

# 数理统计上机课第一次作业

2021 年 4 月 6 日

## 1 第一题：数据处理与可视化

数据为数据集 `crawfish.xlsx`，记录了七座城市共计 2527 家经营小龙虾的餐馆的相关数据。请使用你认为合适的方法完成：

- (1) 初步汇总数据，并简要描述小龙虾销售的基本情况，例如价格、销量情况。
- (2) 删除数据中没有提供人均价格 (`price`) 的店铺，并增加新的一列总销售额 (`total`) 为销量 (`sale`) 与人均价格 (`price`) 的乘积。
- (3) 使用你认为合适的统计图表展示七座城市的总销售额 (`total`) 分布情况，解读你绘制的图表并推断可能的原因。
- (4) 分别绘制合适的图表展现分城市有无 `wifi` (`wifi`: 有 =1)、有无车位 (`park`: 有 =1) 以及不同折扣力度 (`distype`) 的总销售额 (`total`) 的平均值，判断这些因素的不同对总销售额 (`total`) 的平均值是否存在影响？陈述你的观点并适当分析原因。(提示：分组绘制图形时可以参考 `ggplot()` 函数中 `stat_summary()` 函数，注意 `ggplot()` 函数括号内 `aes()` 中参数 `fill` 的设置。)
- (5) 王多鱼是一位投资人，他的资金规模相对不大，近期希望投资小龙虾餐饮业。根据以上的分析，你推荐他投资哪一座城市的小龙虾餐饮业？简单陈述理由。

## 2 第二题：循环与函数

完成以下实验：

- (1) 针对特定函数  $y = \cos(\sin(x))$ ，使用循环编写梯度下降法的程序寻找函数的局部极小值，要求：当迭代值  $|x^{(n+1)} - x^{(n)}| < 10^{-4}$  时迭代停止，自己选择一个合适的迭代起始点，并且选择一个较为合适的学习率。(提示：这个函数的导函数可以直接求出零点，可以结合函数图像选择合适的起始点以及学习率，注意避开局部极小值本身)
- (2) 基于 (1) 中的循环，考虑循环中哪些变量条件可以事先给定，并编写梯度下降法的函数，封装为独立函数后自己选择一个具有多个极小值点的函数 (如  $\sin x$ )，绘制函数的图像，并比较不同的迭代起始位置会对迭代过程产生什么影响。
- (3) 经过以上实验，你觉得梯度下降法在实际使用中应该注意什么？

## 作业提交

作业请使用 Rmarkdown 完成，完成后请输出为 html 格式，代码可以参考 GitHub 主页 (<https://github.com/Huang-qy-Chi>)Class3 中的文件：dplyr 包相关函数的练习.Rmd，电脑配置导致的输出问题请在作业开篇位置说明。**4 月 15 日 0 时**之前请将 html 版文件发送到邮箱 **sltj2021zy@163.com**，如确有特殊情况需要推迟请提前说明。