

天数	课程名称	主要内容
第一天上午	利用可视化技术初步探索并展示数据	➤ 数据绘图中的一些注意事项 ➤ 如何绘制一个最能表达你的意图的图形 ➤ 几种常见图形的绘制（直方图、饼图、柱状图、散点图、箱式图、等高线图等） ➤ 利用 JMP 软件绘制金字塔图 ➤ 利用 JMP 软件轻松绘制地图，显示各地区的差异变化 ➤ 如何绘制风玫瑰图 ➤ 利用 JMP 软件绘制气泡图 ➤ 利用 JMP 软件绘制动态的时空变化图 ➤ 如何绘制热图 ➤ 如何绘制马赛克图 ➤ 利用 JMP 软件生成词云图，进行文本挖掘 ➤ 利用 JMP 软件快速聚类图，对数据进行初步分类 ➤ 如何绘制相关系数色图 ➤ 绘制 ROC 曲线图 ➤ 利用 JMP 软件一键绘制森林图 ➤ 绘制维恩（Venn）图 ➤ 利用 JMP 软件快速制表

第一天下午	基本统计分析方法串讲	➤ t 检验和方差分析有什么关系？ ➤ 方差分析与线性回归有什么关系？ ➤ 文章中常见的一般线性模型是什么意思？ ➤ 一般线性模型与广义线性模型的区别 ➤ 从更高的层面上理解基础统计方法，串起常见的基础统计方法 ➤ 如何从“关系”的角度理解常见的基础统计方法 ➤ 方差分析与秩和检验的关系 ➤ 卡方检验与秩和检验的关系
第二天上午	数据分析建模的“天龙八部”（上）	➤ （1）数据清洗 ➤ 快速发现异常值，对数据进行清洗 ➤ 如何处理异常值 ➤ 快速发现缺失值 ➤ 一键实现对缺失数据的填补 ➤ （2）变量变换 ➤ 什么是虚拟变量/哑变量 ➤ 如何识别应该进行哪种变换？ ➤ 快速发现变量之间的恰当关系 ➤ 连续变量何时应变为分类资料？如何划分？ ➤ （3）变量降维

		➤ 变量太多了应如何处理 ➤ 什么情况下可以直接删除变量？ ➤ 利用主成分分析进行变量的综合 ➤ 主成分回归 ➤ 利用偏最小二乘回归进行变量降维 ➤ （4）初步探索变量与结局之间的关系 ➤ 如何快速发现上百个变量之间的关系，从中初步寻找有意义的变量？ ➤ 如何结合探索结果发现新的研究思路 ➤ 发现所有变量间都没有关联怎么办？——从“垃圾”中寻宝
第二天下午	数据分析建模的“天龙八部”（下）	➤ （5）初步建立影响因素模型/风险预测模型 ➤ 常用的变量筛选技术 ➤ 逐步回归筛选出的危险因素是否可靠？ ➤ 当单因素和多因素分析结果不一致时，如何处理？ ➤ 如果一个因素在单因素分析中有意义，而多因素分析无意义，如何考虑这一因素？ ➤ （6）模型评价与改进 ➤ 常用的模型评价指标有哪些？ ➤ 模型的 R 方太小，是否意味着模型不能用？ ➤ 如何比较不同模型的好坏

		➤ （7）模型诊断 ➤ 如何发现模型中可能存在的问题 ➤ 数据分析建模中常见的几种问题及解决方案 ➤ Logistic 回归中 OR 大于 999，这是正常现象吗？ ➤ （8）建立最终模型 ➤ 什么样的模型才算好的模型？ ➤ 如何利用模型进行预测
--	--	---