**上海财经大学2022 －2023 学年第一学期教学进度表**

 学院 专业 年级 合班 课程名称 数据世界探秘

总课时 32本学期安排课时数 32 （其中讲课 27 课时，习题讲解、讨论 5 课时，其他 课时）

讲课教师 吴纯杰 助教 陈紫薇

课程代码 103250 课程序号 0024 讲课教师 吴纯杰

| 周次 | 月/日 ～月/日 | 讲课 | 课堂练习、讨论及课外作业 |
| --- | --- | --- | --- |
| 时数 | 内容 | 时数 | 内容 |
| 1 | 2022年9月5日 | 2 | **第一章 欢迎来到数据科学时代**1. 引言
2. 数据科学的内涵和发展
3. 数据科学的学科地位
4. 数据科学的成熟度曲线
5. 数据科学的理论体系
 |  | 第一章结束后布置团队作业1 |
| 3-4 | 9月19日9月26日 | 4 | **第二章 数据及数据来源**1. 数据定义和类型
2. 传统的数据来源
3. 大数据数据来源
 | 1 | **第二章结束后团队作业1课堂汇报**第二章结束后布置团队作业2 |
| 6 | 10月10日 | 2 | **第三章 数据分析的道与术**第一节 数据分析定义与价值第二节 统计建模工具2.1 SPSS2.2 SAS* 1. R语言
	2. Python
 |  |  |
| 7-9 | 10月17日10月24日10月31日 | 6 | **第三章 数据分析的道与术**第三节 探索性数据分析框架3.1 数值描述分析3.2 数据可视化3.3 数据清洗和转换3.4 实例分析 | 1 | **第三章第三节结束后团队作业2课堂汇报**第三章第三节结束后布置团队作业3 |
| 10-11 | 11月7日11月14日 | 6 | **第三章 数据分析的道与术**第四节 统计建模技术4.1 抽样分布4.2 假设检验4.3 预测分析4.4 聚类分析 | 1 | 第三章第四节开始前布置课程论文**第三章结束后团队作业3课堂汇报** |
| 12-13 | 11月21日11月28日 | 4 | **第四章 数据陷阱**1. 有偏的样本
2. 生存者偏差
3. 精心挑选的平均数
4. 相关关系的误解
5. 辛普森悖论
6. 统计显著与经济显著
 | 1 | 第四章开始前布置团队作业4**第四章结束后团队作业4课堂汇报** |
| 14-16 | 12月5日12月12日12月19日 | 8 | **第五章 数据世界专题课** 第一节 经济管理大数据案例及其应用 第二节 以国家战略为主导向，探统计分析之新思维 第三节 工欲善其事，必先利其器--解读大创项目和建模比赛 第四节 社会生活中的统计思维 |  | **第五章结束后提交结课论文。** |
| 指定教材和教学参考书 | 指定教材：王汉生，数据思维：从数据分析到商业价值，中国人民大学出版社，2018年参考书目：（1）格鲁斯，《数据科学入门》，人民邮电出版社，2016年。（2）朝乐门，《数据科学理论与实践》，清华大学出版社，2017年。（3）杰弗里˙班尼特，威廉˙L.˙布里格斯，马里奥˙F.˙崔奥拉，《妙趣横生的统计学》，人民邮电出版社，2016年。（4）吴喜之，《统计学：从概念到数据分析（第二版）》，高等教育出版社，2016年。（5）Robert I. Kabacoff，《R语言实战（第2版）》，人民邮电出版社，2016年。（6）Norman Matloff，《R语言编程艺术》，机械工业出版社，2013年。（7）吴喜之，《复杂数据统计方法——基于R的应用（第三版）》，中国人民大学出版社，2017年。（8）Megan Squire，《干净的数据：数据清洗入门与实践》，人民邮电出版社，2016年。（9）Winston Chang，《R数据可视化手册》，人民邮电出版社，2014年。（10）Nathan Yau，《鲜活的数据：数据可视化指南》，人民邮电出版社，2012年。（11）戴维·萨尔斯伯格，《女士品茶》，江西人民出版社，2016年。（12）周英、卓金武、卞月青，《大数据挖掘：系统方法与实例分析》，机械工业出版社，2016年。（13）刘小平，《统计学——理论、案例、实训》，电子工业出版社，2017年。（14）城市数据团，《数据不说谎：大数据之下的世界》，清华大学出版社，2017年。（15）达莱尔·哈夫，《统计数字会撒谎》，中国城市出版社，2009年。（16）戴维·穆尔、威廉·诺，《统计学的世界》，中信出版社，2017年 |

注：本表一式四份，一份送教研室，一份送讲课班级，一份送教务处，一份讲课教师自留。 教研室主任