

项目 1：大数据概述

实训目标

通过班级实地微型调研、表格分类梳理等实操，独立区分三类数据、拆解大数据 4V 特征，依托真实调研数据完成现象解读，全程动手完成数据整理、统计与分析实操，锻炼数据实操与客观研判能力。

实训环境

Windows10 及以上、Office/WPS 表格、浏览器。

操作步骤

- 1.查询价格小于 5 元的商品 ID、商品编号、商品名称、商品种类名称、单价和库存量，并把查询结果保存到新表 product_bak 中（使用 CREATE TABLE...SELECT 语句）。
- 2.查询价格大于 20 元的商品记录，将查询结果集插入到 product_bak 表中。
- 3.将商品种类为“饮料”的商品单价统一下调 3%。
- 4.将“S02”销售员的订单及订单明细全部删除。

1、步骤 1：开展三项即时调查

依次提出以下三个问题，由参与者通过举手或其他明确方式作出“是/否”回应：

- （1）今天是否食用了早餐？
- （2）此刻是否感觉精神状态良好？
- （3）是否曾使用手机或电脑查询学习或生活相关信息？

每项问题结束后，统计作出肯定回应的人数。

2、步骤 2：构建简易数据记录表

在书写板面上绘制如下表格，并填入统计结果，假设结果如表 1-1 所示：

表 1-1 即时调查统计结果

调查内容	肯定回应人数	总参与人数	占比（估算）
今天是否食用了早餐？	34	40	约 85%
此刻是否感觉精神状态良好？	19	40	约 48%
今天是否进行了超过 30 分钟的线上学习？	38	40	约 95%

注：占比简单估算即可，重在体现数量关系，不要求精确计算。

3、步骤 3：解读数据所反映的现象

围绕表格内容，思考以下问题：

- （1）高比例的早餐摄入是否反映出良好的生活习惯？
- （2）精神状态良好的比例较低，可能与作息、课业负担或环境因素有关吗？
- （3）绝大多数参与者使用数字设备获取信息，说明数字技术已深度融入日常学习与生活吗？通过分析可知：看似简单的数据，能够客观呈现群体特征，并为改进决策提供依据。

4、步骤 4：延伸思考与原则反思

进一步探讨：

- （1）若连续多日记录同类信息，是否可能发现规律？例如早餐与上午专注度的关系。
- （2）为何这些数据可以公开讨论？——因为问题设计遵循了非敏感、自愿、匿名、有用的基本原则。

由此认识到：有价值的数据获取，应以尊重个体、保障安全为前提。

注意事项

- 1、区分三类数据核心依据：是否为固定二维表格，JSON 属于半结构化，音视频、纯文本为非结构化；
- 2、4V 中 Value 为价值密度低、潜在价值高，不可混淆；
- 3、调研全程匿名，严禁采集姓名、手机号等隐私信息；
- 4、统计、分析时不得篡改原始调研数字，保证实操数据真实。