



蜜雪冰城案例SPSS操作指南

通过数据分析揭示"最便宜反而最赚钱"的商业逻辑

研究背景与目标

蜜雪冰城作为中国领先的平价茶饮连锁品牌,拥有**22000+家门店**,日销售**1300万杯**。本研究收集了1000家样本门店的经营数据,希望通过相关分析和多元线性回归建模,量化各关键因素对门店盈利能力的影响程度。

我们将深入探讨客单价、选址类型、门店面积、门店密度、日销量、租金等级、供应链成熟度和品牌影响力等8个关键指标如何影响月净利润。

22K+

门店数量

1300万

日销售杯数

1000

样本门店

核心研究变量



客单价

平均每位顾客消费金额,范围6.4-28.6元



选址类型

是否位于学校周边(0=否,1=是)



门店面积

门店营业面积,范围15-85平方米



门店密度

区域内门店集中程度,1.0-5.0分

日销量

门店日均销售量,200-1200单位

租金等级

门店租金水平(1=低,2=中,3=高)

供应链成熟度

供应链体系完善程度,1.0-5.0分

品牌影响力

基于市场份额的影响力指数,1.5-5.0

数据准备与导入

()

打开SPSS软件

启动SPSS统计分析软件

([

导入数据文件

点击文件→打开→数据,选择mixue_data_original.csv文件

({

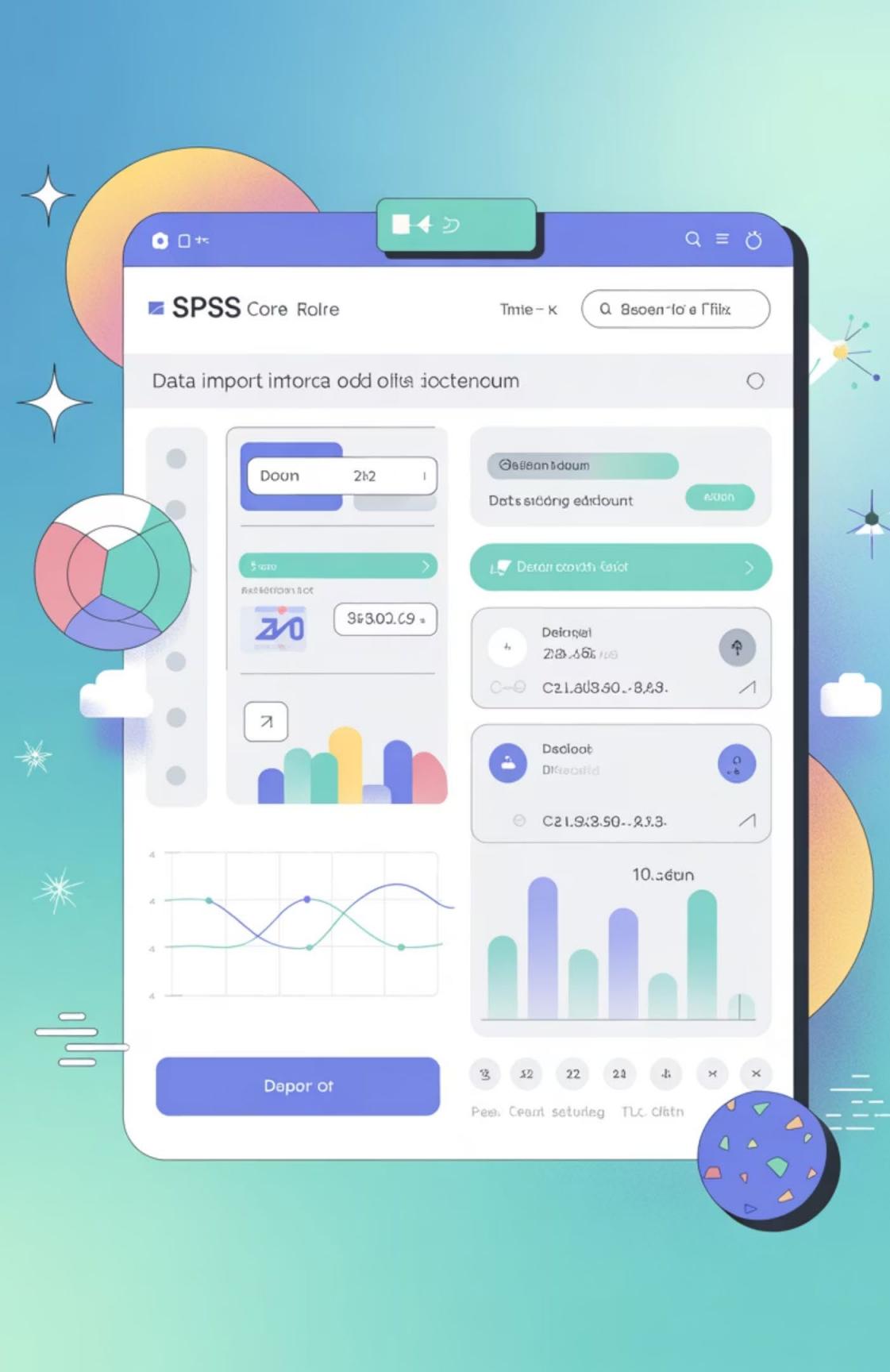
配置导入向导

选择"分隔",分隔符选择"逗号",勾选第一行包含变量名称

(}

完成导入

点击"完成"按钮,数据成功导入SPSS



变量定义与标签设置

在变量视图中进行详细设置,确保每个变量都有正确的测量尺度和标签。这是数据分析的基础步骤,直接影响后续分析结果的准确性。

| 变量名 | 标签 | 测量尺度 | 数据范围 |
|--------------------|---------------|------|------------|
| client_unit_price | 客单价(元) | 标量 | 6.4-28.6 |
| location_type | 选址类型 | 名义 | 0=非学校,1=学校 |
| store_area | 门店面积(m^2) | 标量 | 15-85 |
| daily_sales_volume | 日销量(单位数) | 标量 | 200-1200 |
| monthly_profit | 月净利润(万元) | 标量 | 30.0-70.0 |

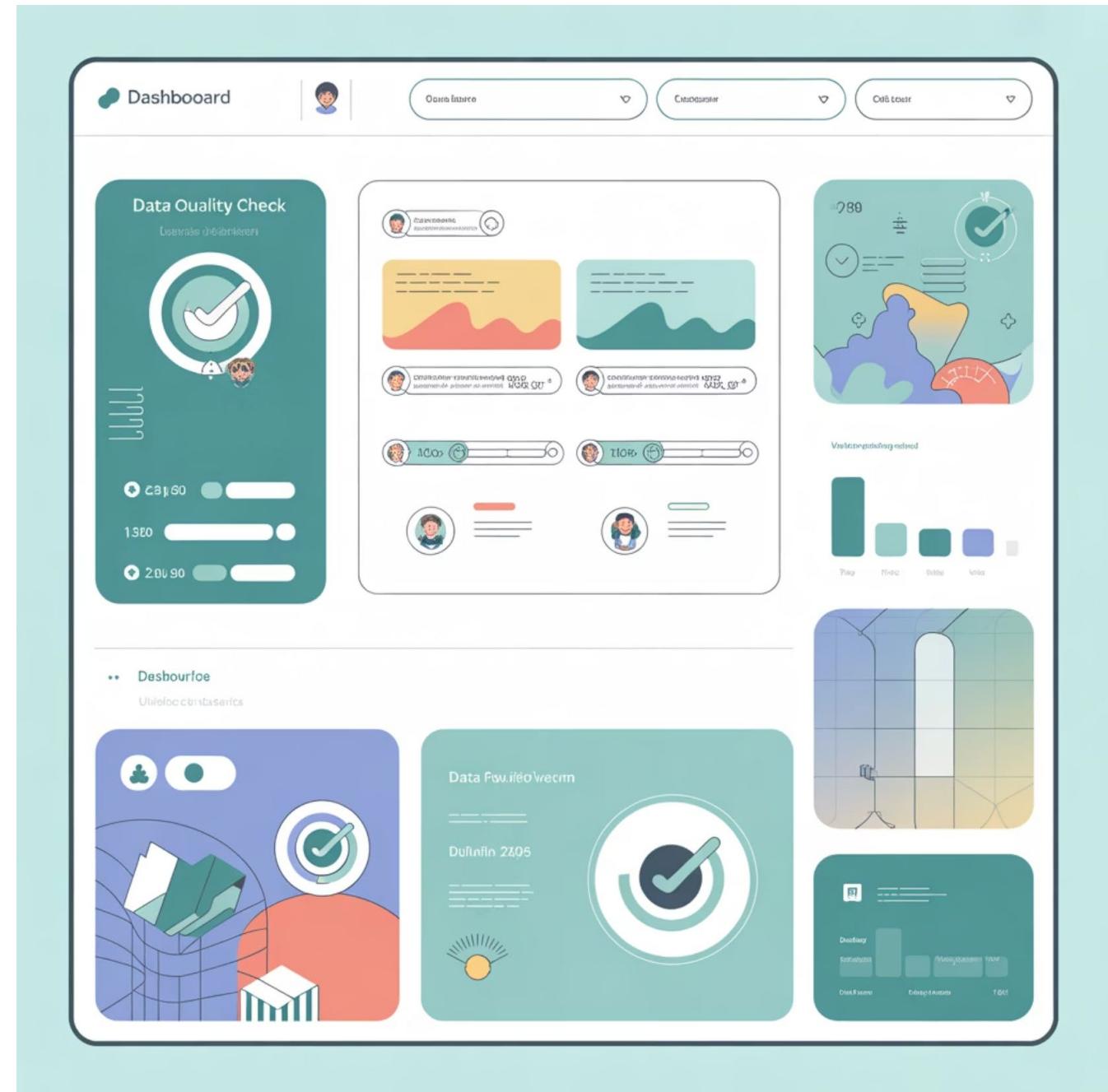
数据质量检查流程

检查样本量

- 点击分析→描述统计→频率
- 选择任一变量查看有效个案数
- 确认样本量为1000

检查缺失值

- 点击分析→描述统计→描述
- 选择所有变量
- 查看有效N,确保数据完整性



描述性统计分析



基本描述统计

点击分析→描述统计→描述,选择关键变量,统计量包括均值、标准差、最小值、最大值、范围、偏度和峰度

分组描述统计

点击分析→比较均值→均值,因变量为月净利润,独立变量为选址类型和租金等级,比较不同组别间的盈利差异

分布检验

点击分析→描述统计→探索,进行正态性检验,查看直方图和正态Q-Q图,确保数据符合分析要求

- 操作目的:了解各变量的基本统计特征,检查数据分布是否正常,识别可能的异常值,为后续回归分析提供分组对比基础。

相关分析操作步骤

皮尔逊相关分析

1. 点击分析→相关→双变量
2. 选择所有连续变量:客单价、门店面积、门店密度、日销量、供应链成熟度、品牌影响力、月净利润
3. 相关系数选择Pearson,显著性检验选择双侧
4. 勾选标记显著相关,点击"确定"

分类变量关联分析

点击分析→描述统计→交叉表,行为选址类型,列为租金等级,统计量勾选卡方检验和相关性,检验选址类型与租金等级是否存在关联。



观察相关系数

查看大小和方向



注意显著性标记

* $p < 0.05$, ** $p < 0.01$



识别强相关关系

$|r| > 0.7$ 为强相关

多元线性回归建模

建立回归模型

点击分析→回归→线性,因变量为月净利润,自变量包括客单价、选址类型、门店面积、门店密度、日销量、租金等级、供应链成熟度、品牌影响力,方法选择"输入"

统计量设置

点击"统计量"按钮,勾选估计、模型拟合、R平方变化、描述统计,确保输出完整的分析结果

残差分析设置

点击"绘图"按钮,Y轴选择ZRESID(标准化残差),X轴选择ZPRED(标准化预测值),勾选正态概率图和直方图



回归分析输出结果

模型摘要

包含R²、调整R²、标准误差,评估模型拟合度

方差分析表

F检验结果,验证模型整体显著性

回归系数表

系数值、标准误、t值、p值,量化各因素影响

共线性统计量

容忍度、VIF值,检测多重共线性问题

模型诊断与验证

多重共线性检验

1

检查自变量间是否存在严重共线性。容忍度 >0.1 为可接受, VIF值 <10 为可接受, 条件索引 <30 为可接受。在回归分析输出的"共线性统计量"表中查看相关指标。

残差诊断

2

验证回归模型的基本假设。查看残差直方图应近似正态分布, Q-Q图的点应接近直线, Kolmogorov-Smirnov检验 $p>0.05$ 为正态。观察标准化残差vs标准化预测值散点图应呈随机分布。

3

异常值检测

标准化残差查看 $>|3|$ 的个案, Cook距离查看 >1 的个案, 识别可能影响模型稳定性的异常观测值。

交叉验证方法

方法一:简单数据分割验证

创建随机分组变量

变换→计算变量,目标变量为validation_group,数值表达式为RV.UNIFORM(0,1)

创建训练集标识

变换→计算变量,目标变量为train_set,表达式为validation_group<=0.7(70%作为训练集)

训练集建模

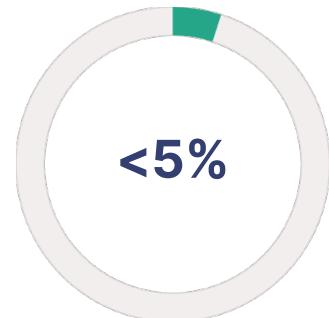
数据→选择个案,条件为train_set=1,运行回归分析并记录R²

测试集验证

选择个案条件为train_set=0,运行回归分析,比较训练集与测试集R²的差异

方法二:重复随机抽样验证

数据→选择个案→随机样本,选择样本大致为总数的70%,在70%数据上运行回归并记录R²。重复5-10次相同过程,每次重新随机抽样,记录每次的R²值,计算平均R²和标准差。



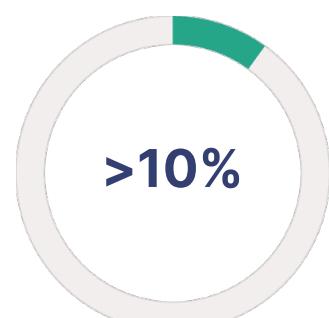
模型稳定

变异系数小于5%



模型较稳定

变异系数在5-10%之间



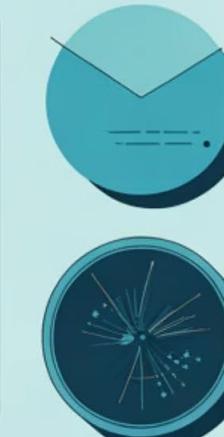
模型不稳定

变异系数大于10%

Statistical Interpretation



3.4



关键指标解读

模型拟合度

- **R²**:决定系数,表示模型解释变异的比例
- **调整R²**:考虑自变量数量的修正R²
- **F检验**:模型整体显著性检验

回归系数

- **系数值**:自变量对因变量的影响程度
- **t值**:系数显著性检验统计量
- **p值**:显著性水平

模型诊断

- **VIF**:方差膨胀因子,检测多重共线性
- **残差图**:检验模型假设

SPSS操作流程总览



操作要点:分类变量必须正确设置测量尺度,回归分析前检查变量分布,注意解读显著性标记,保存分析语法以便重复操作,及时备份数据和结果。

常见注意事项与文件管理

常见注意事项



正确设置测量尺度

分类变量必须设置为名义或序数尺度



检查变量分布

回归分析前确认数据符合正态分布



解读显著性标记

注意*、**、***的含义和差异



保存分析语法

便于重复操作和结果验证

输出文件管理



保存SPSS数据文件

保存为.sav格式,保留所有变量定义



导出结果

导出到Word或Excel,便于报告撰写



保存语法文件

保存为.sps格式,便于重复分析



记录操作步骤

记录关键操作步骤和参数设置