

课程详述

COURSE SPECIFICATION

以下课程信息可能根据实际授课需要或在课程检讨之后产生变动。如对课程有任何疑问，请联系授课教师。

The course information as follows may be subject to change, either during the session because of unforeseen circumstances, or following review of the course at the end of the session. Queries about the course should be directed to the course instructor.

1.	课程名称 Course Title	魅力统计 Fascinating Statistics
2.	授课院系 Originating Department	统计与数据科学系 Department of Statistics and Data Science
3.	课程编号 Course Code	STA101
4.	课程学分 Credit Value	3
5.	课程类别 Course Type	通识选修课 General Education (GE) Elective Courses – 专业导论类
6.	授课学期 Semester	春季 Spring
7.	授课语言 Teaching Language	中英双语 English & Chinese
8.	授课教师、所属学系、联系方式（如属团队授课，请列明其他授课教师） Instructor(s), Affiliation & Contact (For team teaching, please list all instructors)	徐匆 统计与数据科学系 xuc6@sustech.edu.cn
9.	实验员/助教、所属学系、联系方式 Tutor/TA(s), Contact	待公布 To be announced
10.	选课人数限额(可不填) Maximum Enrolment (Optional)	100

11. 授课方式 Delivery Method	讲授 Lectures	习题/辅导/讨论 Tutorials	实验/实习 Lab/Practical	其它(请具体注明) Other (Please specify)	总学时 Total
学时数 Credit Hours	48				48
12. 先修课程、其它学习要求 Pre-requisites or Other Academic Requirements	无				
13. 后续课程、其它学习规划 Courses for which this course is a pre-requisite	统计学、数据科学与大数据技术其他课程				
14. 其它要求修读本课程的学系 Cross-listing Dept.	无				

教学大纲及教学日历 SYLLABUS

15. 教学目标 Course Objectives

本课程针对全校一年级本科生开设，教学目标包括：

- 了解统计学与数据科学中的一些基本概念、知识、应用场景
- 培养统计学与数据科学的思维方式和基本技能。

16. 预达学习成果 Learning Outcomes

本课程每一个章节都通过某个历史经典故事、网络热点事件、行业真实案例等引出对统计学与数据科学知识点的介绍，通过本课程学习，学生预期能够对统计学与数据科学中的基础知识有初步的了解，如数据类型、随机变量、概率分布、参数估计、假设检验等，并对一些统计方法产生的背景和历史有一定认知，并由此知道统计学与数据科学是问题驱动的学科。

17. 课程内容及教学日历（如授课语言以英文为主，则课程内容介绍可以用英文；如团队教学或模块教学，教学日历须注明主讲人）

Course Contents (in Parts/Chapters/Sections/Weeks. Please notify name of instructor for course section(s), if this is a team teaching or module course.)

第一章：引言 (2 学时)

1.1 技术的变迁

1.1.1 统计学的起源，统计学简介

1.1.2 大数据时代新纪元，大数据简介

1.2 分析工具

1.2.1 统计学者的最爱，R 语言简介

1.2.2 全栈工程师的最爱，Python 简介

第二章：统计与科学 (10 学时)

2.1 随机的世界

2.1.1 打台球的物理学家，确定、不确定与随机

2.1.2 上帝掷骰子吗，决定论与随机性

2.1.3 连环杀手的归案，随机与均匀

2.1.4 扔硬币的数学家们，大数定律

2.2 认识概率

2.2.1 游戏奖金的分配，概率论的发展

2.2.2 六连号和十四连号，概率的计算

2.2.3 主持人背后的山羊，条件概率

2.2.4 寻找失踪的核潜艇，贝叶斯概率

2.3 统计思想与模型

2.3.1 女士品茶，假设检验

2.3.2 雷达探测，最大似然

2.3.3 六西格玛的奇迹，小概率

2.3.4 牛顿的苹果，模型拟合

2.4 统计与科学

2.4.1 智多星与神机军师，统计预测

2.4.2 深蓝与阿尔法狗，归纳和演绎

2.4.3 中药与西药，临床试验

2.4.4 所有模型都是错的，科学和否定

第三章：数据与变量 (10 学时)

3.1 数据与空间

- 3.1.1 多维世界的虫子，坐标和向量
- 3.1.2 黑客帝国和变形金刚，矩阵简介
- 3.1.3 九章算术与线性方程，线性方程组
- 3.1.4 二十八宿与黄道十二宫，线性变换

3.2 随机变量与分布

- 3.2.1 伯努利的硬币，随机分布
- 3.2.2 相亲与神奇的 37%，离散型分布
- 3.2.3 棣莫弗的正态，连续型分布
- 3.2.4 酒鬼的步伐，随机过程

3.3 认识数据

- 3.3.1 忒休斯之船，数据、测量与变量
- 3.3.2 从性别到体重，数据的尺度
- 3.3.3 周岁与虚岁，连续变量与离散变量
- 3.3.4 一份体检记录，数据分析的基本数据结构

3.4 数理统计基础

- 3.4.1 管中窥豹与一叶知秋，总体和样本
- 3.4.2 赌徒的诡计，数字特征
- 3.4.3 被平均的工资，统计量
- 3.4.4 小李飞刀与孔雀翎，参数估计

第四章：数据可视化 (6 学时)

4.1 历史上的统计图形

- 4.1.1 河图与洛书，可视化简介
- 4.1.2 伦敦霍乱的防治，空间可视化
- 4.1.3 南丁格尔的玫瑰，玫瑰花瓣图
- 4.1.4 拿破仑远征，自定义统计图形

4.2 数据与可视化

- 4.2.1 女王的裙子，数据可视化
- 4.2.2 画布与宣纸，图形设备
- 4.2.3 深水王子与针眼画师，绘图语言的变迁
- 4.2.4 “挑战者”号航天飞机，直观的可视化

4.3 基础统计图形

- 4.3.1 老忠实喷泉的秘密，分布图
- 4.3.2 统计图形的奠基人，条形图和饼图
- 4.3.3 古老国度的诗云，坐标变换

4.3.4 飞翔的动态气泡图，动态可视化

4.4 数据之间的关系

4.4.1 东上相的轨道，散点图

4.4.2 五十州的最高峰，箱线图

4.4.3 泰坦尼克号的幸存者，马赛克图

4.4.4 切尔诺夫的笑脸，样本关系的可视化

第五章：模型与方法 (10 学时)

5.1 常用统计模型

5.1.1 穿杨与射雕，回归模型

5.1.2 降维攻击，主成分分析

5.1.3 顾客就是上帝，路径模型

5.1.4 股票的走势，时间序列

5.2 机器学习

5.2.1 啤酒和尿布的传说，关联规则

5.2.2 寻找“白富美”，聚类分析

5.2.3 树木与森林，常用分类算法

5.3 人工智能

5.3.1 人工智能的三起两落，AI 发展史

5.3.2 深度学习的前世今生，深度学习简史

5.3.3 神秘的神经，神经网络简介

5.3.4 美丽的滤镜，卷积神经网络与深度学习

5.4 其他分析方法

5.4.1 茶、酒与可乐，随机试验方法

5.4.2 蒙特卡洛和原子弹，蒙特卡洛方法

5.4.3 医生的笔迹，文本分析

5.4.4 沙漠里的飞碟，最优化方法

第六章：大数据时代 (4 学时)

6.1 计算框架

6.1.1 冰箱里的大象，可扩容的数据分析

6.1.2 将兵与将将，并行计算

6.1.3 电老虎和电蚂蚁，大型机和云计算

6.1.4 摩尔定律的未来，GPU 计算框架

6.2 大数据行业应用

- 6.2.1 互联网的兴起, 互联网概览
- 6.2.2 流量的起点, 搜索引擎
- 6.2.3 收入的来源, 精准广告
- 6.2.4 猜你喜欢和投其所好, 推荐系统

第七章：数据的陷阱 (4 学时)

7.1 一叶障目

- 7.1.1 神奇的天蝎座, 规律的背后
- 7.1.2 赢家的诅咒, 悖论与分布
- 7.1.3 飞机的油价, 选择性关注
- 7.1.4 缘分, 频率与巧合

7.2 相关与因果

- 7.2.1 芳华与热饮, 遗漏的关键变量
- 7.2.2 热帖的秘密, 不存在的相关
- 7.2.3 雪与火的城市, 地理决定的因果
- 7.2.4 名字很重要吗, 背后的关键因素

7.3 样本和调查

- 7.3.1 测不准的美国大选, 选择性抽样
- 7.3.2 幸运儿的传奇, 幸存者偏差
- 7.3.3 哈佛校长的辞职, 样本方差的影响

18. 教材及其它参考资料 Textbook and Supplementary Readings

[1] 统计之美. 李舰, 海恩. 电子工业出版社, 2019

[2] 魅力统计. 王静龙、梁小芸. 中国统计出版社, 2012

[3] 数据科学导论 (第 2 版). 杨旭, 汤海京, 丁刚毅. 北京理工大学出版社, 2017

[4] 数据科学实战. 冯凌秉, 王群锋[译](Schutt R., O'Neil, C. 著). 人民邮电出版社, 2015

[5] Data Mining: Practical Machine Learning Tools and Techniques (Fourth Edition) By Ian H. Witten, Eibe Frank, Mark A. Hall, Christopher J. Pal, 2016

课程评估 ASSESSMENT

19. 评估形式 Type of Assessment	评估时间 Time	占考试总成绩百分比 % of final score	违纪处罚 Penalty	备注 Notes
出勤 Attendance		10		
课堂表现				

Class Performance				
小测验 Quiz				
课程项目 Projects				
平时作业 Assignments		40		
期中考试 Mid-Term Test				
期末考试 Final Exam		25		
期末报告 Final Presentation		25		
其它（可根据需要 改写以上评估方式） Others (The above may be modified as necessary)				

20. 记分方式 GRADING SYSTEM

A. 十三级等级制 Letter Grading

课程审批 REVIEW AND APPROVAL

21. 本课程设置已经过以下责任人/委员会审议通过

This Course has been approved by the following person or committee of authority

田国梁, 2023.10.10