

Chapter 5 Programming Assignments

王迦楠 数学科学学院 求数 2101 3210105175

2024 年 2 月 3 日

1 Problem 1

模型概述

1. 列出所有可能的影响”momentum”的因素
2. 进行数据清洗和处理
3. 共线性分析，去除部分重复的因素
4. 使用层次分析法进行分析
5. 计算”momentum”

为了探究”momentum”的原因，我们使用层次分析法进行定量的分析。具体地，”momentum”的大小定义为

$$f_{ij} = \omega \cdot \mathbf{X}_{ijk}$$

其中：

1. f_{ijk} 代表第 i 场比赛（按照表格顺序）中第 j 个 point number 之前的”momentum”
2. \mathbf{X}_{ij} 是一个 n 维的列向量，代表相对应时刻的一些影响因素，后面我们会具体给出
3. ω^n 是一个 n 维的行向量，指影响因素具体的权重，我们会通过层次分析法来得到
4. 这里的公式中，具体会分为两个不同的计算方法，一个分别代表自己发球的场次和对手发球的场次，我们可以写为

$$\omega = \omega_0 \cdot \delta = (\omega^{(0)}\delta^{(0)}, \omega^{(1)}\delta^{(1)}, \dots, \omega^{(n)}\delta^{(n)})$$

代表两个相同维数向量对应位置相乘构成的向量，其中 δ 为一个 0,1 向量，取决于是否是自己的发球轮次，在具体的计算中，我们会分两种情况带入具体考虑。

1.1 影响因素分析

对于 \mathbf{X}_{ij}^a 的具体定义，我们认为，除了是否为发球方，还有许多其他因素会造成影响，包括选手的技术，疲惫程度，以及比赛的实时心态。基于以上三个大点，我们整理了 12 个因素来作为初步的影响因素，具体如下：（画个好看的表）

1.2 数据清洗和处理，以及数据集的标准化

1.3 共线性检验

在对数据进行处理之后，考虑到同一类别的影响因素之间可能存在共线性，比如发球 ace，一发得分率，前一分是否得分可能会相关，跑动距离，击拍数可能会相关，我们使用 stata 对数据进行共线性检测。检测结果表面跑动距离与击拍数之间方差膨胀因子较大，因此我们剔除二者之一的击拍数，使用剩下 11 个数据进行层次分析法。

1.4 层次分析法

我们为两种不同的发球类型（自己发球、对方发球）建立了两个矩阵；每个元素都显示了因子 i 和因子 j 之间的偏好程度。

我们通过计算最大特征值并对其对应的特征向量进行归一化来获得每个部分的权重。我们使用组成比 (CR) 来检验每个矩阵的一致性，给定的数据为:…其中，三个矩阵的组成比分别为 0.0612、0.0223 和 0.0301。因为它们都小于 0.1，所以证实了矩阵的一致性。(需要看看)

We obtain the weight of each section by calculating the maximum eigenvalue and normalizing its corresponding eigenvector. We test the consistency of each matrix using the Consistency Ratio (CR), given $CR = \frac{CI}{RI}$ where $CI = \frac{\lambda_{max} - n}{n - 1}$ and $RI = 0.90$ ($n = 4$). The CRs of the three matrices are 0.0612, 0.0223 and 0.0301, respectively. Because they are all lesser than 0.1, the consistency of the matrices is confirmed. Therefore, the weight of the four sections in our model is shown in Figure 7.

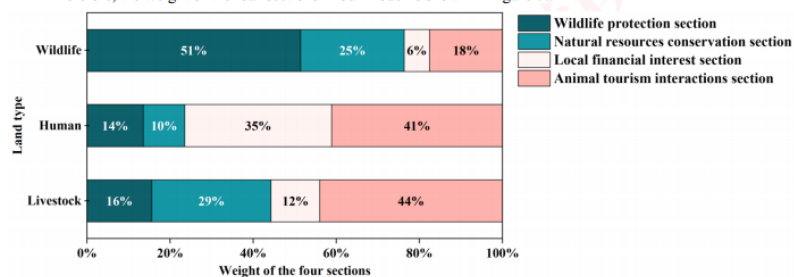


Figure 7. The weight of the four sections in different land types

Our final score is defined as

$$Final\ score = \omega_1 Score1 + \omega_2 Score2 + \omega_3 Score3 + \omega_4 Score4$$

Where ω_1 , ω_2 , ω_3 and ω_4 represent the weight of the four scores, which can be derived according to different land types listed in Figure 7.

做一个这样的图，参考 23b 题 2300136