

SPSS 完全学习手册：从入门到数据分析大师

前言

在数据驱动决策的时代，SPSS (Statistical Product and Service Solutions) 作为一款功能强大的统计分析软件，已成为科研工作者、数据分析人员、学生及职场人士的必备工具。它以直观的图形化操作界面、丰富的统计分析功能和便捷的数据处理能力，广泛应用于社会学、经济学、医学、教育学、心理学等多个领域。

本书专为 SPSS 初学者和有一定基础但希望系统提升的读者打造，摒弃了复杂晦涩的理论堆砌，以“实操为主、理论为辅”的原则，从软件安装与界面认知入手，逐步深入数据录入与清洗、描述性统计分析、推论性统计分析、回归分析、聚类分析等核心模块，每个知识点均搭配详细的操作步骤、案例演示和结果解读，帮助读者快速掌握 SPSS 的使用技巧，真正将统计分析方法落地到实际工作中。

无论你是准备撰写学术论文、开展市场调研，还是需要进行数据可视化呈现，本书都能为你提供全方位的指导。跟随书中的步骤循序渐进学习，你将轻松突破 SPSS 操作瓶颈，实现从“数据小白”到“数据分析高手”的蜕变。

第一部分：SPSS 基础入门

第 1 章：认识 SPSS——软件概述与安装

1.1 SPSS 的发展与应用领域

SPSS 最初由斯坦福大学的三位学生于 1968 年开发，全称为“Statistical Package for the Social Sciences”（社会科学统计软件包），后更名为“Statistical Product and Service Solutions”，但简称 SPSS 仍被广泛使用。经过半个多世纪的迭代升级，SPSS 已从最初的单一统计工具，发展为集数据录入、清洗、分析、可视化于一体的综合性软件。

其应用领域极为广泛：

- 学术研究：用于问卷调查数据处理、实验数据统计分析、论文实证研究等；
- 商业领域：市场调研数据解读、消费者行为分析、销售业绩预测等；
- 医疗卫生：临床试验数据统计、疾病发生率分析、疗效评估等；
- 教育领域：学生成绩分析、教学效果评估、问卷调查研究等。

1.2 SPSS 的版本与安装步骤

1.2.1 常见版本

目前 SPSS 的主流版本包括 26.0、27.0、28.0 及最新的 29.0 版本，各版本在核心功能上差异不大，主要优化了操作界面、增加了部分高级分析模块和数据可视化效果。对于初学者而言，选择任一版本均可满足学习需求，本书将以 SPSS 28.0 为例进行操作演示。

1.2.2 安装步骤 (Windows 系统)

1. 准备 SPSS 28.0 安装包（可通过官方网站购买或正规渠道获取），双击安装文件，进入安装向导界面，点击“下一步”；
2. 阅读许可协议，勾选“我接受协议”，点击“下一步”；
3. 选择安装类型，推荐“完整安装”（适合初学者），如需自定义安装路径，可点击“浏览”修改，点击“下一步”；
4. 等待文件解压与安装，过程约 5-10 分钟（取决于电脑配置）；
5. 安装完成后，弹出许可证激活界面，输入购买的许可证密钥，点击“激活”；
6. 激活成功后，点击“完成”，重启电脑即可启动 SPSS。

1.2.3 安装常见问题解决

- 安装失败提示“缺少 .NET Framework”：需先安装对应版本的 .NET Framework（可通过微软官网下载）；
- 激活失败：检查许可证密钥是否正确，或确认安装包是否完整，若仍无法激活，联系官方客服咨询；
- 启动后闪退：关闭电脑中的杀毒软件，或重新安装 SPSS，确保安装路径无中文。

1.3 SPSS 的界面认知

启动 SPSS 后，将进入软件主界面，主要由以下 5 个核心部分组成：

1.3.1 数据编辑器窗口

这是 SPSS 的核心操作窗口，分为“数据视图”和“变量视图”两个标签页：

- 数据视图：用于录入和查看原始数据，每行代表一个案例（如一个调查对象、一次实验数据），每列代表一个变量（如性别、年龄、成绩）；
- 变量视图：用于定义变量属性，包括变量名称、类型、宽度、小数位数、标签、值标签、缺失值、测量尺度等，是数据录入前的关键设置。

1.3.2 输出窗口

执行统计分析操作后，结果将自动显示在输出窗口中，包括统计表格、图表、文字说明等。用户可对输出结果进行编辑、复制、保存（格式为.spv），或导出为 Word、Excel、PDF 等格式。

1.3.3 菜单栏

位于界面顶部，包含“文件”“编辑”“视图”“数据”“分析”“图形”“实用程序”“窗口”“帮助”9 个主菜单，几乎所有 SPSS 功能都可通过菜单栏调用。

1.3.4 工具栏

位于菜单栏下方，包含常用功能的快捷按钮，如新建数据文件、打开文件、保存文件、撤销、重做、运行分析、插入变量、插入案例等，点击即可快速执行对应操作。

1.3.5 状态栏

位于界面底部，显示当前操作状态，如“就绪”“正在处理数据”等，帮助用户了解软件运行情况。

第 2 章：数据录入与管理 —— 数据分析的基础

2.1 变量定义：变量视图的核心设置

在录入数据前，需先在“变量视图”中定义变量属性，确保数据录入的规范性和准确性，避免后续分析出错。以下是变量视图中关键属性的设置方法：

2.1.1 变量名称

- 命名规则：最多 64 个字符，可包含字母、数字、下划线，不能以数字或下划线开头，不能包含空格和特殊字符（如!、@、# 等）；
- 示例：性别、年龄、学历、月收入、满意度。

2.1.2 类型

用于设置变量的数据类型，常用类型包括：

- 数值：用于存储整数、小数（如年龄、收入、成绩）；
- 字符串：用于存储文本信息（如姓名、地址、学历名称）；
- 日期：用于存储日期数据（如出生日期、调查日期），需选择对应的日期格式（如 YYYY-MM-DD）。

2.1.3 宽度与小数位数

- 宽度：变量的总字符数（数值型变量包括整数部分、小数点和小数部分）；

- 小数位数：数值型变量的小数保留位数，如年龄可设为 0 位小数，收入可设为 2 位小数。

2.1.4 标签

变量名称的详细说明，当变量名称过于简洁时，标签可帮助用户理解变量含义（如变量名称为“sex”，标签可设为“性别”）。

2.1.5 值标签

用于定义分类变量的取值含义（仅适用于数值型分类变量），如性别变量：1=“男”，2=“女”；学历变量：1=“初中及以下”，2=“高中 / 中专”，3=“大专”，4=“本科”，5=“硕士及以上”。设置后，在数据视图中输入 1、2 等数字时，会自动显示对应标签，方便数据录入和查看。

2.1.6 缺失值

定义缺失数据的处理方式，常用“无”“离散缺失值”“范围缺失值”：

- 无：不设置缺失值，所有数据均视为有效；
- 离散缺失值：输入具体的缺失值（如 99、999），表示这些数值为无效数据；
- 范围缺失值：设置缺失值的范围（如 ≤ 0 ），范围内的数值视为无效数据。

2.1.7 测量尺度

定义变量的测量类型，直接影响统计分析方法的选择：

- 定类变量：用于分类，无顺序和大小之分（如性别、职业、民族），测量尺度设为“名义”；
- 定序变量：有分类且有顺序，但无固定间隔（如满意度：非常不满意、不满意、一般、满意、非常满意），测量尺度设为“有序”；
- 定距变量：有顺序、有固定间隔，但无绝对零点（如温度、成绩）；
- 定比变量：有顺序、有固定间隔且有绝对零点（如年龄、收入、身高），定距和定比变量的测量尺度均设为“度量”。

2.2 数据录入：数据视图的操作技巧

变量定义完成后，切换到“数据视图”进行数据录入，以下是实用操作技巧：

2.2.1 手动录入

- 逐行录入：每行对应一个案例，点击单元格直接输入数据，分类变量可通过下拉菜单选择值标签（需先设置值标签）；
- 批量录入：若数据已存在于 Excel 中，可直接复制 Excel 中的数据，粘贴到 SPSS 数据视图中（注意变量顺序一致）。

2.2.2 数据录入规范

- 数值型变量：直接输入数字，避免输入文本（如“二十”需改为 20）；
- 字符串变量：输入文本时，避免多余空格（如“男”需改为“男”）；
- 缺失数据：按变量视图中设置的缺失值规则录入（如 99），或直接留空（SPSS 默认留空为缺失值）。

2.2.3 数据录入快捷操作

- 插入变量：点击工具栏“插入变量”，在当前变量列前新增一列；
- 插入案例：点击工具栏“插入案例”，在当前行前新增一行；
- 复制 / 粘贴数据：选中单元格、行或列，使用 Ctrl+C、Ctrl+V 快捷键复制粘贴；
- 查找替换：点击“编辑”→“查找”，输入查找内容和替换内容，可批量修改错误数据。

2.3 数据清洗：处理缺失值、异常值与重复值

数据录入后，需进行数据清洗，确保数据质量，避免影响分析结果的准确性。

2.3.1 缺失值处理

1. 查看缺失值：点击“分析”→“描述统计”→“描述”，将需要检查的变量移入“变量”框，点击“选项”，勾选“缺失值”，运行后可在输出窗口查看各变量的缺失值数量和比例；
2. 缺失值处理方法：
 - 删除法：若缺失值比例低于 5%，且缺失数据为随机分布，可直接删除包含缺失值的案例（点击“数据”→“选择个案”，勾选“排除个案：基于缺失值”）；
 - 替换法：对于数值型变量，用均值、中位数或众数替换缺失值（点击“分析”→“缺失值分析”→“替换缺失值”，选择替换方法）；
 - 插值法：通过回归分析、线性插值等方法估算缺失值（适合缺失值比例在 5%-20% 之间）。

2.3.2 异常值处理

异常值是指与其他数据差异过大的极端值，可能由数据录入错误或真实异常情况导致。

1. 检测异常值：
 - 箱线图法：点击“图形”→“旧对话框”→“箱线图”，选择“简单箱线图”，将变量移入“Y 轴”，运行后，箱线图外的点即为异常值；
 - Z 分数法：点击“分析”→“描述统计”→“描述”，勾选“将标准化值另存为变量”，运行后，Z 分数绝对值大于 3 的案例视为异常值；

1. 处理异常值:

- 核实修正: 若为录入错误, 修改为正确数据;
- 删除法: 若异常值为真实数据但影响分析结果, 可删除该案例 (需谨慎使用);
- 转换法: 对变量进行对数转换、标准化处理, 降低异常值的影响。

2.3.3 重复值处理

1. 查找重复值: 点击“数据”→“标识重复个案”, 将唯一标识变量 (如身份证号、编号) 移入“定义匹配个案的依据”, 运行后, SPSS 会新增一列标识重复案例;
2. 删除重复值: 点击“数据”→“删除重复个案”, 选择唯一标识变量, 保留“第一个匹配个案”或“最后一个匹配个案”, 删除其余重复案例。

2.4 数据保存与格式转换

2.4.1 数据保存

- 保存为 SPSS 格式: 点击“文件”→“保存”, 默认格式为.sav (SPSS 数据文件), 可直接在 SPSS 中打开编辑;
- 另存为其他格式: 点击“文件”→“另存为”, 可选择 Excel (.xlsx)、CSV (.csv)、文本文件 (.txt) 等格式, 方便与其他软件共享数据。

2.4.2 格式转换

- Excel 转 SPSS: 点击“文件”→“打开”→“数据”, 选择 Excel 文件 (.xlsx), 勾选“第一行作为变量名”, 点击“确定”, 即可将 Excel 数据导入 SPSS;
- CSV 转 SPSS: 同理, 选择 CSV 文件, 设置编码格式 (如 UTF-8), 即可导入;
- SPSS 转 Excel/CSV: 按“数据保存”中的“另存为”操作即可。

第二部分：核心统计分析功能

第 3 章：描述性统计分析 —— 数据的基本特征解读

描述性统计分析是数据分析的第一步, 通过统计量和图表描述数据的基本特征, 包括集中趋势、离散程度、分布形态等, 帮助用户快速了解数据全貌。

3.1 常用描述性统计量

3.1.1 集中趋势统计量

用于描述数据的中心位置, 常用指标包括:

- 均值 (Mean) : 所有数据的平均值, 适用于定距 / 定比变量, 易受异常值影响;
- 中位数 (Median) : 将数据从小到大排序后, 中间位置的数值, 适用于定序 / 定距 / 定比变量, 不受异常值影响;
- 众数 (Mode) : 数据中出现频率最高的数值, 适用于所有类型变量, 可用于判断数据的常见取值。

3.1.2 离散程度统计量

用于描述数据的分散程度, 常用指标包括:

- 标准差 (Std. Deviation) : 数据偏离均值的平均程度, 标准差越大, 数据离散程度越高, 适用于定距 / 定比变量;
- 方差 (Variance) : 标准差的平方, 同样反映数据离散程度;
- 极差 (Range) : 最大值与最小值的差值, 简单直观但易受异常值影响;
- 四分位距 (IQR) : 上四分位数 (Q3) 与下四分位数 (Q1) 的差值, 不受异常值影响, 适用于定序 / 定距 / 定比变量。

3.1.3 分布形态统计量

用于判断数据是否符合正态分布, 常用指标包括:

- 偏度 (Skewness) : 描述数据分布的不对称程度, 偏度 = 0 为对称分布, 偏度 > 0 为右偏分布 (长尾在右), 偏度 < 0 为左偏分布 (长尾在左);
- 峰度 (Kurtosis) : 描述数据分布的陡峭程度, 峰度 = 0 为正态分布 (mesokurtic), 峰度 > 0 为尖峰分布 (leptokurtic), 峰度 < 0 为平峰分布 (platykurtic)。

3.2 SPSS 操作: 描述性统计分析步骤

以“某班级学生数学、语文、英语成绩”数据为例, 演示描述性统计分析操作:

1. 打开 SPSS 数据文件, 确保成绩变量为“度量”尺度;
2. 点击“分析”→“描述统计”→“描述”, 弹出描述性统计对话框;
3. 将“数学成绩”“语文成绩”“英语成绩”移入“变量”框;
4. 点击“选项”, 勾选需要的统计量 (如均值、标准差、中位数、范围、偏度、峰度), 勾选“显示顺序”为“按变量顺序”, 点击“继续”;
5. 点击“确定”, 运行分析, 在输出窗口查看结果。

3.3 结果解读: 统计表格与图表分析

3.3.1 统计表格解读

输出结果中的“描述统计量”表格包含各变量的样本量（N）、均值、标准差、最小值、最大值、中位数、偏度、峰度等信息：

- 样本量（N）：无缺失值的案例数量，若 N 小于总案例数，说明存在缺失值；
- 均值与中位数：若两者接近，说明数据分布较对称；若差异较大，可能存在异常值或偏态分布；
- 标准差：如数学成绩标准差为 10.2，语文成绩标准差为 8.5，说明数学成绩的离散程度高于语文成绩；
- 偏度与峰度：若偏度绝对值峰度绝对值可认为数据近似正态分布；若超出该范围，需考虑数据转换或选择非参数检验方法。

3.3.2 数据可视化：常用描述性图表

除了统计量，图表能更直观地展示数据特征，SPSS 中常用的描述性图表包括：

1. 直方图（含正态曲线）

- 操作：点击“图形”→“旧对话框”→“直方图”，将变量移入“变量”框，勾选“显示正态曲线”，点击“确定”；
- 解读：观察直方图形状，若与正态曲线贴合较好，说明数据近似正态分布；若呈现明显的左偏或右偏，需结合偏度指标判断。

1. 箱线图

- 操作：点击“图形”→“旧对话框”→“箱线图”，选择“简单箱线图”，将变量移入“Y 轴”，点击“确定”；
- 解读：箱线图的箱体代表四分位距（Q1-Q3），中间的线为中位数，箱体上下的须代表极差（排除异常值），箱体外部的点为异常值。

1. 条形图

- 操作：点击“图形”→“旧对话框”→“条形图”，选择“简单条形图”，将分类变量移入“X 轴”，数值变量移入“Y 轴”，点击“确定”；
- 解读：适用于展示分类变量的分布情况（如不同学历的人数分布）。

第 4 章：推论性统计分析 —— 从样本推断总体

推论性统计分析基于样本数据推断总体的特征，核心是假设检验，包括参数检验和非参数检验，常用于比较两组或多组数据的差异、分析变量间的关联等。

4.1 假设检验的基本原理

4.1.1 核心概念

- 总体与样本：总体是研究对象的全部集合，样本是从总体中抽取的部分个体；
- 原假设 (H_0)：默认的、无差异的假设（如“两组数据的均值相等”“变量间无关联”）；
- 备择假设 (H_1)：与原假设对立的假设（如“两组数据的均值不相等”“变量间有关联”）；
- 显著性水平 (α)：判断假设是否成立的标准，常用 $\alpha=0.05$ （即 5% 的犯错概率）；
- P 值：假设原假设成立时，观察到当前样本结果的概率，若 $P < \alpha$ ，拒绝原假设；若 $P \geq \alpha$ ，接受原假设。

4.1.2 假设检验的步骤

1. 建立原假设 (H_0) 和备择假设 (H_1)；
2. 选择合适的检验方法（根据变量类型、样本量、数据分布等）；
3. 计算检验统计量和 P 值；
4. 对比 P 值与显著性水平 α ，做出统计决策；
5. 结合实际问题，解读检验结果。

4.2 参数检验：基于正态分布的检验方法

参数检验要求数据满足正态分布、方差齐性等前提条件，适用于定距 / 定比变量，常用方法包括独立样本 t 检验、配对样本 t 检验、单因素方差分析 (ANOVA)。

4.2.1 独立样本 t 检验：比较两组独立样本的均值

适用场景：比较两个独立群体的某个指标是否存在显著差异（如男性与女性的收入差异、实验组与对照组的实验结果差异）。

操作步骤（以“男性与女性的数学成绩差异”为例）：

1. 打开数据文件，确保“数学成绩”为“度量”尺度，“性别”为“名义”尺度（值标签：1 = 男，2 = 女）；
2. 点击“分析”→“比较均值”→“独立样本 t 检验”，弹出对话框；
3. 将“数学成绩”移入“检验变量”框，将“性别”移入“分组变量”框；
4. 点击“定义组”，输入分组值（组 1=1，组 2=2），点击“继续”；
5. 点击“确定”，运行分析。

结果解读：

- 方差齐性检验 (Levene 检验)：若 $P > 0.05$ ，说明两组方差齐性，看“假设方差相等”行的结果；若 $P < 0.05$ ，说明方差不齐，看“假设方差不相等”行的结果；

- t 检验结果：关注 t 值、自由度 (df) 和 P 值 (Sig. 双侧)，若 $P < 0.05$ ，拒绝原假设，说明两组均值存在显著差异；若 $P \geq 0.05$ ，接受原假设，说明两组均值无显著差异。

4.2.2 配对样本 t 检验：比较两组相关样本的均值

适用场景：比较同一群体在不同条件下的指标差异（如学生课前与课后的成绩差异、同一产品使用前后的满意度差异）。

操作步骤（以“学生课前与课后的数学成绩差异”为例）：

1. 打开数据文件，确保“课前成绩”“课后成绩”均为“度量”尺度；
2. 点击“分析”→“比较均值”→“配对样本 t 检验”，弹出对话框；
3. 将“课前成绩”和“课后成绩”移入“成对变量”框（成对 1：课前成绩→变量 1，课后成绩→变量 2）；
4. 点击“选项”，设置置信区间百分比（默认 95%），点击“继续”；
5. 点击“确定”，运行分析。

结果解读：

- 首先查看配对样本的相关性分析，若 $P < 0.05$ ，说明两组数据存在显著相关，适合进行配对 t 检验；
- 配对 t 检验结果：关注 t 值、自由度 (df) 和 P 值 (Sig. 双侧)，若 $P < 0.05$ ，说明两组均值存在显著差异（如课后成绩显著高于课前成绩）；若 $P \geq 0.05$ ，说明两组均值无显著差异。

4.2.3 单因素方差分析 (ANOVA)：比较多组独立样本的均值

适用场景：比较三个及以上独立群体的某个指标是否存在显著差异（如不同学历群体的收入差异、不同教学方法的学生成绩差异）。

操作步骤（以“不同学历群体的月收入差异”为例）：

1. 打开数据文件，确保“月收入”为“度量”尺度，“学历”为“名义”尺度（值标签：1 = 初中及以下，2 = 高中 / 中专，3 = 大专，4 = 本科，5 = 硕士及以上）；
2. 点击“分析”→“比较均值”→“单因素 ANOVA”，弹出对话框；
3. 将“月收入”移入“因变量”框，将“学历”移入“因子”框；
4. 点击“事后检验”，勾选常用的事后检验方法（如 LSD、SNK），用于进一步分析具体哪两组存在差异，点击“继续”；
5. 点击“选项”，勾选“描述性”“方差齐性检验”，点击“继续”；
6. 点击“确定”，运行分析。

结果解读：

- 方差齐性检验 (Levene 检验)：若 $P > 0.05$ ，说明多组方差齐性，可进行后续 ANOVA 分析；若 $P \leq 0.05$ ，需选择校正的方差分析方法（如 Welch 检验）；
- ANOVA 结果：关注 F 值、自由度 (df1、df2) 和 P 值 (Sig.)，若 $P \leq 0.05$ ，拒绝原假设，说明至少有两组均值存在显著差异；若 $P > 0.05$ ，接受原假设，说明多组均值无显著差异；
- 事后检验结果：若 ANOVA 结果显著，查看事后检验表格，P 的组别对说明存在显著差异（如本科与大专的收入存在显著差异，本科与硕士及以上的收入无显著差异）。

4.3 非参数检验：不依赖正态分布的检验方法

当数据不满足正态分布、方差齐性等参数检验前提，或变量为定类 / 定序变量时，需使用非参数检验，常用方法包括卡方检验、曼 - 惠特尼 U 检验、威尔科克森符号秩检验、克鲁斯卡尔 - 沃利斯 H 检验。

4.3.1 卡方检验：分析分类变量的分布或关联

适用场景：

- 拟合优度检验：检验分类变量的实际分布是否与理论分布一致（如性别比例是否为 1:1、不同年龄段的人数分布是否符合预期）；
- 独立性检验：检验两个分类变量是否存在关联（如性别与学历是否有关联、职业与满意度是否有关联）。

操作步骤（以“性别与满意度的独立性检验”为例）：

1. 打开数据文件，确保“性别”（1 = 男，2 = 女）和“满意度”（1 = 非常不满意，2 = 不满意，3 = 一般，4 = 满意，5 = 非常满意）均为“名义”或“有序”尺度；
2. 点击“分析”→“描述统计”→“交叉表”，弹出对话框；
3. 将“性别”移入“行”框，将“满意度”移入“列”框；
4. 点击“统计量”，勾选“卡方”，点击“继续”；
5. 点击“单元格”，勾选“观察值”“期望值”，点击“继续”；
6. 点击“确定”，运行分析。

结果解读：

- 关注“皮尔逊卡方”的卡方值、自由度 (df) 和 P 值 (Sig.)，若 $P \leq 0.05$ ，拒绝原假设，说明两个分类变量存在显著关联（如性别与满意度有关联，女性满意度高于男性）；若 $P > 0.05$ ，接受原假设，说明两个分类变量无显著关联；
- 注意：若单元格的期望值小于 5 的数量超过总单元格数的 20%，需合并相邻类别后重新检验，避免结果偏差。

4.3.2 曼 - 惠特尼 U 检验：独立样本 t 检验的非参数替代

适用场景：比较两组独立样本的分布是否存在显著差异（数据不满足正态分布，或变量为定序变量），对应参数检验中的独立样本 t 检验。

操作步骤（以“男性与女性的满意度差异”为例，满意度为定序变量）：

1. 打开数据文件，“满意度”为“有序”尺度，“性别”为“名义”尺度；
2. 点击“分析”→“非参数检验”→“旧对话框”→“2 个独立样本”，弹出对话框；
3. 将“满意度”移入“检验变量列表”，将“性别”移入“分组变量”，定义组 1=1，组 2=2；
4. 勾选“曼 - 惠特尼 U”检验方法，点击“确定”，运行分析。

结果解读：

- 关注曼 - 惠特尼 U 值、Z 值和 P 值（Sig. 双侧），若 $P < 0.05$ ，拒绝原假设，说明两组样本的分布存在显著差异（如女性的满意度显著高于男性）；若 $P \geq 0.05$ ，接受原假设，说明两组样本的分布无显著差异。

4.3.3 威尔科克森符号秩检验：配对样本 t 检验的非参数替代

适用场景：比较两组相关样本的分布是否存在显著差异（数据不满足正态分布），对应参数检验中的配对样本 t 检验。

操作步骤（以“学生课前与课后的满意度差异”为例）：

1. 打开数据文件，“课前满意度”“课后满意度”均为“有序”尺度；
2. 点击“分析”→“非参数检验”→“旧对话框”→“2 个相关样本”，弹出对话框；
3. 将“课前满意度”和“课后满意度”移入“成对变量”框；
4. 勾选“威尔科克森”检验方法，点击“确定”，运行分析。

结果解读：

- 关注 Z 值和 P 值（Sig. 双侧），若 $P < 0.05$ ，拒绝原假设，说明两组样本的分布存在显著差异（如课后满意度显著高于课前满意度）；若 $P \geq 0.05$ ，接受原假设，说明两组样本的分布无显著差异。

4.3.4 克鲁斯卡尔 - 沃利斯 H 检验：单因素 ANOVA 的非参数替代

适用场景：比较三个及以上独立样本的分布是否存在显著差异（数据不满足正态分布，或变量为定序变量），对应参数检验中的单因素 ANOVA。

操作步骤（以“不同学历群体的满意度差异”为例）：

1. 打开数据文件，“满意度”为“有序”尺度，“学历”为“名义”尺度；
2. 点击“分析”→“非参数检验”→“旧对话框”→“K 个独立样本”，弹出对话框；
3. 将“满意度”移入“检验变量列表”，将“学历”移入“分组变量”，定义分组范围（如 1-5）；

4. 勾选“克鲁斯卡尔 - 沃利斯 H”检验方法，点击“确定”，运行分析。

结果解读：

- 关注卡方值、自由度 (df) 和 P 值 (Sig.)，若 $P < 0.05$ ，拒绝原假设，说明至少有两组样本的分布存在显著差异；若 $P \geq 0.05$ ，接受原假设，说明多组样本的分布无显著差异；
- 若结果显著，可通过两两比较（如调整显著性水平的曼 - 惠特尼 U 检验）进一步分析具体哪两组存在差异。

第 5 章：回归分析 —— 变量间的因果关系探究

回归分析是研究变量间因果关系的核心方法，通过建立回归模型，分析自变量对因变量的影响程度和方向，常用于预测和解释变量间的关系。

5.1 回归分析的基本概念

5.1.1 自变量与因变量

- 因变量 (Dependent Variable)：被预测或被解释的变量（如收入、成绩、满意度），需为定距 / 定比变量；
- 自变量 (Independent Variable)：影响因变量的变量（如年龄、学历、工作年限），可为定类、定序或定距 / 定比变量（定类 / 定序变量需进行编码处理）。

5.1.2 回归模型的类型

- 一元线性回归：仅包含一个自变量的线性回归模型，用于分析单个变量对因变量的影响；
- 多元线性回归：包含两个及以上自变量的线性回归模型，用于分析多个变量对因变量的综合影响；
- 非线性回归：自变量与因变量呈非线性关系（如二次函数、指数函数）的回归模型；
- 逻辑回归：因变量为二分类变量（如是否购买、是否患病）的回归模型。

5.1.3 回归分析的前提假设（线性回归）

- 线性性：自变量与因变量之间存在线性关系；
- 独立性：残差（实际值与预测值的差值）相互独立；
- 正态性：残差服从正态分布；
- 同方差性：残差的方差在所有自变量取值范围内保持一致（方差齐性）。

5.2 一元线性回归分析

适用场景：分析单个自变量对因变量的线性影响（如工作年限对收入的影响、学习时间对成绩的影响）。

5.2.1 操作步骤（以“工作年限对月收入的影响”为例）

1. 打开数据文件，确保“月收入”（因变量）和“工作年限”（自变量）均为“度量”尺度；
2. 检验线性关系：点击“图形”→“旧对话框”→“散点/点图”，选择“简单散点图”，将“月收入”移入“Y轴”，“工作年限”移入“X轴”，点击“确定”，若散点图呈现明显的线性趋势，适合进行线性回归；
3. 点击“分析”→“回归”→“线性”，弹出对话框；
4. 将“月收入”移入“因变量”框，将“工作年限”移入“自变量”框；
5. 点击“统计量”，勾选“估计值”“模型拟合度”“共线性诊断”“残差”，点击“继续”；
6. 点击“图形”，将“ZRESID”移入“Y轴”，“ZPRED”移入“X轴”，勾选“正态概率图”，点击“继续”；
7. 点击“确定”，运行分析。

5.2.2 结果解读

1. 模型拟合度检验 (R^2)： R^2 为决定系数，取值范围 0-1， R^2 越接近 1，说明模型拟合效果越好（如 $R^2=0.65$ ，说明工作年限能解释 65% 的月收入变异）；
2. 回归方程显著性检验 (F 检验)：若 $P<0.05$ ，说明回归模型整体显著（即工作年限对月收入的线性影响显著）；
3. 回归系数检验 (t 检验)：关注自变量的回归系数 (B)、标准误、t 值和 P 值，若 $P<0.05$ ，说明自变量对因变量的影响显著；回归系数 B 的正负表示影响方向（如 $B=2000$ ，说明工作年限每增加 1 年，月收入平均增加 2000 元）；
4. 残差分析：通过残差散点图和正态概率图检验前提假设，若残差散点图无明显规律（随机分布），说明满足线性性和同方差性；若正态概率图的点接近对角线，说明残差服从正态分布；
5. 最终回归方程：月收入 = 截距 + B × 工作年限（如月收入 = 3000 + 2000 × 工作年限）。

5.3 多元线性回归分析

适用场景：分析多个自变量对因变量的综合线性影响（如年龄、学历、工作年限对收入的共同影响）。

5.3.1 操作步骤（以“年龄、学历、工作年限对月收入的影响”为例）

1. 打开数据文件，“月收入”为因变量（度量），“年龄”（度量）、“学历”（名义，需进行虚拟变量编码）、“工作年限”（度量）为自变量；

2. 虚拟变量编码（针对学历变量）：点击“数据”→“编码”→“重新编码为不同变量”，将“学历”移入“输入变量→输出变量”，设置输出变量名称（如“学历_本科”），点击“值”，将“本科”编码为 1，其他学历编码为 0，同理编码“学历_硕士及以上”“学历_大专”，以“初中及以下”为参照组；
3. 点击“分析”→“回归”→“线性”，将“月收入”移入“因变量”框，将“年龄”“学历_大专”“学历_本科”“学历_硕士及以上”“工作年限”移入“自变量”框；
4. 选择自变量进入模型的方法（常用“逐步”法，自动筛选显著的自变量）；
5. 重复一元线性回归的“统计量”“图形”设置，点击“确定”，运行分析。

5.3.2 结果解读

1. 模型拟合度检验（调整 R^2 ）：多元回归中需关注调整 R^2 （避免因自变量数量增加导致 R^2 虚高），调整 R^2 越接近 1，模型拟合效果越好；
2. 回归方程显著性检验（F 检验）：P.05，说明模型整体显著；
3. 回归系数检验（t 检验）：每个自变量的 P，说明该自变量对因变量的影响显著（控制其他自变量不变）；回归系数 B 表示该自变量每变化 1 单位，因变量的平均变化量（如“学历_本科”的 $B=3000$ ，说明在年龄和工作年限相同的情况下，本科学历比初中及以下学历的月收入平均高 3000 元）；
4. 共线性检验：查看“共线性诊断”中的方差膨胀因子（VIF）， $VIF>10$ 说明存在严重多重共线性（自变量间高度相关），需删除或合并相关自变量；
5. 最终回归方程：月收入 = 截距 + $B1 \times$ 年龄 + $B2 \times$ 学历_大专 + $B3 \times$ 学历_本科 + $B4 \times$ 学历_硕士及以上 + $B5 \times$ 工作年限。

5.4 逻辑回归分析

适用场景：因变量为二分类变量（如是否购买产品：是 = 1，否 = 0；是否患病：是 = 1，否 = 0），分析自变量对因变量分类结果的影响。

5.4.1 操作步骤（以“年龄、性别、收入对是否购买某产品的影响”为例）

1. 打开数据文件，“是否购买”为因变量（二分类，1 = 是，0 = 否，名义尺度），“年龄”（度量）、“性别”（名义，1 = 男，2 = 女）、“月收入”（度量）为自变量；
2. 点击“分析”→“回归”→“二元逻辑回归”，弹出对话框；
3. 将“是否购买”移入“因变量”框，将“年龄”“性别”“月收入”移入“协变量”框；
4. 点击“分类”，将“性别”移入“分类协变量”，设置参照组（如“最后一个类别”），点击“继续”；
5. 点击“统计量”，勾选“分类表”“方程中的变量”“Hosmer-Lemeshow 拟合度”，点击“继续”；

6. 点击“确定”，运行分析。

5.4.2 结果解读

1. 模型拟合度检验：

- Hosmer-Lemeshow 检验：若 $P > 0.05$ ，说明模型拟合效果较好（模型预测值与实际值无显著差异）；
- 伪 R^2 （如 Cox & Snell R^2 、Nagelkerke R^2 ）：取值范围 0-1，越接近 1，模型拟合效果越好（通常伪 R^2 达到 0.2-0.4 即为较好拟合）；

1. 回归系数检验（Wald 检验）：关注各自变量的 Wald 值、自由度和 P 值，若 $P > 0.05$ ，说明该自变量对因变量的分类有显著影响；回归系数 B 的正负表示影响方向，Exp (B)（优势比）表示自变量每变化 1 单位，因变量发生某类事件的优势比（如“月收入”的 Exp (B)=1.5，说明月收入每增加 1 单位，购买产品的优势是原来的 1.5 倍）；

2. 分类表：展示模型的预测准确率（如准确率 = 85%，说明模型能正确预测 85% 的案例）。

第三部分：高级分析与数据可视化

第 6 章：聚类分析与因子分析 —— 数据降维与分类

6.1 聚类分析：数据的自动分类

聚类分析是一种无监督学习方法，根据数据的相似性将样本或变量自动分组，使组内数据相似度高，组间数据相似度低，常用于客户细分、市场定位、数据分类等场景。

6.1.1 聚类分析的类型

- 样本聚类（Q 型聚类）：对案例（样本）进行聚类（如对客户进行细分）；
- 变量聚类（R 型聚类）：对变量进行聚类（如将多个相关变量合并为几个综合变量）。

6.1.2 常用聚类方法

- 系统聚类法（Hierarchical Clustering）：逐步合并或拆分聚类，形成聚类树，适用于样本量较小（ < 100 ）的情况；
- K - 均值聚类法（K-Means Clustering）：预先指定聚类数量 k，将样本分配到 k 个聚类中，适用于样本量较大（ > 100 ）的情况。

6.1.3 系统聚类法操作与解读（以“客户细分”为例）

1. 打开数据文件，选择客户的特征变量（如年龄、月收入、消费频率、消费金额，均为度量尺度）；

2. 数据标准化：点击“分析”→“分类”→“系统聚类”，弹出对话框；
3. 将特征变量移入“变量”框，选择“个案”（样本聚类）；
4. 点击“方法”，选择聚类方法（如“组间平均链锁法”），选择标准化方法（如“Z分数”，消除变量量纲影响），点击“继续”；
5. 点击“绘制”，勾选“树状图”，点击“继续”；
6. 点击“确定”，运行分析。

结果解读：

- 树状图（Dendrogram）：展示样本的聚类过程，横轴为样本编号，纵轴为聚类距离，根据树状图可手动确定聚类数量（如从树状图中可看出分为3类客户）；
- 聚类结果：根据确定的聚类数量，将样本分为不同组别，可通过“分析”→“分类”→“系统聚类”→“保存”，勾选“聚类成员”，保存聚类结果到数据文件中，进一步分析各组的特征（如第一组为年轻低收入客户，第二组为中年中等收入客户，第三组为老年高收入客户）。

6.1.4 K - 均值聚类法操作与解读

1. 打开数据文件，选择特征变量（需标准化）；
2. 点击“分析”→“分类”→“K - 均值聚类”，弹出对话框；
3. 将特征变量移入“变量”框，输入聚类数量 k（如 k=3）；
4. 点击“迭代与分类”，设置最大迭代次数（默认 20）和收敛准则（默认 0.02），点击“继续”；
5. 点击“保存”，勾选“聚类成员”“与聚类中心的距离”，点击“继续”；
6. 点击“确定”，运行分析。

结果解读：

- 聚类中心：展示每个聚类的中心值（标准化后的均值），可通过中心值判断各聚类的特征；
- 聚类成员：保存每个样本的聚类归属，可统计各聚类的样本数量和比例；
- 与聚类中心的距离：每个样本到所属聚类中心的距离，距离越小，样本与聚类中心的相似度越高；
- 方差分析：若 P05，说明各聚类在所选特征变量上存在显著差异，聚类效果较好。

6.2 因子分析：数据降维与潜在变量提取

因子分析是一种数据降维方法，将多个相关的观测变量提取为少数几个不相关的潜在变量（因子），这些因子能反映原始变量的大部分信息，常用于简化数据结构、构建量表维度等场景。

6.2.1 因子分析的前提条件

- 相关性：原始变量间需存在一定的相关性（若变量间无相关性，无法提取共同因子）；
- KMO 检验：KMO 值 > 0.6 ，说明适合进行因子分析；KMO 值 < 0.6 ，不适合进行因子分析；
- 巴特利特球形度检验：若 $P < 0.05$ ，说明变量间的相关矩阵非单位矩阵，适合进行因子分析。

6.2.2 操作步骤（以“客户满意度量表数据降维”为例，量表包含 10 个观测变量）

1. 打开数据文件，10 个满意度观测变量均为度量尺度；
2. 点击“分析”→“降维”→“因子分析”，弹出对话框；
3. 将 10 个观测变量移入“变量”框；
4. 点击“描述统计”，勾选“KMO 和巴特利特球形度检验”“相关矩阵”，点击“继续”；
5. 点击“提取”，选择提取方法（如“主成分分析法”），设置因子提取标准（如“特征值 > 1 ”），点击“继续”；
6. 点击“旋转”，选择旋转方法（如“最大方差旋转”，使因子载荷矩阵更易解释），点击“继续”；
7. 点击“得分”，勾选“保存为变量”，选择得分方法（如“回归法”），点击“继续”；
8. 点击“确定”，运行分析。

6.2.3 结果解读

1. 前提条件检验：KMO 值 > 0.6 ，巴特利特球形度检验 $P < 0.05$ ，适合进行因子分析；
2. 因子提取结果：
 - 总方差解释表：展示各因子的特征值、方差贡献率和累计方差贡献率，累计方差贡献率 $> 60\%$ ，说明提取的因子能解释大部分原始变量信息（如提取 3 个因子，累计方差贡献率 = 75% ）；
 - 因子载荷矩阵（旋转后）：因子载荷表示观测变量与因子的相关性，绝对值 > 0.6 的观测变量可归为该因子（如变量 1、2、3 的载荷在因子 1 上 > 0.6 ，变量 4、5、6 的载荷在因子 2 上 > 0.6 ，变量 7、8、9、10 的载荷在因子 3 上 > 0.6 ）；
1. 因子命名：根据各因子包含的观测变量含义，给因子命名（如因子 1 命名为“产品质量满意度”，因子 2 命名为“服务态度满意度”，因子 3 命名为“价格满意度”）；
2. 因子得分：SPSS 会生成每个样本的因子得分（如 FAC1_1、FAC2_1、FAC3_1），可用于后续分析（如聚类分析、回归分析）。

第 7 章：数据可视化 —— 让数据更直观

数据可视化是将数据通过图表形式呈现，帮助用户快速发现数据规律、传递信息，SPSS 提供了丰富的可视化工具，包括条形图、折线图、饼图、直方图、散点图、雷达图等。

7.1 基础图表制作

7.1.1 条形图：展示分类变量的分布或比较

适用场景：比较不同类别的数据大小（如不同学历的人数、不同地区的销售额）。

操作步骤：

1. 点击“图形”→“旧对话框”→“条形图”，选择“简单条形图”“分组条形图”或“堆积条形图”；
2. 选择“个案组摘要”或“变量摘要”，点击“定义”；
3. 将分类变量移入“X 轴”，数值变量移入“Y 轴”，点击“确定”；
4. 双击图表进入编辑模式，可修改图表标题、坐标轴标签、颜色、字体等格式。

7.1.2 折线图：展示数据的变化趋势

适用场景：展示数据随时间或其他连续变量的变化趋势（如月度销售额变化、学生成绩随学习时间的变化）。

操作步骤：

1. 点击“图形”→“旧对话框”→“折线图”，选择“简单折线图”或“多线折线图”；
2. 选择“个案组摘要”，点击“定义”；
3. 将时间变量或连续变量移入“X 轴”，数值变量移入“Y 轴”，点击“确定”；
4. 编辑图表：添加趋势线、修改坐标轴刻度、调整线条样式等。

7.1.3 饼图：展示分类变量的占比

适用场景：展示各类别占总体的比例（如不同产品的市场份额、不同年龄段的人口占比）。

操作步骤：

1. 点击“图形”→“旧对话框”→“饼图”，选择“个案组摘要”；
2. 点击“定义”，将分类变量移入“定义饼图的类别”，将数值变量移入“切片的大小”（如人数、金额）；
3. 勾选“显示百分比”，点击“确定”；
4. 编辑图表：修改饼图颜色、添加数据标签、调整百分比显示格式等。

7.1.4 直方图：展示数值变量的分布形态

适用场景：展示数值变量的分布情况（如收入分布、成绩分布），判断是否符合正态分布。

操作步骤：

1. 点击“图形”→“旧对话框”→“直方图”，将数值变量移入“变量”框；
2. 勾选“显示正态曲线”，点击“确定”；
3. 编辑图表：调整组距、修改坐标轴标签、添加标题等。

7.2 高级图表制作

7.2.1 散点图：展示变量间的关系

适用场景：分析两个或多个数值变量间的相关性（如身高与体重的关系、学习与成绩的关系）。

操作步骤：

1. 点击“图形”→“旧对话框”→“散点 / 点图”，选择“简单散点图”“矩阵散点图”或“分组散点图”；
2. 点击“定义”，将自变量移入“X 轴”，因变量移入“Y 轴”（矩阵散点图可同时展示多个变量间的关系）；
3. 点击“确定”，双击图表进入编辑模式，可添加拟合线（线性、二次函数等），直观展示变量间的关系。

7.2.2 雷达图：展示多变量的综合特征

适用场景：比较多个样本在多个变量上的表现（如不同产品在质量、价格、服务、外观等维度的评分比较）。

操作步骤：

1. 点击“图形”→“旧对话框”→“雷达图”，选择“简单雷达图”或“分组雷达图”；
2. 点击“定义”，将多个评价变量移入“雷达轴”，将样本分组变量移入“分组变量”（如产品类型）；
3. 点击“确定”，编辑图表：调整雷达轴刻度、修改线条颜色、添加图例等。

7.2.3 箱线图：展示数据的离散程度与异常值

适用场景：比较多组数据的离散程度和异常值（如不同班级学生的成绩分布比较）。

操作步骤：

1. 点击“图形”→“旧对话框”→“箱线图”，选择“简单箱线图”或“分组箱线图”；
2. 点击“定义”，将数值变量移入“Y 轴”，将分组变量移入“X 轴”；
3. 点击“确定”，编辑图表：修改箱体颜色、添加异常值标签、调整坐标轴格式等。

7.3 图表导出与美化

7.3.1 图表导出

1. 选中 SPSS 输出窗口中的图表，右键点击“复制”；
2. 打开 Word、Excel 或 PowerPoint，右键点击“粘贴”，即可将图表导出到其他软件；
3. 若需导出为图片格式（如 PNG、JPG），右键点击图表，选择“导出”，设置保存路径和图片格式，点击“确定”。

7.3.2 图表美化技巧

1. 标题与标签：使用简洁明了的标题，坐标轴标签注明单位（如“月收入（元）”）；
2. 颜色搭配：选择专业、协调的颜色（如商务汇报使用蓝色、灰色系，学术论文使用黑白或低饱和度颜色）；
3. 字体与字号：统一字体（如宋体、Arial），标题字号大于坐标轴标签字号，确保可读性；
4. 去除冗余元素：删除不必要的网格线、图例（若图表简单易懂），使图表更简洁；
5. 数据标签：关键数据可添加标签（如百分比、具体数值），但避免标签过多导致混乱。

第四部分：实战案例与常见问题解答

第 8 章：SPSS 实战案例分析

8.1 案例一：学生成绩分析（描述性统计 + 推论性统计）

案例背景

某中学随机抽取 50 名高一学生，收集其数学、语文、英语三门课程的成绩，以及性别、是否参加课外辅导班等信息，需分析：

1. 三门课程成绩的基本特征（集中趋势、离散程度、分布形态）；
2. 男女学生的数学成绩是否存在显著差异；
3. 参加与未参加课外辅导班的学生，英语成绩是否存在显著差异。

分析步骤与结果

1. 数据录入与变量定义：
 - 变量包括：性别（1 = 男，2 = 女，名义）、是否参加辅导班（1 = 是，0 = 否，名义

（注：文档部分内容可能由 AI 生成）