

## 课程建设经验分享

# 数据分析与经济决策

连玉君 (中山大学)

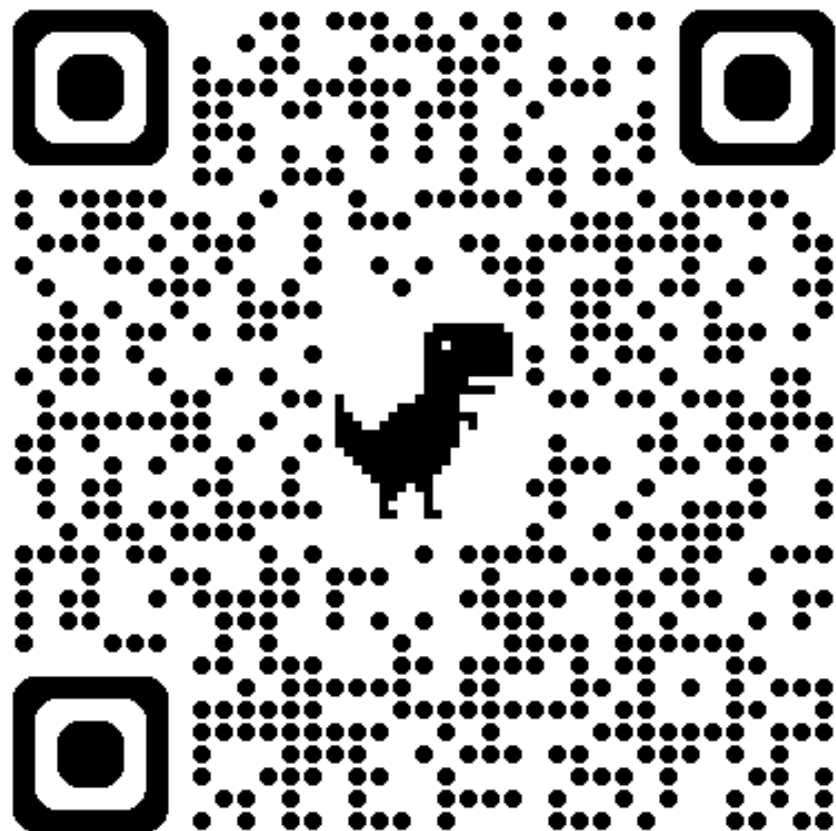
arlionn@163.com

课程主页:

<https://lianxhcn.github.io/ds>

Github 仓库:

<https://github.com/lianxhcn/ds>





- 整体感受
- 课程内容
- 教什么?
- 如何教?
- 讨论

# 整体感受

- 敬畏之心：和学生一起学习
  - Stata/R → Python; LLM; GenAI
  - 课前准备：确定主题、摸底、与相关课程老师沟通
- 让学生卷起来
  - 作业设计、课堂讨论、私下讨论 (线人)
- 思路转变
  - 钓鱼
  - 推理和拆解

名称	修改日期
王璐 24210166-V1-ex01_王璐.ipynb	2025/5/25 19:45
幸经凤-v2-ex2-个人作业.ipynb	2025/5/24 13:25
赵晓燕-v1-ex01-赵晓燕.jpynb.ipynb	2025/5/24 11:04
余正-V1-V1.ipynb	2025/5/24 10:10
李锦婷-v2-李锦婷_v2.ipynb	2025/5/24 10:06
薛艳婷-v1-ex01薛艳婷.ipynb.ipynb	2025/5/24 2:35
王泓北-V1-个人作业-王泓北.ipynb	2025/5/24 1:18
黄金石-v1-黄金石24210128.ipynb	2025/5/24 1:06
曾媚-v1-ex01_曾媚.ipynb.ipynb	2025/5/24 0:05
黄欣24210130-v1-黄欣个人作业.ipynb	2025/5/24 0:04
张昀-v1-ex01_张昀.ipynb	2025/5/23 23:40
吴薇-V2-吴薇_V2_第一次数据个人作业.i...	2025/5/23 23:39
邱钦壮24210157-V1--个人作业.ipynb	2025/5/23 23:35
刘亦宋-v1-ex01_刘亦宋.ipynb.ipynb	2025/5/23 23:19
梁钟方-v1-ex01_梁钟方.ipynb	2025/5/23 22:57

万诗晴的-v2-ex01_万诗晴.ipynb	2025/5/23 9:38
张翼-V1-V1-ex01_张翼.ipynb	2025/5/23 9:15
蒋俊豪-v1-ex01_蒋俊豪.ipynb.ipynb	2025/5/23 2:48
莫贤彬-v1-莫贤彬-24210155.ipynb	2025/5/23 2:34
黄伊姿-v4-第一次个人作业-黄伊姿.ipynb	2025/5/23 1:31
刘英健-v1-刘英健-个人作业v1.ipynb	2025/5/23 1:31
文佳琪-v1-个人作业-第一次.ipynb	2025/5/23 1:24
00_HW01_典型问题.ipynb	2025/5/23 0:36
王一明-v1-ex01_王一明-个人作业.ipynb	2025/5/22 23:51
江兆鹏-v1-江兆鹏-v1-24210134-第一...	2025/5/22 23:51
黎婉珊-V2-ex01_黎婉珊V2.ipynb	2025/5/22 23:44
傅烨豪-v1-ex01-傅烨豪.ipynb	2025/5/22 23:13
肖伟慧-v1-ex01_肖伟慧.ipynb	2025/5/22 23:10
陈贵斌-v1-ex01_陈贵斌.ipynb	2025/5/22 23:05
柯玉蕴-V1-ex01_柯玉蕴.ipynb.ipynb	2025/5/22 23:02
陈雪华-V5-ex01_陈雪华.ipynb	2025/5/22 22:22

# 教指委的课程指南



## 一、课程概述

- 《数据分析与经济决策》是数字经济专业硕士研究生的核心课程之一。
- 课程目标：为学生提供全面的数据分析技能，支持数字经济领域的决策与实践。
- 学习内容涵盖：
  - 数据分析在经济决策中的作用与意义
  - 数据预处理、统计推断与因果推断
  - 机器学习与深度学习 \*
  - 贝叶斯方法 \*
  - 网络分析与社交媒体数据分析 \*
  - 大语言模型及自然语言处理 \*
- 强调数据分析方法在辅助经济决策中的应用。
- 学生将具备扎实的数据处理、分析、解释及决策能力。
- 为未来从事科研、政策制定或企业决策打下坚实基础。

### 三、课程目标

1. 熟练掌握数据分析方法；
2. 运用机器学习和深度学习算法；
3. 掌握网络和社交媒体数据分析能力；
4. 了解经济决策的基本原理和方法；
5. 培养数据驱动决策的能力。



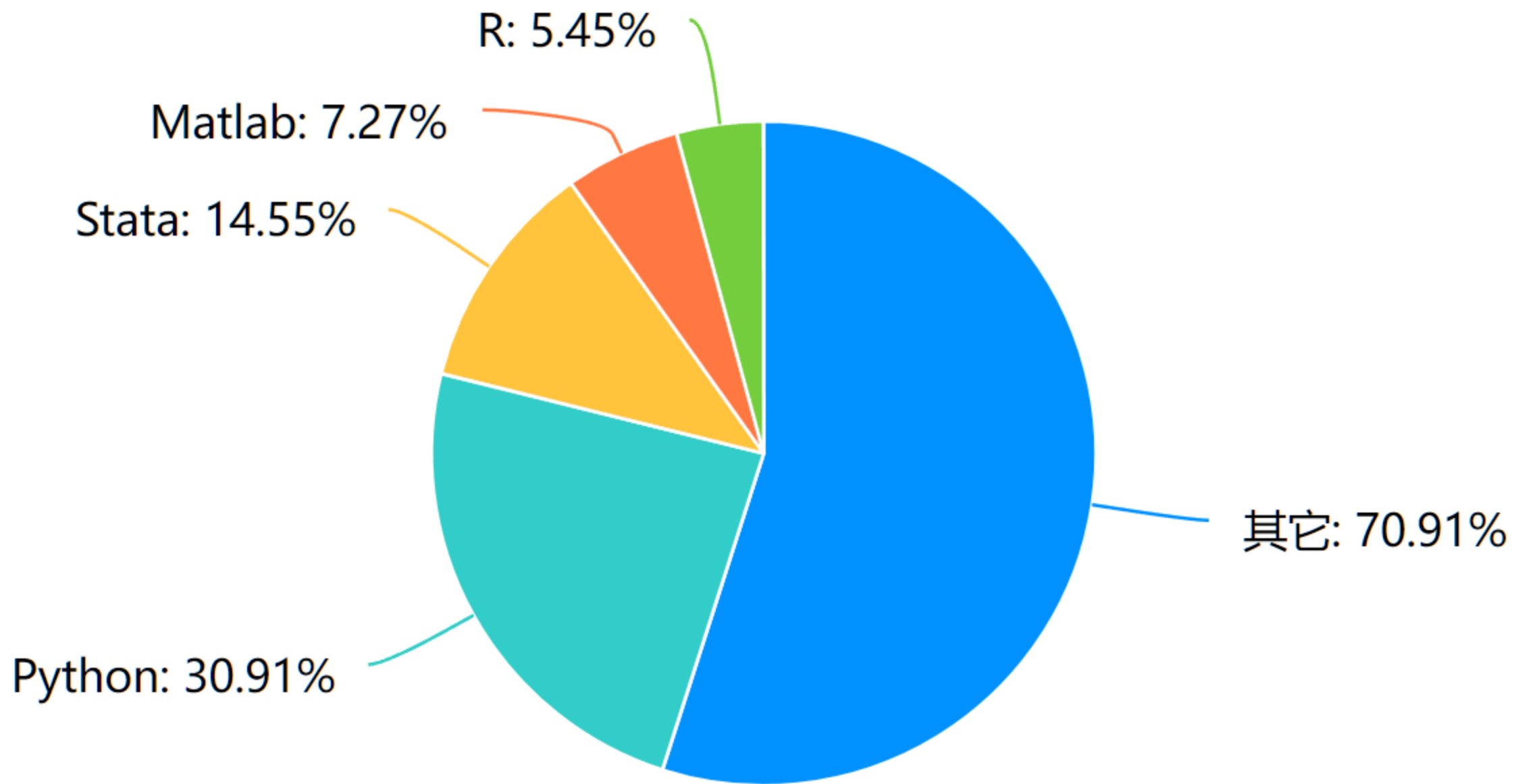
## 五、授课方式

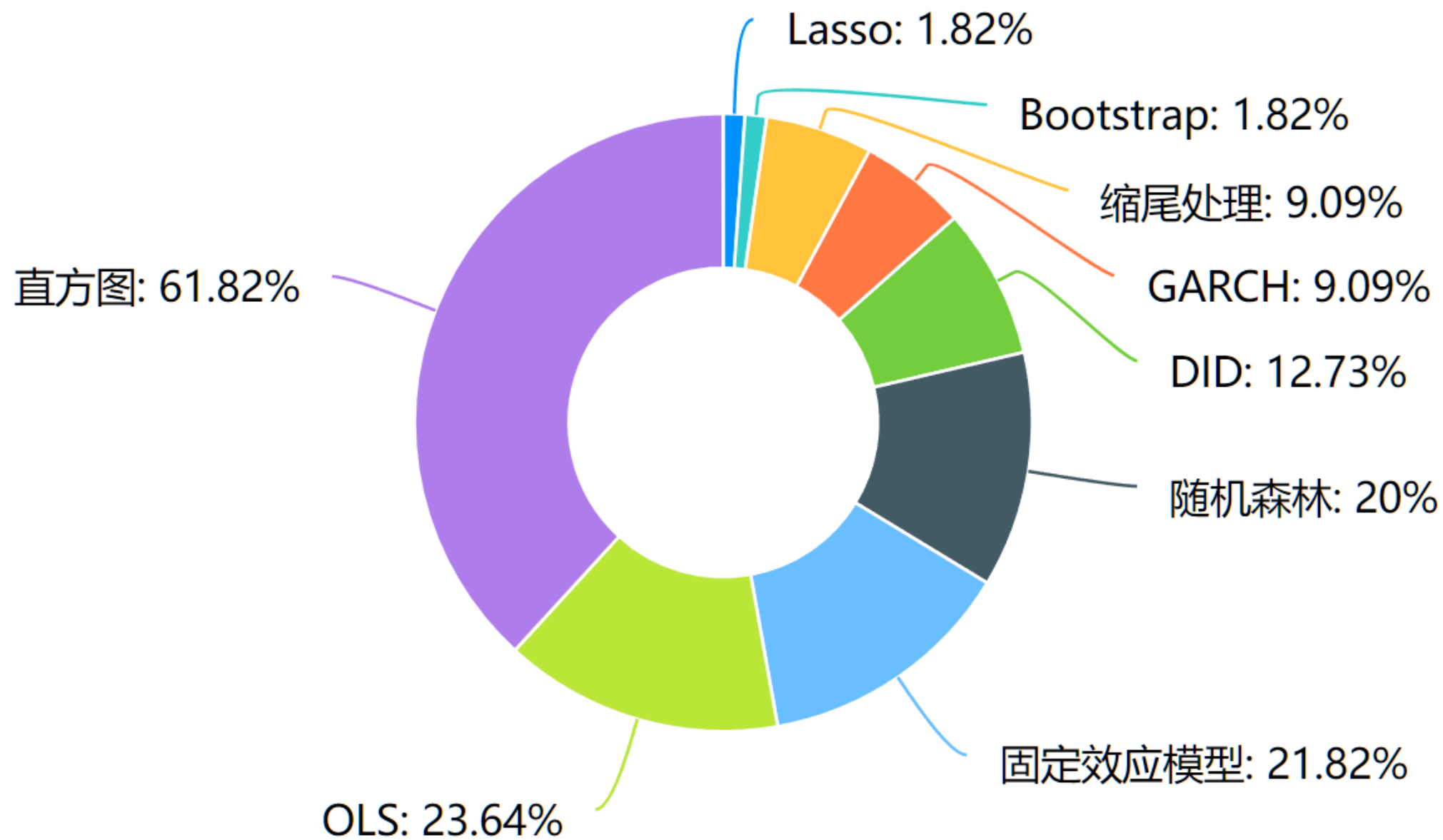
- 本课程采用多种教学方法，包括课堂教师讲授和课堂互动讨论相结合。
- 基础理论知识由教师进行课堂讲授，而应用和数据分析则通过课堂讨论互动的方式实现。
- 另外，积极引入在线学习平台和资源，为学生提供更丰富的学习资源和学习方式。
- 利用智能辅助教学工具，以及数据可视化工具和软件，来提升学习效果。

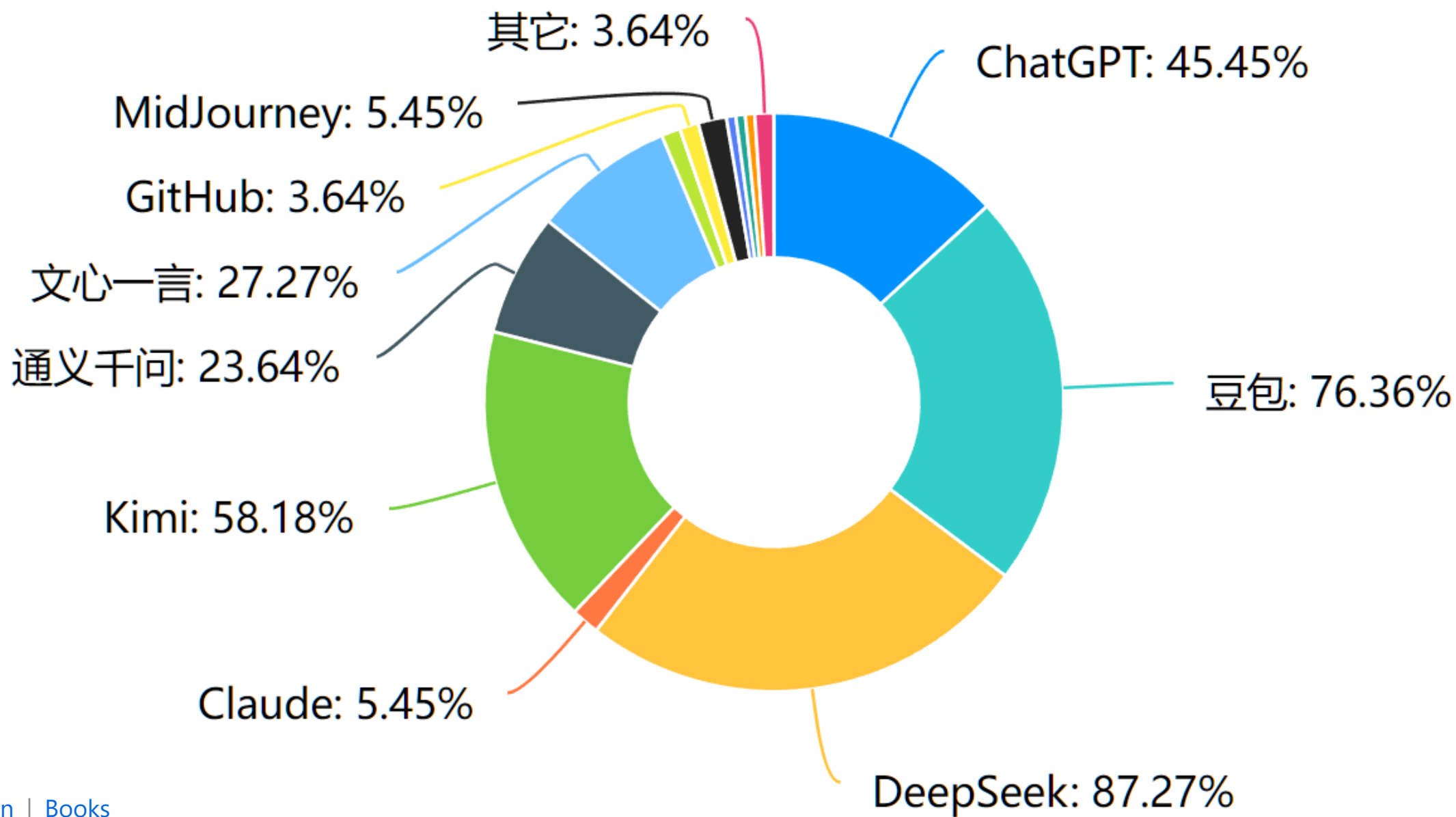
# 课前摸底

学生人数：55 人；age：24-49 岁

1. 你会用哪些软件？（多选题）
2. 你了解以下哪些方法？（多选题）
3. 你期望学习哪方面的内容？（多选）
4. 你经常使用哪些 AI 工具？
5. 你更喜欢哪种授课方式？







选项	小计	比例
上市公司财务数据分析（合并、清洗、基本统计和回归分析）	40	<div><div></div></div> 72.73%
股票和投资组合分析（投资组合收益-风险分析）	35	<div><div></div></div> 63.64%
全球宏观数据（各国GDP，M2，通胀率，失业率，贸易等）	32	<div><div></div></div> 58.18%
文本分析（政府公告、企业公告的爬取、主题分析、情感分析）	32	<div><div></div></div> 58.18%
因果推断和政策评价方法（某项政策的短期和长期影响，事件研究法）	30	<div><div></div></div> 54.55%
单个企业的财务分析（指标分析、背景分析等）	28	<div><div></div></div> 50.91%
可视化分析（常用二维图，三维图和动图的绘制）	27	<div><div></div></div> 49.09%
其它	4	<div><div></div></div> 7.27%
本题有效填写人次	55	

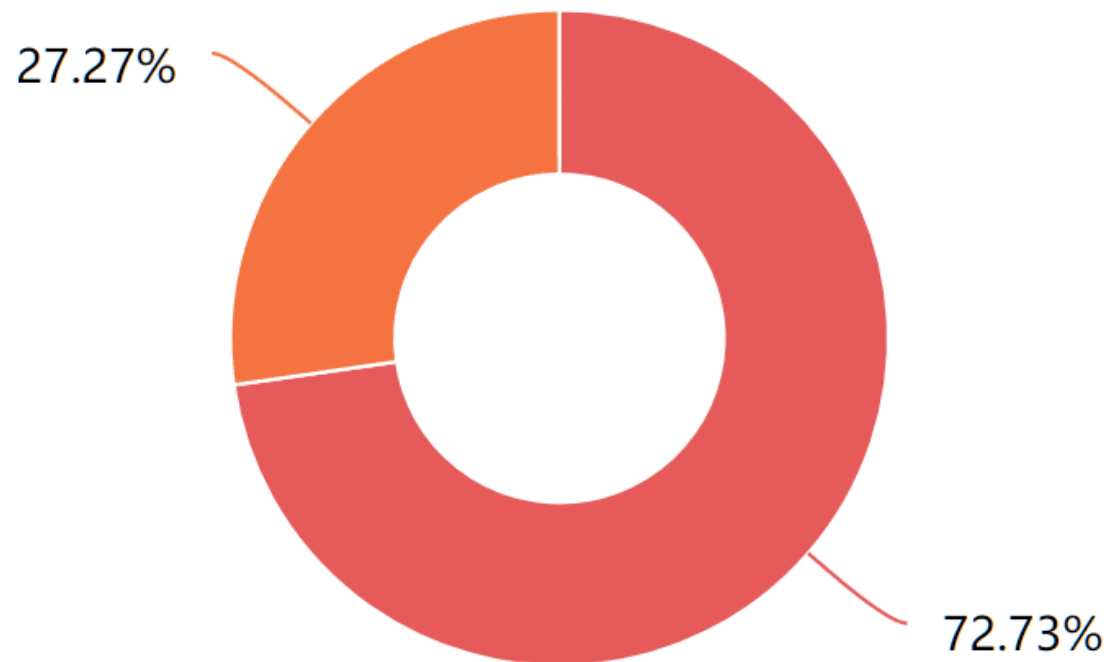
## 教学模式选择

### 9-1 模式

- 90% 以上由老师讲授，学生课后完成作业

### 6-4 模式

1. 老师讲授核心概念和原理、数据分析流程等
2. **小组作业**：一个小型的数据处理和分析项目
3. 课堂上留出大概 2/5 的时间，由学生报告，并与同学和老师做详细的讨论。





## 学生背景 (1)

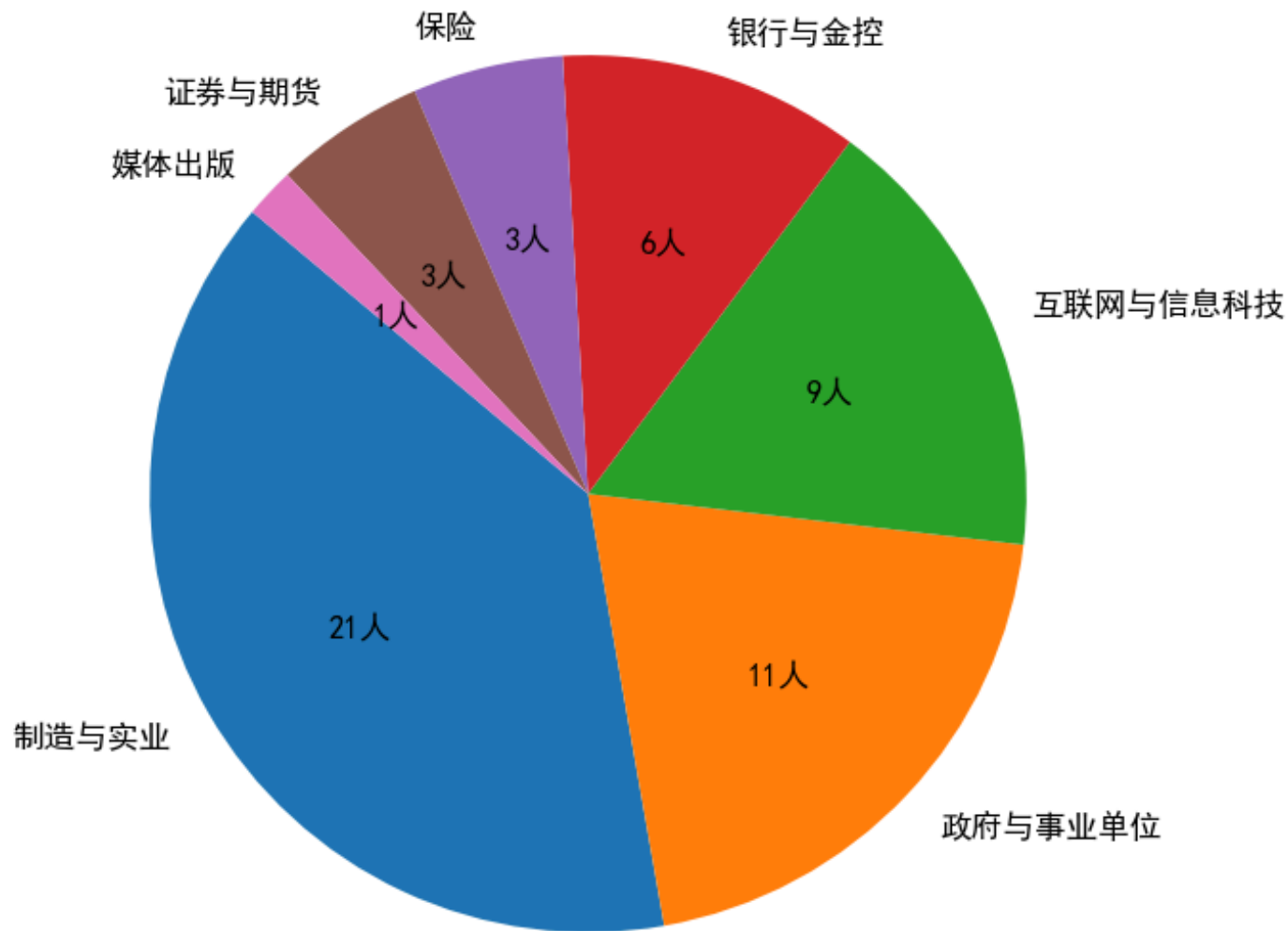
### 制造与实业: 21 人

- 中国联合网络通信集团
- 广东裁成律师事务所
- 广州安迅经济发展有限公司
- 广东天禾农资股份有限公司

### 政府与事业单位: 11 人

- 广州市烟草专卖局
- 佛山市商务局
- 萍乡市发展和改革委员会
- 广州市荔湾区财政局

学员单位分类分布 (人数)



## 学生背景 (2)

### 互联网与信息技术

- 数字广东网络建设有限公司
- 荣耀终端有限公司
- 深圳市积加跨境网络科技有限公司
- 阿里巴巴集团

### 银行与金控

- 交通银行股份有限公司广州花都支行
- 中国工商银行股份有限公司广东省分行
- 广发银行股份有限公司信用卡中心

### 保险

- 富德财产保险股份有限公司
- 大家人寿保险股份有限公司广东分公司
- 中国人寿保险股份有限公司广东省分公司

### 证券与期货

- 招商证券股份有限公司
- 广发期货
- 中信证券

### 媒体出版

- 南方财经全媒体集团

# 教什么？

- 教指委的课程指南
- 我们实际教授的内容

# 教指委 · 课程指南 (1)

## 1. 基础

- 数据获取与清洗
- 描述性统计与可视化
- 统计推断：估计、检验、方差分析

## 2. 因果与机器学习

- 因果推断：实验与政策评估
- 机器学习：监督 / 非监督
- 分类与聚类：树、KNN、K-means

## 教指委·课程指南（2）

### 3. 高级方法

- 深度学习与神经网络
- 贝叶斯方法与风险管理
- 网络分析与社交媒体数据
- 大语言模型（LLM）与文本分析

# 我们教的主要内容

- 数据分析的 **目的**
  - **搞清楚事实**：数据清理和可视化
  - **搞清楚关系**：回归分析、因果推断
  - **预测**：时间序列分析、机器学习
  - **辅助决策**：优化、模拟
- 数据分析的 **流程**
  - 目标 → 数据 → 方法 → 结果 → 决策
- 数据分析的 **工具**
  - AI 工具 + Python + Jupyter Notebook
  - Github + Github Desktop
  - Markdown + Marp

## 主要模块 (1)

- 数据的获取
  - 常用数据库：CSMAR、Wind、CEIC
  - API：yfinance、tushare、akshare .....
  - 爬虫
- 数据清洗和格式化
  - 项目文档结构
  - 格式化数据与非格式化数据
  - 数据变换、二次编码、离群值
- 数据可视化
  - 常用图表 (密度函数图、箱线图、分仓散点图)
  - 可视化的一些基本原则



## 主要模块 (2)

- 探索性数据分析 (EDA)
  - 了解数据的基本特征、变量之间的关系
  - 初步建模
- 常用统计和计量模型
  - 入手：数据类型和分布特征
  - 假设检验和统计推断：传统 → Bootstrap + 交叉验证
  - 线性模型 → GLM → 离散选择和受限因变量
  - 机器学习方法
    - 分类、回归、聚类、降维
    - 树模型、随机森林、XGBoost

## 主要模块 (3)

- 数据分析案例
  - 上市公司财务分析：负债率
  - 宏观经济分析：GDP、失业率、通胀率
  - 金融市场分析：股票收益率、波动率、风险溢价
  - 消费行为分析：客户细分、购买预测
  - 文本数据分析：情感分析、主题建模
  - 网络数据分析：社交网络、影响力传播

如何教？

## 如何教？整体思路

🍏 搭好戏台 → 缺啥补啥 → 先让代码跑起来

- AI 辅助教学
- 原理 + 流程 + 规范
- 统计软件和工具的选择
  - Python + Jupyter Notebook
  - Github + GitHub Copilot + Github Desktop
  - 好处：提高编程效率，便于协作和版本控制
- Markdown
  - 一定要多写：想不清楚的东西一定写不清楚
  - 使用 Marp 制作幻灯片

## 如何教？作业

- 个人作业：每周一次 (2-3 小时)
  - [ex\\_P01.md](#) | [ex\\_P02.md](#)
- 小组作业：2-3 个 (每组 4-5 人)
  - 根据兴趣选择案例 (最好能提供 [备选主题](#))
  - 展示和讨论 (2-3 个小组做同一个案例)
- 老师：设计作业 v.s. 布置作业

## 如何教？收作业

- 目前：坚果云 [交作业](#)
- 以后：github [助教工作指南](#)

## 如何教？教材和讲义

- 教材：理论基础扎实、结构完整
  - [2 课程简介和资源](#)
- AI 辅助生成，参见 连玉君, 2025, [Empirical Research with AI](#)
  - 关键：学会写提示词 - 结构、逻辑
- Github 仓库
  - [2 课程简介和资源](#)
  - [6100+ 仓库](#)
- 在线讲义：Quarto + GitHub Pages
  - 连玉君，2025，[Quarto Book](#)
  - [用 Quarto book 写的书](#)



## Python

- [Problem Solving with Python](#) by Kazarinoff
- [Automate the Boring Stuff with Python](#) by Sweigart
- [Python Data Science Handbook](#) by VanderPlas
- [Think Python: How to Think Like a Computer Scientist](#) by Downey
- [A Bite of Python](#)
- [Dive Into Python 3](#) by Pilgrim

# 讨论

## 讨论 1: 彼此的优势

### 老师

- 理论基础
- 工具地图和分析流程
- 前沿工具

### 学生

- 实践经验 → 应用场景 (提问)
- 案例分析能力
- 内卷的潜力: 对新工具的适应能力

## 讨论 2：教学模式

- 环境配置：让代码跑起来 (信心)
- 学生没有能力或者需要花很多时间才能理解的内容
  - 线性回归 → 非参数估计 (KNN, 核密度函数图, 随机森林)
  - 条件期望 + 条件概率 → GLM (广义线性模型) → Logit/Duration
- 分析流程和规范 (经验)
  - EDA → 可视化 → 回归分析 → 机器学习
  - 离群值
  - 非结构化数据 → 结构化数据
- 阅读和检索能力 → 知道周围在发生什么 → 趋势敏感性

## 一些典型的大纲

- Harvard University, [Data Analysis Courses](#), 课程分类很细致，提供了大量免费在线课程
- Washington State UNIVERSITY, [Data Analytics Course Syllabi](#) (19 门课)
  - [Statistical Modeling for Data Analytics](#)

## 教学模式

- 思路一：教方法和模型 → 学生自行选择案例分析对象
- 思路二：案例导向 → 学生根据自己的需要来学习

## 讨论 3：作业

### 个人作业：

- 环境配置
- 基础知识和概念
- 🍎：要让学生「卷起来」
- 量大，有一定难度

### 小组作业：

- 根据兴趣选择案例
- 展示和讨论 (2-3 个小组做同一个案例)
- 作业库？

## 讨论 4：教材和讲义

- 联合编写教材和讲义
- Github + Quarto (协作)



## 讨论 5：案例库 🍏

- MBA 教学经验：中欧案例库
  - 岭院的师资培训：MIT Sloan 管理学院 (5 个月) + 中欧案例培训
  - MBA 教学经验：MBA-CF
- Kaggle 数据平台 (<https://www.kaggle.com/datasets>)
  - 深度不够、案例背景资料缺乏
- 学生的资源
  - 案例报告 / 小组作业
  - 毕业论文
  - 校企合作
- 年度案例大赛或案例征集

<https://lianxhcn.github.com/ds>