

## 频数频率统计

### 一、核心思想

频数频率统计：通过计算数据出现的次数（频数）和占比（频率），挖掘数据背后的规律或热点。

文旅场景：统计游客行为（如景点打卡次数）、关键词出现频率（如“亲子游”在搜索中占比）、消费偏好（如门票收入占比）等。

技术本质：将杂乱的数据转化为“计数游戏”，用数字回答“什么最热门”“哪里有问题”。

### 二、核心步骤（以景区游客行为分析为例）

#### 数据清洗

去噪：删除无效数据（如测试账号的异常操作）。

结构化：将原始日志（如“2025-03-25 游客 A 打卡黄山”）转化为表格（日期、游客 ID、景点）。

#### 统计计算

频数：景点“黄山”被打卡 1000 次，“九华山”被打卡 800 次。

频率：“黄山”打卡次数占比  $1000 / (1000 + 800) \approx 55.6\%$ 。

#### 结果呈现

可视化：柱状图（景点打卡次数对比）、饼图（消费项目占比）、热力图（区域游客密度）。

### 三、常用算法与文旅适配

算法名称	核心逻辑	文旅应用示例
简单计数	直接统计单个指标出现次数	统计“西湖”在游记中被提及次数（频数），或占所有景点提及的比例（频率）
分组统计	按维度分组后统计（如时间、地域）	按季度统计“三亚”游客量（Q1: 50 万, Q2: 30 万），识别淡季 / 旺季规律
频率矩阵	多维度交叉统计（如性别 × 消费项目）	女性游客中购买“温泉票”的比例（30%）高于男性（15%），指导精准营销

算法名称	核心逻辑	文旅应用示例
滑动窗口计数	动态统计时间窗口内的高 频事件	国庆假期每小时统计景区拥堵指数（实时调整限流措 施）

## 四、文旅游业典型应用

### 游客画像分析

输入：游客消费记录（餐饮、住宿、门票）

输出：高频消费项目（如“温泉票”占比40%）→开发“温泉+住宿”套餐。

价值：提升客单价，优化产品组合。

### 舆情热点监测

输入：社交媒体提及的景点名称

输出：“阿那亚”频数突然上升→快速策划宣传活动（如“网红打卡地攻略”）。

价值：抓住流量红利，提高景点曝光度。

### 服务效率优化

输入：景区投诉内容（如“卫生间排队”“指示牌少”）

输出：“卫生间”投诉占比达60%→增加临时厕所或优化导览图。

价值：针对性解决痛点，提升游客满意度。

## 五、技术挑战与对策

### 挑战：

数据稀疏性：小众景点（如“查济古镇”）因游客少导致统计结果偏差。

实时性要求：节假日需分钟级更新统计结果，传统方法难以支撑。

多源数据整合：线上评论、线下闸机数据、APP行为日志格式不统一。

### 对策：

平滑处理：对低频数据使用拉普拉斯平滑（如“查济古镇”默认基础权重）。

分布式计算：用Spark/Flink实时处理数据（如每秒处理10万条游客行为记录）。

数据标准化：建立统一标签体系（如“投诉类型”分为“设施”“服务”“体验”）。

## 六、行业实践案例

某文旅局通过频数频率统计优化资源配置：

景区热度排名：统计各景点打卡次数 → 对“冷门但优质”景点（如“呈坎八卦村”）增加宣传预算。

游客来源分析：按城市统计游客占比 → 发现“上海”游客占比 35%，针对性开通上海直达高铁。

消费结构优化：统计餐饮消费频率 → 发现“徽菜”占比仅 20%，鼓励景区内徽菜馆推出特色套餐。

总结：频数频率统计是文旅大数据的“基础温度计”，通过简单的计数和占比分析，能快速定位热点（如网红景点）、识别短板（如服务投诉集中点）、验证假设（如女性游客更爱温泉）。在文旅领域，它不仅能辅助决策（如资源分配），还能为复杂算法（如推荐系统、预测模型）提供数据基础。未来，结合 AI 的智能统计（如自动识别异常波动原因）将进一步提升文旅运营的精细化水平。