

# 解决行业问题的统计思维

## 课程纲要

《解决行业问题的统计思维》是免费的在线课程，适用于所有希望掌握实用技能，使用数据来更好解决问题的人。本课程由七个模块组成，共计约30小时，课程进度可自行掌控。每个模块都包括教学短视频、JMP演示、问题和练习。要了解详细信息和报名，请访问 [jmp.com/statisticalthinking](http://jmp.com/statisticalthinking)。

每个模块所涵盖的主题如下所述。



### 模块 1：统计思维与问题解决

学习如何绘制过程图、定义项目及界定其范围，以及确定解决问题所需的数据。

#### 统计思维

- 什么是统计思维

#### 问题解决

- 问题解决的概览
- 解决问题的统计方法
- 问题类型

#### 定义问题

- 定义问题
- 目标与关键绩效指标
- 白色聚合物案例研究

#### 定义过程

- 什么是过程？
- 制定 SIPOC 图
- 制定输入/输出过程图
- 自上而下流程图和部署流程图

#### 识别潜在的根本原因

- 识别潜在原因的工具
- 头脑风暴
- 多轮投票
- 使用亲和图
- 因果图
- 五个为什么
- 因果矩阵

#### 编制和收集数据

- 用于解决问题的数据收集
- 数据类型
- 可操作的定义
- 数据收集的策略
- 导入数据进行分析



### 模块 2：探索性数据分析

学习如何利用图形来描述数据，利用交互式可视化来查找和沟通数据中的内容。

#### 描述数据

- 描述性统计量简介
- 数据类型
- 直方图
- 集中趋势和位置的测量
- 离差的测量 — 极差和四分位数间距
- 离差的测量 — 方差和标准差
- 可视化连续型数据
- 描述类别型数据

#### 概率概念

- 概率概念简介
- 样本和总体

- 了解正态分布
- 正态性检验
- 中心极限定理

#### 用于解决问题的探索性数据分析

- 探索性数据分析简介
- 探索连续型数据：增强工具
- Pareto 图
- 堆叠条形图和数据过滤
- 树状图和马赛克图
- 使用格状图和叠加变量
- 气泡图和热图
- 探索性数据分析工具的汇总

## 模块 2：探索性数据分析（续）

### 与数据沟通

- 与数据沟通简介
- 创建有效可视化
- 评估可视化的有效性
- 设计有效可视化
- 用动画进行可视化沟通
- 针对受众进行设计
- 了解目标受众
- 设计用于沟通的可视化
- 设计可视化：注意事项

### 保存和分享结果

- 保存和分享结果简介
- JMP 内部的保存和分享结果
- JMP 外部的保存和分享结果
- 决定使用何种格式

### 准备数据进行分析

- 数据表的要点
- 常见的数据质量问题
- 识别数据表中的问题
- 识别问题（单次单变量）
- 重构数据进行分析
- 合并数据
- 衍生新变量
- 处理日期



## 模块 3：质量方法

学习用于量化、控制和减少产品、服务或过程中变异的工具。

### 统计过程控制

- 控制图简介
- 单值和移动极差控制图
- 常见原因变异与特殊原因变异
- 特殊原因的检验
- 均值-极差控制图与均值-标准差控制图
- 合理子组化
- 三因子控制图
- 分阶段的控制图

### 过程能力

- 客户的声音
- 过程能力指标
- 能力的短期和长期估计
- 了解过程改进的能力

- 估计过程能力：示例
- 计算非正态数据的能力
- 估计多个变量的过程能力
- 识别性能差的过程
- 行业视角

### 测量系统研究

- 什么是测量系统分析 (MSA)?
- 语言和术语
- 设计测量系统研究
- 设计和执行 MSA
- 分析 MSA
- 研究测量系统准确度
- 改进测量过程



## 模块 4：利用数据做出决策

学习从数据中得出推论，构建统计区间，执行假设检验，并理解样本大小与功效之间的关系。

### 估计

- 统计推断简介
- 什么是置信区间?
- 估计平均值
- 可视化抽样变异
- 构建置信区间
- 了解置信水平和 Alpha 风险
- 预测区间
- 容差区间
- 比较区间估计

### 统计检验的基础

- 统计检验简介
- 统计决策
- 了解原假设和备择假设
- 原假设下的抽样分布
- p 值和统计显著性

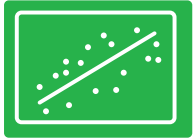
## 模块 4：利用数据做出决策（续）

### 连续型数据的假设检验

- 执行单样本 t 检验
- 了解 p 值和 t 比值
- 等价性检验
- 比较两个平均值
- 不等方差检验
- 配对的观测
- 单因子 ANOVA（方差分析）
- 多重比较
- 统计显著性与实际显著性

### 样本大小与功效

- 样本大小与功效简介
- 平均值置信区间的样本大小
- 统计检验的结果
- 统计功效
- 探索样本大小与功效
- 计算单样本 t 检验的样本大小
- 计算双样本 t 检验和方差分析的样本大小



## 模块 5：相关性与回归

学习如何研究成对变量之间的线性关联，如何拟合和解释线性和 Logistic 回归模型。

### 相关性

- 什么是相关性？
- 解释相关性

### 简单线性回归

- 回归分析简介
- 简单线性回归模型
- 最小二乘法
- 可视化最小二乘法
- 回归模型的假定
- 解释回归结果
- 使用曲率拟合模型

### 多元线性回归

- 什么是多元线性回归？
- 拟合多元线性回归模型

- 在“解释建模”中解释结果
- 残差分析和离群值
- 带类别型预测变量的多元线性回归
- 带交互项的多元线性回归
- 变量选择
- 多重共线性

### Logistic 回归的介绍

- 什么是 Logistic 回归？
- 简单 Logistic 模型
- 简单 Logistic 回归示例
- 解释 Logistic 回归结果
- 多元 Logistic 回归
- 带交互项的 Logistic 回归
- 常见问题



## 模块 6：实验设计

学习实验设计 (DOE) 的语言，以及了解如何在 JMP 中设计、开展和分析实验。

### 实验设计的介绍

- 什么是实验设计？
- 开展特设实验和单次单因子 (OFAT) 实验
- 为什么使用实验设计？
- 实验设计的术语
- 实验设计的类型

### 析因实验

- 设计析因实验
- 分析重复的完全析因
- 分析未重复的完全析因

### 筛选实验

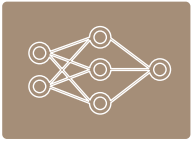
- 筛选重要效应
- 了解部分析因设计
- 定制筛选设计

### 响应曲面实验

- 响应曲面设计简介
- 分析响应曲面实验
- 创建定制响应曲面设计
- 序贯实验

### 实验设计指南

- 实验设计指南简介
- 定义问题和目标
- 识别响应
- 识别因子和因子水平
- 识别限制和约束
- 准备开展实验
- 案例研究



## 模块 7：预测性建模和文本挖掘

学习如何识别可能的关系，构建预测性模型以及从自由格式的文本中导出数值。

### 预测性建模的要点

- 预测性建模简介
- 过拟合与模型验证
- 评估模型性能：预测模型
- 评估模型性能：分类模型
- 受试者操作特征 (ROC) 曲线

### 决策树

- 决策树简介
- 分类树
- 回归树
- 带验证的决策树
- 随机 (Bootstrap) 森林

### 神经网络

- 什么是神经网络？
- 解释神经网络
- 带神经网络的预测性建模

### 广义回归

- 广义回归简介
- 使用最大似然法拟合模型
- 惩罚回归简介

### 模型比较和选择

- 比较预测模型

### 文本挖掘的介绍

- 文本挖掘简介
- 处理文本数据
- 管理词条列表
- 可视化和探索文本数据
- 分析（挖掘）文本数据

[jmp.com/statisticalthinking](http://jmp.com/statisticalthinking)



Statistical Discovery.™ From SAS.