性的温度、熵等概念与微观的不确定性的分子运动统一起来，使概率统计成为了宏观与微观的桥梁，确定性与不确定性的中介。



# 统计学的起源

- ◆ 统计学的初衷是作为政府以及管理阶层的工具。
  - ◆ 在拉丁语、意大利语、德语等语言中，统计学（statistics）的词根都来自国家（state），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”。
  - ◆ 英国学者葛朗特在1662年发表的著作《关于死亡公报的自然和政治观察》，标志着这门学科的诞生。



# 葛朗特的《关于死亡公报的自然和政治观察》

- ◆ 中世纪欧洲流行黑死病，死亡的人不少。自1604年起，伦敦教会每周发表一次“死亡公报”，记录该周内死亡的人的姓名、年龄、性别、死因。以后还包括该周的出生情况（依据受洗人名单）。几十年来，积累了很多资料。
- ◆ 葛朗特是第一个对这一庞大的资料加以整理和利用的人，他原是一个小店主的儿子，后来子承父业，靠自学成才。他因这一部著作被选入当年成立的英国皇家学会，反映学术界对他这一著作的承认和重视。
- ◆ 这著作篇幅很小，主要内容为8个表，从现在的观点看只是一种例行的数据整理工作，但在当时则是有原创性的科研成果，其中所提出的一些概念在某种程度上可以说沿用至今，如**数据简约**（大量的、杂乱无章的数据，须注过整理、约化才能突出其中所包含的信息）、**频率稳定性**（一定的事件，如“生男”、“生女”，在较长时期中有一个基本稳定的比率，这是进行统计性推断的基础）、**数据纠错**、**生命表**（反映人群中寿命分布的情况，至今仍是保险与精算的基础概念）等。





# 统计学的起源

- ◆ 另一个重要源头来自天文和测地学中的误差分析问题。
  - ◆ 比利时天文学家兼统计学家凯特勒19世纪的工作，对促成现代数理统计学的诞生起了很大的作用。



# 统计学的发展

- ◆ 法国数学家兼天文家**勒让德**19世纪初（1805）在研究慧星轨道计算时发明的“**最小二乘法**”。
- ◆ 德国大学者**高斯**1809年在研究行星绕日运动时提出用**正态分布（也称高斯分布）**刻画测量误差的分布。
  - ◆ 现今仍在常用的许多统计方法，就是建立在“所研究的量具有或近似地具有正态分布”这个假定的基础上，而经验和理论（概率论中“中心极限定理”）都表明这个假定的现实性。





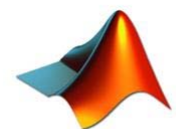
# 统计学的发展

- ◆ 19世纪后期由英国遗传学家兼统计学家**高尔顿**发起，并经现代统计学的奠基人之一费歇尔·K·皮尔逊和其他一些英国学者所发展的统计相关与回归理论。
  - ◆ 现实世界中的现象往往涉及众多变量，它们之间有错综复杂的关系，且许多属于非决定性质，相关回归理论的发明，提供了一种通过实际观察去对这种关系进行定量研究的工具，有着重大的认识和实用意义。



# 统计计算

- ◆ 计算机在20世纪后半叶的大量应用对统计科学产生了极大的影响。
  - ◆ 早期统计模型常常回避**非线性模型**，但强劲的计算机及其算法导致非线性模型（如神经网络）和新式算法（如广义线性模式、等级线性模型）的大量应用。
  - ◆ 计算机性能的增强使得需要大量计算的**再取样算法**成为时尚，如置换检验、自助法。**Gibbs取样法**也使得**贝叶斯模型**更加可行。
- ◆ 计算机革命使得统计在未来更加注重“**实验**”和“**经验**”
  - ◆ 大量普通或专业的统计软件如Excel、R、SAS、SPSS等。



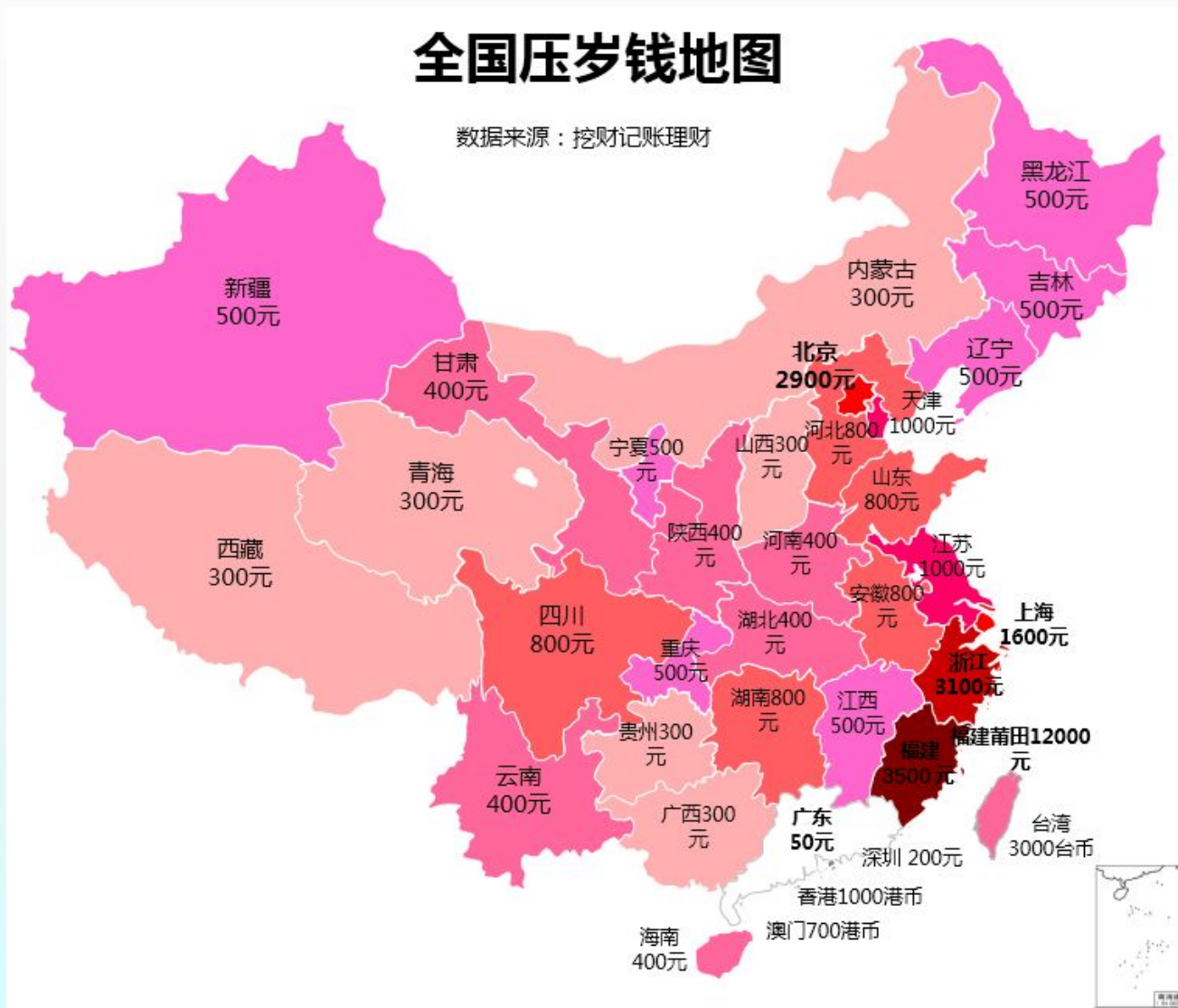
MATLAB



# 统计滥用

## 全国压岁钱地图

数据来源：挖财记账理财



# 统计滥用

- ◆ 谎言，该死的谎言，统计数字。

Lies, damned lies, and statistics.

——西方谚语（马克·吐温）

- ◆ 统计数字是精确说谎；
- ◆ 模糊说谎是“借用众人”的谬误逻辑，即造成“如果许多人都这么认为，那一定是对的”的假象。
- ◆ 对政府提供的统计数据可靠性的质疑，已经成为一个全球性的经常性话题。
  - ◆ 当前中国统计体系的问题主要表现在关键数据不准确、重要数据不统计和报告数据不重要。

——黄益平