



第04章数据规整

顾立平

要点



注重视觉效果与数据清晰度平衡，
避免过度立体导致信息传达混乱。



立体饼图设计原则

在同一直方图展示两组数据，通
过高度或面积对比，凸显数据差
异与相似性，促进深入理解。

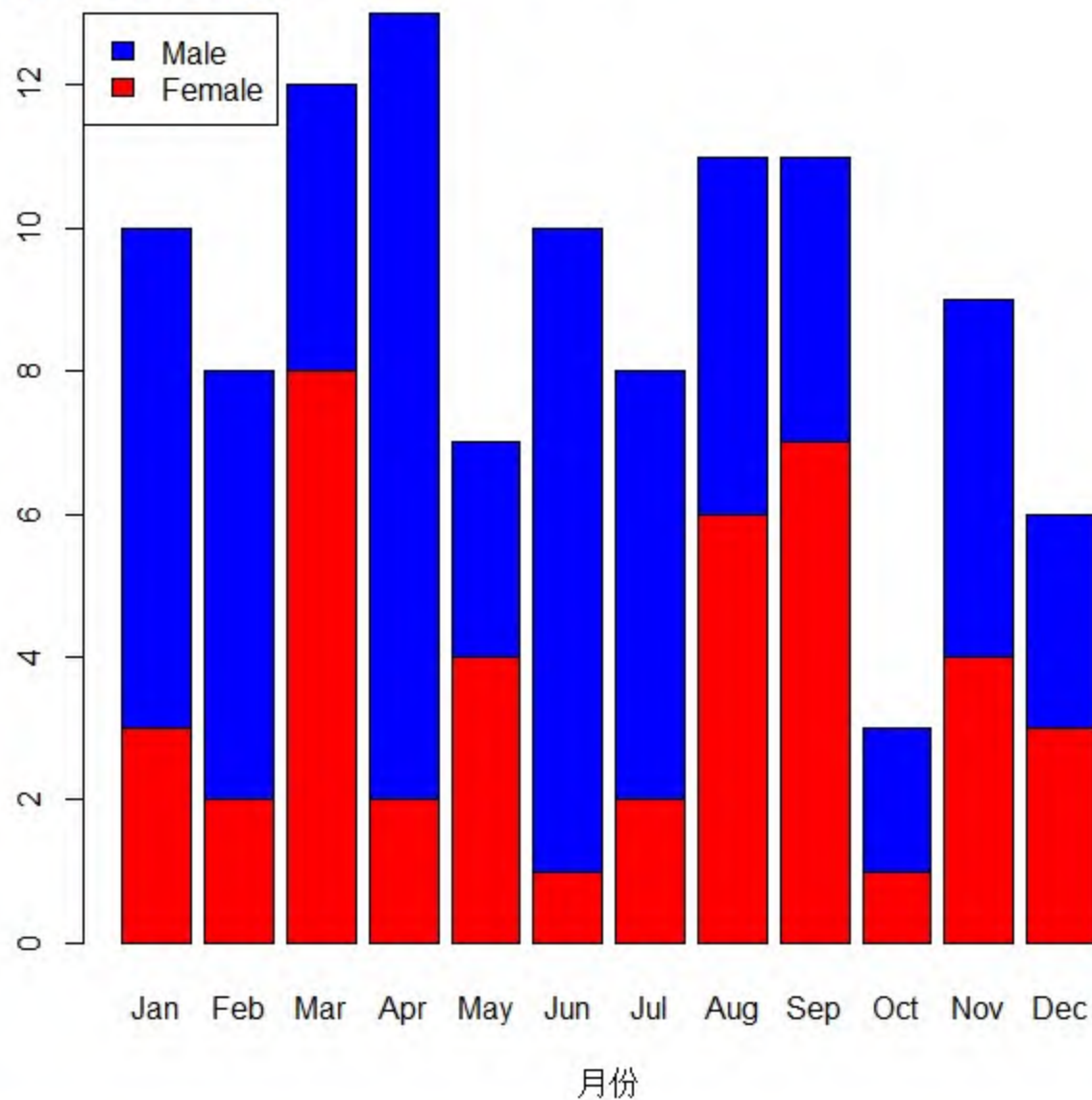


直方图比较分析



030 直方图上的两组比较

出生月份样本



科学领域的应用



生物学中的直方图应用

对比基因型对生理指标的影响，如野生型与突变型植物的光合作用效率，揭示突变的生理效应。

生态学中的物种多样性分析

通过直方图比较不同生态系统中物种多样性的分布，为生态保护提供科学依据。

医学研究中的临床分析

展示不同治疗组的生理指标分布，如血糖或血压，以评估治疗效果和安全性。

流行病学的疾病分布比较

使用直方图比较疾病在不同人群中的发病率，揭示疾病流行规律和影响因素。



科学领域的 应用



01

物理化学实验
可视化

在物理实验中比较不同条件下结果的分布，化学分析中评估样品的均一性和稳定性。



02

商业领域的市
场分析

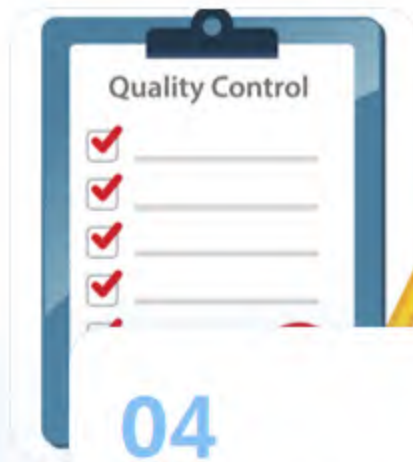
比较不同产品或服务的市场销量或份额，帮助企业了解市场需求和竞争态势。



03

财务管理中的
企业绩效评估

通过直方图比较部门或项目的成本、收入和利润分布，评估企业财务状况和绩效。



04

质量控制中的
生产质量监控

对比不同批次产品的质量指标分布，识别生产过程中的问题，选择优质供应商。

要点

直方图比较与箱线图展示

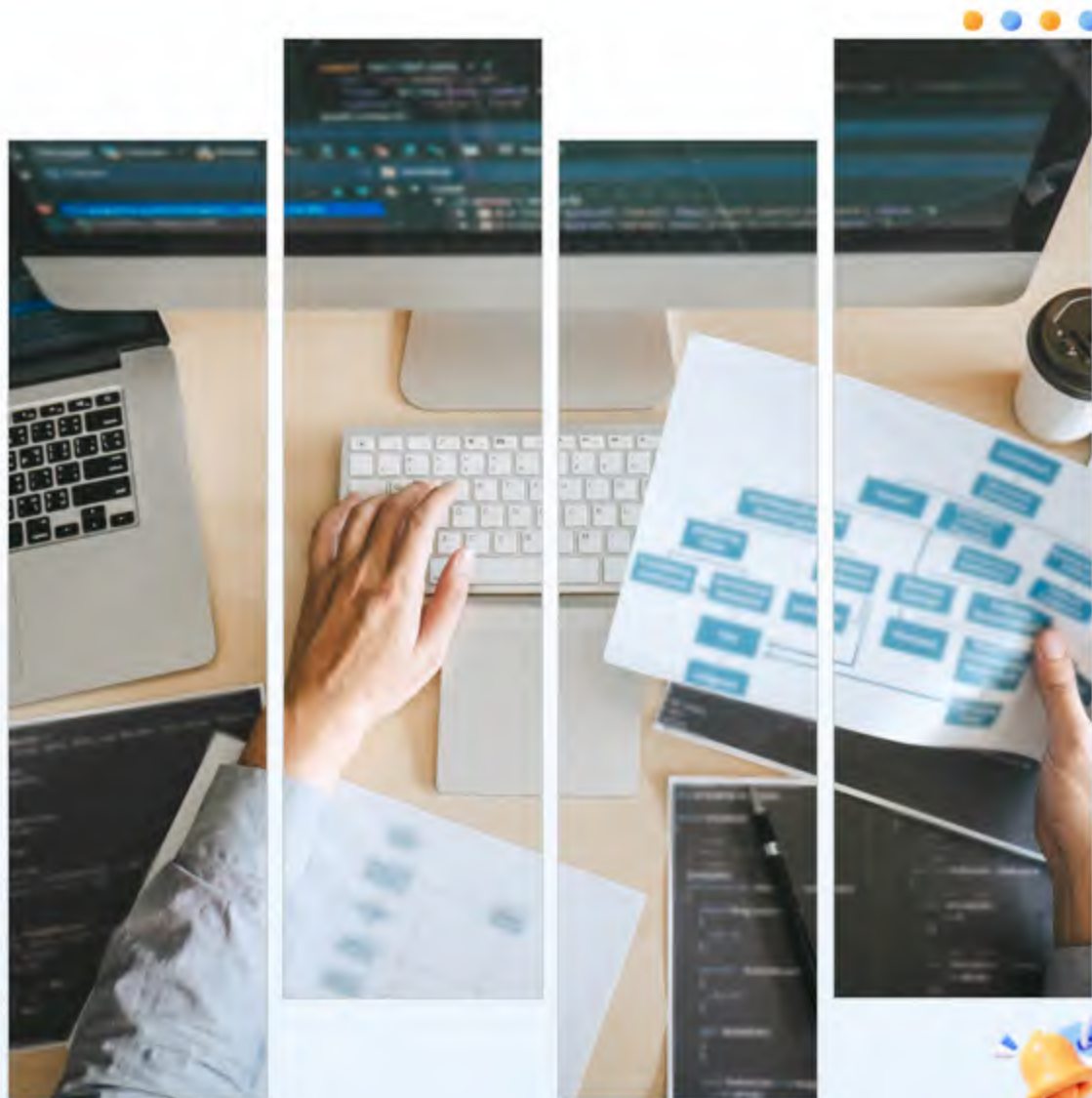
科学与商业领域，用于数据分布比较和详细描述，揭示数据背后模式，支持决策分析。

栈与循环应用

栈在编译、语法分析和数据库事务管理中关键，循环用于数值模拟、数据分析和用户界面交互，两者分别解决不同问题。

分布描述与解释

结合箱线图提供数据分布的直观展示，通过文字或数值描述增强理解，应用于医学研究、市场分析和质量控制等场景。



031 栈 (Loops)



```
R Console
> # 创建一个字符向量Month, 包含了一年中的12个月份
> Month=c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "Dec")
>
> # 初始化一个长度为12的数值型向量FBMCOUNTS, 用于存储每个月份在FBM中出现的次数
> FBMCOUNTS=vector(mode="numeric",length=12)
>
> # 使用for循环遍历1到12的整数, 对应一年中的12个月份
> for (i in c(1:12)){
+   # 使用which函数找出FBM中等于当前月份 (Month[i]) 的元素的索引
+   # length函数计算这些索引的数量, 即当前月份在FBM中出现的次数
+   # 将这个次数存储在FBMCOUNTS向量的第i个位置
+   FBMCOUNTS[i]=length(which(FBM==Month[i]))
+ }
>
> # 初始化一个长度为12的数值型向量MBMCOUNTS, 用于存储每个月份 (以数字表示) 在MBM中出现的次数
> MBMCOUNTS = vector(mode="numeric", length=12)
>
> # 使用for循环遍历1到12的整数, 对应一年中的12个月份 (在这里, 月份是以数字1到12表示)
> for (i in c(1:12)){
+   # 使用which函数找出MBM中等于当前月份数字i的元素的索引
+   # length函数计算这些索引的数量, 即当前月份在MBM中出现的次数
+   # 将这个次数存储在MBMCOUNTS向量的第i个位置
+   MBMCOUNTS[i]=length(which(MBM==i))
+ }
>
```



总结



箱线图与分布描述

科学商业领域常用，展示数据分布，结合文字数值解释，支持决策。



直方图叠加分布曲线

直方图基础上加曲线，直观比较实际与理论分布，揭示数据特征和拟合度。



要点



图形展示数据分布频率，矩形条高度对应间隔内频数，宽度表示间隔。

直方图基本概念

生物学至商业领域，直方图分析基因影响、物种多样性和市场趋势，助力实验可靠性、疾病研究及商业决策。

直方图应用领域

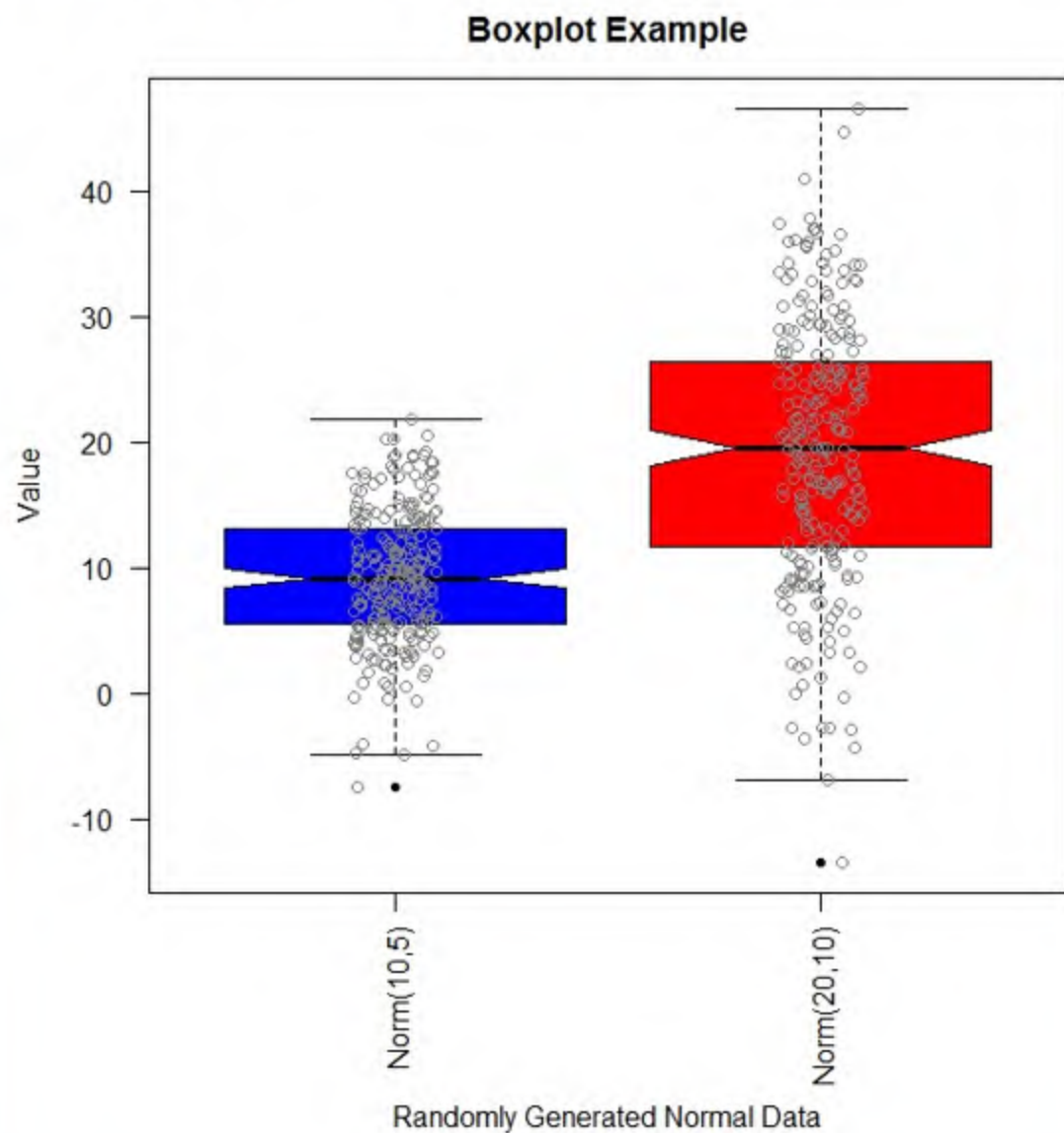


叠加分布曲线意义

比较实际与理论分布，理论曲线如正态分布，便于评估两者吻合度。



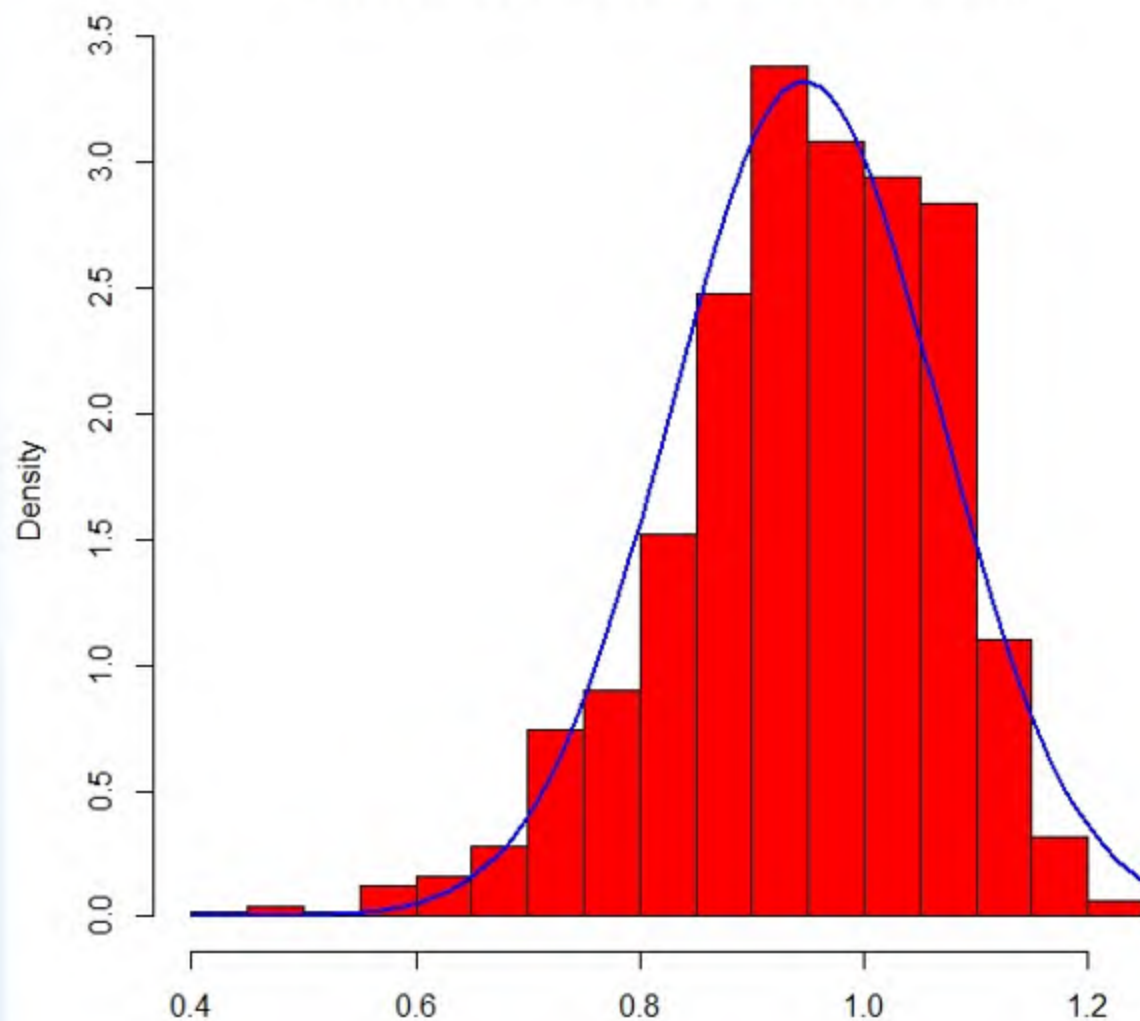
032 箱线图加上分布描述



033 直方图 (Histogram) 上叠加分布曲线



1000 Random Weibull Distribution Values





谢谢

gulp@mail.las.ac.cn