

# 从零开始学数学建模

## 熵权法

主讲人：北海

b站/公众号：数学建模BOOM

# 熵权法

- 模型简介
- 适用赛题
- 典型例题与原理讲解
- 代码求解

## □ 模型简介

## □ 从考试成绩说起

- 张三、李四和王五的考试成绩如下表所示
- 观察这份成绩表，有哪门科目**最容易拉开差距**？哪门课又**最不容易拉开差距**？

考试科目	张三	李四	王五
语文	88	92	79
数学	61	100	83
英语	90	82	95
物理	97	92	95
化学	100	100	100
生物	81	92	89
总分	517	558	541

微信公众号：数学建模BOOM

## □模型简介

## □熵的概念

- 信息论中，熵是对**不确定性**的一种度量，可判断一个事件的**随机性**及**无序**程度
- 用熵值判断某个指标的**离散程度**，指标的**离散程度越大**，该指标对**综合评价**的影响越大

## □直观看熵：考试成绩

- 某一门课“不易拉开差距”，意味着这一门课对最终总成绩排名影响很小
- 所有人的化学成绩相同，意味着“整齐划一”，相应混乱程度就低

## □熵的现实意义

- 本模型中，混乱程度低对应着**熵值接近1**，评价总成绩时可给该指标赋予低权重
- **即使这门课的成绩不算进总成绩（权重为0），总成绩的排名也是不会变的**
- 相对的，数学的分数差距较大，如果去掉这门课，总成绩排名可能出现变化

考试科目	张三	李四	王五
数学	61	100	83
化学	100	100	100

# 熵权法

- 模型简介
- 适用赛题
- 典型例题与原理讲解
- 代码求解

## □适用赛题

### □ 数据全面、缺少文献或主观依据的题目

- 例如**评价河流水质**，已知河流的含氧量、PH值、细菌密度、生物密度等数据
- 但**缺乏评价水质的文献资料**，或者文献内的**说法不一**
- 即文献很难帮助我们确定影响水质最重要的因素是哪一个
- 也很难告诉我们其余指标的重要程度如何衡量
- 此时即可使用熵权法，**根据数据本身建立评价体系**

### □ 注意事项

- 熵权法与其他方法（层次分析法、TOPSIS法等）最大的区别就是**完全客观**
- 追求“**公平公正**”的情况，可优先考虑熵权法
- 但有时“完全客观”也是缺点，**难以将数据之外的因素考虑进去**

# 熵权法

- 模型简介
- 适用赛题
- 典型例题与原理讲解
- 代码求解

## □ 典型例题

□ 根据下表给出的10个学生8门课的成绩，给出这10个学生评奖学金的评分排序

学生编号	语文	数学	物理	化学	英语	政治	生物	历史
1	93	66	86	88	77	71	90	94
2	97	99	61	61	75	87	70	70
3	65	99	94	71	91	86	80	93
4	97	79	98	61	92	66	88	69
5	85	92	87	63	67	64	96	98
6	63	65	91	93	80	80	99	74
7	71	77	90	88	78	99	82	68
8	82	97	76	73	86	73	65	70
9	99	92	86	98	89	83	66	85
10	99	99	67	61	90	69	70	79

建模BOOM



## □数据标准化

### □ 标准化的原因

- 评价体系中，存在数值越大越好的**正向指标**，和数值越小越好的**负向指标**
- 不同指标数量级也可能不同；且求熵的公式中用到对数函数，变量不允许有负值

### □ 变量定义

- 设 $x_{ij}$ 为第 $i$ 个学生的第 $j$ 门课程的成绩；本题中所有指标（各科成绩）都是正向指标

### □ 正向指标标准化

$$a_{ij} = \frac{x_{ij} - \min\{x_{1j}, \dots, x_{nj}\}}{\max\{x_{1j}, \dots, x_{nj}\} - \min\{x_{1j}, \dots, x_{nj}\}}$$

### □ 负向指标标准化

$$a_{ij} = \frac{\max\{x_{1j}, \dots, x_{nj}\} - x_{ij}}{\max\{x_{1j}, \dots, x_{nj}\} - \min\{x_{1j}, \dots, x_{nj}\}}$$

- 标准化之后， $a_{ij}$ 所有值在 $[0,1]$ 区间内，且都是数值越大、现实意义越好

微信公众号：数学建模BOOM

# 熵权法

## □ 指标的熵值和变异程度

### □ 每个评价对象在各个指标中的比重

- 可理解为统计意义上某种情况出现的概率

$$p_{ij} = \frac{a_{ij}}{\sum_{i=1}^n a_{ij}}$$

### □ 熵值

- 对于第 $j$ 个指标，其熵值 $e_j$ 为：

$$e_j = -\frac{1}{\ln n} \sum_{i=1}^n p_{ij} \ln p_{ij}$$

### □ 变异系数

- 第 $j$ 个指标的变异系数： $g_j = 1 - e_j$
- 显然熵值越大、变异系数越小，代表该指标越有序，该指标的信息量也就越小

微信公众号：数学建模BOOM

## □ 熵值和变异程度

## □ 极端例子

- 假设10个学生每个人的语文成绩都是100
- 如果不经标准化，直接根据前文公式，计算  $p_{i1} = \frac{a_{ij}}{\sum_{i=1}^n a_{ij}} = 1/10$
- 第1个指标(语文成绩)的熵值:  $e_1 = -\frac{1}{\ln n} \sum_{i=1}^n p_{i1} \ln p_{i1} = 1$
- 该指标的变异系数:  $g_j = 1 - e_j = 0$

## □ 意义分析

- 该指标的变异系数为0，代表所有数据都一样，其信息量为0
- 那么该指标毫无用途，在评价体系中去掉该指标，对评价结果不会有影响
- 因此，数据差异越小、熵值越大、变异系数越小、信息量越小
- 极端情况就是数据全相等、无差异，最终求得的变异系数为0

## □ 权重与评分

### □ 变异系数求权重

- 计算第 $j$ 个指标的权重：

$$w_j = \frac{g_j}{\sum_{j=1}^m g_j}$$

- 指标的变异系数越大、信息量越大，相应指标的权重也越大

### □ 综合评分

- 计算第 $i$ 个评价对象的综合评价值

$$s_i = \sum_{j=1}^m w_j p_{ij}$$

- 该公式对不同科目加权求和，得到每个人的平均值，评价值越大越好
- $p_{ij}$ 和 $w_j$ 都是原始数据（成绩）求得的，完全客观，不掺杂主观成分

微信公众号：数学建模BOOM

# 熵权法

- 模型简介
- 适用赛题
- 典型例题与原理讲解
- 代码求解

## □代码求解

### □ 调用数据和公式细节

- 使用熵权法时需要数据全面，有的数据是比赛提供的，有的是需要自己找
- 数据往往以excel文件存储，可在MATLAB中用readmatrix函数调用
- 注意调用的excel文件需要与代码文件在同一文件夹下！！
- 接下来到MATLAB实时脚本文件shangquan.mlx中讲解代码

微信公众号：数学建模BOOM

## □ 写出你的笔记

## □ 费曼学习法



费曼学习法

- 费曼学习法：以教代学
- 只有当你能够教会别人，才代表你真正学会了！

## □ 有奖征集：每学完一期课程，整理笔记，发布在各平台

- 将你每节课所学到的，整理出一套笔记
- 尽量不要照搬或截图课程的内容
- 可自行发布在知乎/CSDN等等各类平台

① 确定主题开始学习

② 理解所学内容

③ 把所学内容讲给别人

④ 把讲不清楚的地方去学明白

- 符合以下要求的文章，且文章点赞超过100或浏览量超1万的，可获取半价退款奖励（联系北海的QQ：1980654305）
- 1、标题设为：XXXX（模型或算法）——北海数学建模课程笔记
- 2、文章首行写：本文为北海的数模课程学习笔记，课程出自微信公众号：数学建模BOOM。

## □ “从零开始学数学建模” 系列课程

- 本期课程视频出自**b站up**：数学建模BOOM
- 全套课程请关注**微信公众号**：数学建模BOOM，回复“课程”

# END

微信公众号：数学建模BOOM