地理数据的类型

（1）根据地理数据本身性质不同：定性数据、定量数据。

（2）根据地理数据来源及表征系统的特征不同：社会－经济数据、环境与自然资源数据；

（3）空间数据与属性数据

- ✓ **空间数据：**用于描述地理实体、地理要素、地理现象、地理事件及地理过程产生、存在和发展的地理位置、区域范围及空间联系。(点线面及其拓扑关系)
- ✓ **属性数据：**用于描述地理实体、地理要素、地理现象、地理事件、地理过程的有关属性特征。
  - ✓ **数量标志数据**
    - ① **间隔尺度数据:**以有量纲的数据形式表示测度对象在某种单位(量纲)下的绝对量。
    - ② **比例尺度数据:**以无量纲的数据形式表示测度对象的相对量。
  - ✓ **品质标志数据**
    - ① **有序尺度数据:**当测度标准不是连续的量，只是表示其顺序关系的数据。
    - ② **二元数据:**用 0、1 表示地理事物、地理现象或地理事件的是非判断问题。
    - ③ **名义尺度数据:**数字表示地理实体、地理要素、地理现象或地理事件的状态类型。

地理数据的表达方式：（1）表格（2）地理数据矩阵

地理数据的基本特征：1 数量化、形式化与逻辑化 2 不确定性 3 多种时空尺度 4 多维性

形式化、逻辑化与数量化，是所有地理数据的共同特征。

- ✓ **定量化的地理数据**是建立地理数学模型的基础，其作用为：确定模型的**参数**、给定模型运行的初值条件;检验模型的**有效性**。
- ✓ 地理计算学，对于地理数据的形式化、逻辑化提出了更高的要求，要求“整体”和“大容量”的地理数据具有**统一的数据形式和交换标准**。

不确定性是地理数据的**基本特征**之一。地理数据不确定性的来源：

- ✓ 地理系统**本身的复杂性**从本质上决定着地理数据的不确定性。
- ✓ 各种原因所导致的**数据误差**。

多种时空尺度

- ✓ **从空间尺度上来看**：全球尺度的、洲际尺度的、国家尺度的，也有流域尺度的、地区尺度的、城市尺度的、社区尺度的。
- ✓ **从时间尺度上来看**，如历史年代、天、月、季度、年等。

多维性：地理数据的这种多维性，被人们描述为**地理数据立方体**

对于一个地理对象的具体意义要从**空间、属性、时间**三个方面综合描述:

- ✓ **空间方面，**描述该地理对象所处的地理位置和空间范围，一般需要 <u>2～3 个变量</u>；
- ✓ **属性方面，**描述该地理对象的具体内容，<u>至少需要 1 个以上</u>；
- ✓ **时间方面，**描述该地理对象产生、发展和存在的时间范围 ，需要 <u>1 个变量</u>

地理数据的变换：

1 定性数据转换成定量数据

1 有序数据转换 :不用连续的量, 而是用表示其次序或等级关系的数据来作为测度标准。

2 二元数据转换：即 0、1 转换:　1－是、有；0－否、无

2 定量数据本身转换－数值标准化方法

对数变换、开（立）平方变换、取倒数变换、概率变换、模数变换、指数变换、滑动平均法、差分算子法等

地理数据采集与处理：

地理数据的采集：

渠道来源：野外调查、室内化验分析、定位或半定位观测、从地形图、航片、卫片上提取地理信息、从有关部门收集观测或统计资料

采集地理数据过程中需要注意的问题…

地理数据处理，是所有地理问题研究的**核心环节**。

✓ 从理论上讲，在地理学中，数学方法的运用主要有两个目的：

（1）运用数学语言对地理问题的描述，建立地理数学模型，从**更高、更深层次**上**揭示地理问题的机理**；

（2）运用有关数学方法，通过**定量化的计算和分析**，对地理数据进行处理，从而**揭示有关地理现象的内在规律**。因此，从一定意义上来说，地理数据处理也是计量地理学的任务之一。

✓ 地理信息系统的**核心功能**是地理数据处理，它实现了空间数据与属性数据的完美结合。数学方法确实是其强有力的支撑。

地理计算学（Geocomputation）的实质是借助于现代化的计算理论、计算方法和计算技术，通过对“整体”和“大容量”的地理数据进行处理，揭示复杂地理系统的运行机制，探索和寻求新的地理系统理论。

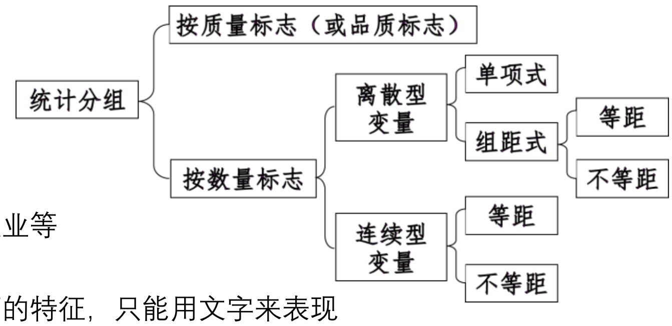
地理数据整理

检查资料； 统计分组。

分组：通过对地理数据分组来研究各组数据出理的频数(次数)和频率，以便概括数据总体结构及其分布特征。

统计指标:总体，可量性，无论是数量指标还是质量指标，都能用数值表示

统计标志：数量标志具有可量性，品质标志不具有可量性。



统计分组——统计标志分类：

1、**按变异情况**，可分为：

不变标志：国籍（人口普查）

变异标志：性别、年龄、民族、职业等

2、**按性质，**可分为：

质量标志：表明总体单位属性方面的特征，只能用文字来表现

数量标志：表明总体单位数量方面的特征，其标志表现可以用数值表示，即标志值

-离散型变量：是指其数值只能用自然数或整数单位计算。

-连续型变量：指在一定区间内可以任意取值的变量。

按组距分组

[1] 求变数的全距（R=max{X}-min{X}）

[2] 确定组数（n=**1+3.32lgN**,经验参考 Sturges 公式）

[3] 计算组距 h=R/n

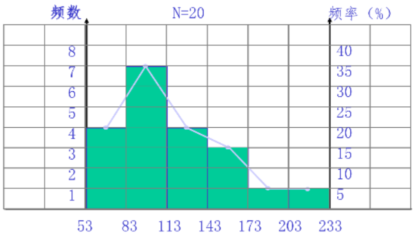
[4] 确定组限，组限修正（边界点），其中**第一组的下限=最小值-1/2 组距**

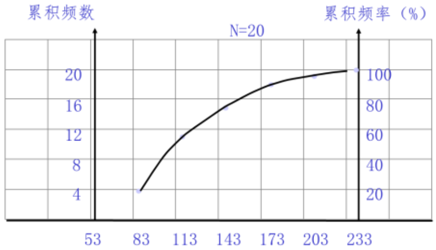
[5] 计算组中值 **m =（下限+上限）/ 2**

[6] 频数分布图表的绘制

频数分布表、频数分布图（直方图、多边形图、累积频数）

组序	组距	组中值	频数	累积频数	频率（%）	累积频率（%）
1	53-82	67.5	4	4	20	20
2	83-112	97.5	7	11	35	55
3	113-142	127.5	4	15	20	75
4	143-172	157.5	3	18	15	90
5	173-202	187.5	1	19	5	95
6	203-233	218	1	20	5	100





（1）“不重不漏”（2）“组距合适”

（一）集中性的代表值

1.平均数：

算术平均数 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=\frac{x_1+x_2+\cdots+x_n}{n}$ $\bar{x}=\frac{f_1x_1+f_2x_2+\cdots+f_nx_n}{\sum f_i}=\frac{1}{N}\sum_{i=1}^{n}f_ix_i$

几何平均数 $\bar{x}_g=\sqrt[n]{x_1\cdot x_2\cdot\cdots\cdot x_n}=\sqrt[n]{\prod_{i=1}^{n}x_i}$ $\ln\bar{x}_g=\frac{1}{n}(\ln x_1+\ln x_2+\cdots+\ln x_n)=\frac{1}{n}\sum_{i=1}^{n}\ln x_i$

环比公式（平均发展速度） $\bar{x}_g=\sqrt[n]{\frac{x_n}{x_0}}$

平均增长速度 ＝ 平均发展速度 — 1

与**历史同时期**比较，例如 2005 年 7 月份与 2004 年 7 月份相比称其为**同比；与上一统计段**比较，例如 2005 年 7 月份与 2005 年 6 月份相比较称其为**环比。**

2.中位数 $M_e=L_{M_e}+\frac{(\frac{N}{2}-f_{m-1})\cdot h}{f_m}$

3.众数 $M_0=L_{M_0}+\frac{f_{m+1}}{f_{m-1}+f_{m+1}}\times h$

单峰正偏态分配上 M0<Me<avg(X)

（二）离散性的代表值

a.绝对离散度

1.**离差** $d_i=x_i-\bar{x}$

2.**离差平方和** $\sum_{i=1}^{n}d_i^2=\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}x_i^2-\frac{(\sum_{i=1}^{n}x_i)^2}{n}$

3.**方差:总体方差** $\sigma^2=\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$ (μ为总体的均值) **样本方差** $S^2=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$

4.**标准差**

b.相对离散度 $C_v=\frac{S}{\bar{x}}=\frac{1}{\bar{x}}\times\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}=\sqrt{\frac{\sum_{i=1}^{n}(\frac{x_i}{\bar{x}}-1)^2}{n-1}}=\sqrt{\frac{\sum_{i=1}^{n}(k_i-1)^2}{n-1}}$ $k_i=\frac{x_i}{\bar{x}}$

- 两系列数据单位不同、均值相差较大、标准差相同时

变差系数 Cv：离差系数、变异系数(ki 为模比系数)

偏度系数 g1：测度地理数据分布的不对称性情况

峰度系数 g2：它测度了地理数据在均值附近的集中程度

$g_1=\sum_{i=1}^{n}\frac{1}{n}\left(\frac{x_i-\bar{x}}{\sigma}\right)^3$

$g_2=\sum_{i=1}^{n}\frac{1}{n}\left(\frac{x_i-\bar{x}}{\sigma}\right)^4-3$

地理数据库：

组成：计算机硬件设备、地理数据库软件（DBMS）、地理数据文件、地理数据库应用程序

建立步骤：(1）用户需求调查（2）系统分析（3）系统设计（4）系统调试和运行

空间分布的测度：(1)点状分布类型(2)线状分布类型：分支、回路、区划
(3)面状分布类型：离散区域分布类型、连续区域分布类型

点状分布的测度

三种模式：凝聚、均等（离散）

随机：均匀平面的泊松随机过程（每个位置上出现点的概率相同；点出现的位置相互独立；）

凝聚：出现一些点分布的高密度区域，点密度的空间变化大；

均等（离散）：比完全随机分布更少聚集性（散布、均一、等距；排斥、竞争；低离差）

邻近平均距离：最邻近距离越小，说明点在空间分布越密集;

1.顺序法(点型分布为**凝聚型**时适用)

1.各点依次作为基准点：
计算该点 dih（到其它全部点距离）和 dib（到区域边界最短距离）;
找出满足 dih≤ dib 的距离；若有 p 个，按顺序排列：di1,di2,…dip,p=0,1,2,..,n-1

2.n 个点得到 n 个序列，取每个序列最小值求均值，第 k 级最邻近距离则取每个序列第 k 小

2.区域法(点型分布为**随机型、均等型**时适用)

1.各点依次作为基准点，以之为中心将研究区分割成 k 个等大的齿轮状区域（一般 6 个）
量度 dih（区内中点到最邻近点的距离），dib（到区域边界的最短距离）;
找出满足 dih≤ dib 的距离；若有 m 个，按顺序排列：di1,di2,…dim,m<=k

2. n 个点得到 n 个序列，取每个序列最小值求均值，第 k 级最邻近距离则取每个序列第 k 小

邻近指数（R） $R=\frac{\bar{d}_1}{\bar{d}_E}$

$\bar{d}_E=\frac{1}{2\sqrt{D}}$ 为理论的随机分布型的最邻近平均距离（泊松分布时理想平均距离）。

D=n/A，为点的密度，其中 A 为区域面积，n 为区域内点的个数。

R=1，随机型分布； R<1，趋向于凝集型分布； R>1，趋向于离散型的均匀分布。

1.可以把要讨论的点的空间分布图式放在一个从凝集的、通过随机的一直到均匀分布的**连续的定量范围**之内，此尺度范围为：0-2.149。一般在 0.33-1.67 之间。

2.对于一个固定地域，点的空间分布随时间而变化，可通过 R 尺度分析去判断：
1 其空间分布**比原先的是更凝集还是更趋于分散**；2 定量的表达出其**凝集或分散的程度**。

对中心位置的测度

中项中心：作一对经纬线使得划分的四部分点数一致，交点即中项中心

中项中心总是偏向**分布点密度较大的一侧**，选择这样的中心，可以使中心与多数分布点之间取得较好的联系。寻找中项中心的过程比较简便，应用也较广。

平均中心（分布重心）：作 x，y 轴； 确定每一点的坐标； 计算坐标均值。

当某一空间现象的空间均值**显著区别于区域几何中心**，就说明这一空间现象的不均衡分布，或称“**重心偏离**”。偏离方向指示了空间现象的“高密度”部位，偏离的距离则指示了**均衡程度**。

离散程度的测度

① 对于平均中心（中项中心）的离散程度

d_b^2/d^2：趋于1时，区域间具有最大差异性；
接近于0时，各区域具有与整个模型相同的平均中心

1 计算标准距离：离平方的均值开根号

2 划分小区，计算小区内标准距离 dω，再计算全区域中小区间标准距离 db,d2=两者平方和

3 中项中心四个区域面积分别与总面积之比 $I_d=\frac{Q}{A}$

4 按点与选择的中心之间的距离进行分组，画出频率累积曲线，读出占 50%的累积频率半径。

②对于任意指定中心的离散程度

③各点之间离散程度的测定：邻近指数 R

线状分布的测度

网络图：是仅由一些点以及点之间的连线所组成的图形。(**顶点、边**)

有向图：如果 G 的每条边给定了方向，即：　　(v，e)≠（e，v)，则称 G 为有向图；

无向图：如果 G 的每条边都没有方向，即：　　(v，e)=（e，v)，则 G 称为无向图。

赋权图：如果图 G =（V，E）中的每一条边（vi，vj）都相应地赋有一个数值 wij，则称 G 为赋权图，其中 wij 称为边（vi，vj）的**权值**。除了可以给图的边赋权外，也可以给图的顶点赋权。即，对图 G 中的每一顶点 vj，也可以赋予一个**载荷** a（vj）。

图的定义只关注点之间是否连通，而不关注点之间的连结方式。

对于许多地理问题，当它们被抽象为图论意义下的网络图时，问题的核心就变成了**网络图上的优化计算**问题。其中，最为常见的是关于**路径**和**顶点**的优选计算问题。

在路径的优选计算问题中，最常见的是最短路径问题；

而在顶点的优选计算问题中，最为常见的是中心点和中位点选址问题。

最短路径问题

权值：空间距离、经济距离、时间距离

标号法（Dijkstar 算法）

不仅可以求出起点到终点的最短路径及其长度

而且可以求出起点到其他任何一个顶点的最短路径及其长度；

同时适用于求解有向图或无向图上的最短路径问题。

1.画表初始化 0 或正无穷，　取第一行最小值（出发点为 0），该店为当前点

2.更新一步最短值比较与原来的值，取最小值，该点最短路径确定

3.以 2 中确定点为当前点，重复 2

画表时注意： 1.某点最短路径确定时，应圈起来，下一行时不再补格子，若多行同值，圈出所有同值的格子以便回溯 2.选点时还要记录该点最短距离

服务点的最优区位

1.求 G 的最短距离矩阵，2.求每一顶点最大服务距离（1 中行最大值）3.求 2 中值最小的点

2 中可加入顶点载荷

面状分布的测度

1）　离散区域分布测度

空间罗伦兹曲线：分组（区域），两列属性，按属性比值降序排列，做累积百分比图

基尼系数：洛伦兹曲线与对角线所夹面积

< 0.2 ：收入绝对平均；0.2-0.3：收入比较平均；0.3-0.4：收入相对合理；

0.4-0.5：收入差距较大；>0.5 ：收入差距悬殊

集中化指数 $I = \frac{C-R}{M-R}$

洛伦兹曲线中：C:各点 x 轴（总产值）之和，R:各点 y 轴（部门产值）之和，

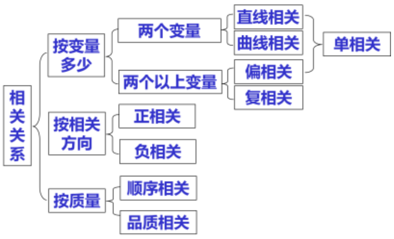
M=n(最不均匀时的 C)

2）　连续区域分布测度

高程曲线：相对高度=（绝对高度—高程最低值）/（高程最高值—高程最低值）

相关分析与回归分析

1.相关分析；2.回归分析；3.趋势面分析



1.相关分析：**简单线性相关程度的测度**

自变量和因变量的判定

地理相关

地理要素之间关系的类型：函数关系（完全相关）、相关关系（统计相关）、独立

1.1 相关表

　　某一变量按其取值大小顺序排列

1.2 相关图（散点图）

1.3 相关系数

　　A）Pearson's R 積差相關：离差积和与离差平方和积开方之比

$$r=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2\cdot\sum(y_i-\bar{y})^2}}=\frac{l_{xy}}{\sqrt{l_{xx}\cdot l_{yy}}}$$

$$l_{xx}=\sum x_i^2-\frac{(\sum x_i)^2}{n}\qquad l_{yy}=\sum y_i^2-\frac{(\sum y_i)^2}{n}\qquad l_{xy}=\sum x_i y_i-\frac{(\sum x_i)(\sum y_i)}{n}$$

0-0.3-0.5-0.8-0.8-1 完全不-微弱-低度-显著-高度-完全

显著性检验，f=n-2

　　B）Spearman's R 等级相关：

（1）　未有相同等級者：$r_s=1-\frac{6\sum D^2}{N(N^2-1)}$　（D 為二變數對稱之等級差）

（2）　有相同等級者：$r_s=\frac{\sum x^2+\sum y^2-\sum D^2}{2\sqrt{\sum x^2\sum y^2}}$

$\sum x^2=\frac{N^3-N}{12}-\sum Tx$　$\sum y^2=\frac{N^3-N}{12}-\sum Ty$　$\sum T=\frac{t^3-t}{12}$　t：表示得到相同等第的人數。

　　C）多要素相关：相关矩阵，偏相关与复相关

1.两两算相关（Pearson's R 单相关）得到相关矩阵

2.一阶偏相关（控制 z 后的 XY 相关性）：$r_{xy,z}=\frac{r_{xy}-r_{xz}r_{yz}}{\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)}}$

3.多阶偏相关（M 可迭代）$r_{xy,mz}=\frac{r_{xy,m}-r_{xz,m}r_{yz,m}}{\sqrt{(1-r_{xz,m}^2)(1-r_{yz,m}^2)}}$

①偏相关系数分布的范围在-1 到 1 之间；

②偏相关系数的绝对值越大，表示其偏相关程度越大；

③偏相关系数的绝对值必小于或最多等于由同一系列资料所求得的复相关系数。

显著性检验 $t=\frac{r\sqrt{n-q-2}}{\sqrt{1-r^2}}$

r 为偏相关系数，n 为样本数，q 为阶数，统计量服从 n-q-2 个自由度的 t 分布

2.回归分析

作用：确定数学表达式（建立回归模型）、预测

（1）根据变量的多少 $\begin{cases}\text{一元回归}\\\text{多元回归}\end{cases}$

（2）根据所建回归模型形式 $\begin{cases}\text{线性回归}\\\text{非线性回归}\end{cases}$

2.1.线性回归与非线性回归

通过变换坐标轴（取 ln）将线性模型转化为幂指对模型

2.2 一元线性回归：最小二乘法 $y_i = A + Bx_i + \varepsilon_i$

回归值 $\hat{y}_i = a + bx_i$

$$\delta_i = y_i - \hat{y}_i \quad \text{Min } Q = \sum \delta_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - a - bx_i)^2$$

2.2.1 解：

$$\begin{cases}\dfrac{\partial Q}{\partial a} = -2\sum (y_i - a - bx_i) = 0\\[2mm] \dfrac{\partial Q}{\partial b} = -2\sum (y_i - a - bx_i)x_i = 0\end{cases} \quad \text{即} \quad \begin{cases}\sum (y_i - a - bx_i) = 0\\ \sum (y_i - a - bx_i)x_i = 0\end{cases}$$

$$a = \frac{1}{n}\sum y_i - b\frac{1}{n}\sum x_i$$

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\sum x_i y_i - \dfrac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \dfrac{(\sum x_i)^2}{n}} = \frac{l_{xy}}{l_{xx}}$$

2.2.2 效果检验

A 标准估计误差（标准标准差）$S = \sqrt{\dfrac{\sum (y - \hat{y})^2}{n-2}}$

B 显著性检验 (方差分析)1.离差平方和 $S_{总} = l_{yy} = \sum (y_i - \bar{y})^2$

又有 $y_i - \bar{y} = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})$

$$\sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2 + 2\sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y})$$

得 $\sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2$

记为 S 总=U+Q 或 $SS_T = SS_R + SS_E$

即 **离差平方和=回归平方和+剩余（残差）平方和**

$$\begin{aligned} SS_R &= \sum (\hat{y} - y)^2 \\ &= \sum (\hat{y} - y)(a + bx - a + b\bar{x}) \\ &= b\sum (\hat{y} - y)(\hat{x} - x) \\ &= bl_{xy} \end{aligned}$$

$$Q = \sum (y_i - \hat{y}_i)^2 = l_{yy} - U = l_{yy} - bl_{xy}$$

方差分析

$$U / l_{yy} = \frac{bl_{xy}}{l_{yy}} = \frac{l_{xy}^2}{l_{xx} \cdot l_{yy}} = r^2$$

$$\Rightarrow U = r^2 l_{yy}$$

$$Q = (1 - r^2) l_{yy}$$

总平方和的自由度也可以分解为：

$$\underset{N-1}{f_{总}} = \underset{1}{f_U} + \underset{N-2}{f_Q}$$

$$y - \bar{y} = (y - \hat{y}) + (\hat{y} - y)$$

$$\Rightarrow \Sigma(y - \bar{y})^2 = \Sigma(y - \hat{y})^2 + \Sigma(\hat{y} - \bar{y})^2 + 2\Sigma(y - \hat{y})(\hat{y} - \bar{y})$$

$$\begin{aligned} \because \quad \Sigma(y - \hat{y})(\hat{y} - \bar{y}) &= \Sigma(y - \hat{y})(a + bx - \bar{y}) \\ &= \Sigma(y - \hat{y})[(a - \bar{y}) + bx] \\ &= (a - \bar{y})\Sigma(y - \hat{y}) + b\Sigma(y - \hat{y})x \\ &= (a - \bar{y})\Sigma(y - a - bx) + b\Sigma(y - a - bx)x \end{aligned}$$

根据最小二乘法原理，有：

$$\Sigma(y - a - bx) = 0，\quad \Sigma(y - a - bx)x = 0$$

$$\therefore \quad \Sigma(y - \hat{y})(\hat{y} - \bar{y}) = 0$$

$$\therefore \quad \Sigma(y - \bar{y})^2 = \Sigma(y - \hat{y})^2 + \Sigma(\hat{y} - \bar{y})^2$$

变差来源	平方和	自由度	方差	F检验
回归（因素x）	$U = bl_{xy}$	1	$S_U = U/1$	$F = \dfrac{U/1}{Q/(N-2)}$
剩余（随机因素）	$Q = l_{yy} - bl_{xy}$	N-2	$S_Q = Q/(N-2)$	
总和	$S_{总} = l_{yy}$	N-1		

$F > F_\alpha(1, N-2)$时显著

2.4 多元回归模型 $y_a = \beta_0 + \beta_1 x_{1a} + \beta_2 x_{2a} + \cdots + \beta_k x_{ka} + \varepsilon_a$

$b_0, b_1, \cdots, b_k$ 为 $\beta_0, \beta_1, \beta_2, \cdots, \beta_k$ 拟合值，称偏回归系数 $\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k$

$$Q = \sum_{a=1}^{n}(y_a - \hat{y}_a)^2 == \sum_{a=1}^{n}[y_a - (b_0 + b_1 x_{1a} + b_2 x_{2a} + \cdots + b_k x_{ka})]^2 \to \min$$

由最小二乘法：

$$\begin{cases} \dfrac{\partial Q}{\partial b_0} = -2\sum_{a=1}^{n}(y_a - \hat{y}_a) = 0 \\ \dfrac{\partial Q}{\partial b_j} = -2\sum_{a=1}^{n}(y_a - \hat{y}_a)x_{ja} = 0 \quad (j = 1, 2, \cdots, k) \end{cases}$$

展开得正规方程

$$\begin{cases} nb_0 + (\sum_{a=1}^{n} x_{1a})b_1 + (\sum_{a=1}^{n} x_{2a})b_2 + \cdots + (\sum_{a=1}^{n} x_{ka})b_k = \sum_{a=1}^{n} y_a \\ (\sum_{a=1}^{n} x_{1a})b_0 + (\sum_{a=1}^{n} x_{1\alpha}^2)b_1 + (\sum_{a=1}^{n} x_{1a}x_{2a})b_2 + \cdots + (\sum_{a=1}^{n} x_{1a}x_{ka})b_k = \sum_{a=1}^{n} x_{1a}y_a \\ (\sum_{a=1}^{n} x_{2a})b_0 + (\sum_{a=1}^{n} x_{1a}x_{2a})b_1 + \sum_{a=1}^{n}(x_{2a}^2)b_2 + \cdots + (\sum_{a=1}^{n} x_{2a}x_{ka})b_k = \sum_{a=1}^{n} x_{2a}y_a \\ \cdots\cdots\cdots\cdots \\ (\sum_{a=1}^{n} x_{ka})b_0 + (\sum_{a=1}^{n} x_{1a}x_{ka})b_1 + (\sum_{a=1}^{n} x_{2a}x_{ka})b_2 + \ldots + (\sum_{a=1}^{n} x_{ka}^2)b_k = \sum_{a=1}^{n} x_{ka}y_a \end{cases}$$

$$\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2, \cdots \beta_k], \quad \mathbf{b} = [b_0, b_1, b_2, \ldots, b_k]^T, \quad \mathbf{x} = [1, x_1, x_2, \cdots x_k]^T$$

$$\mathbf{y} = \mathbf{x}^T\boldsymbol{\beta} + \varepsilon, \quad \hat{y} = \mathbf{x}^T\mathbf{b}$$

$$\min \ Q = \frac{1}{2}\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = \frac{1}{2}\|\boldsymbol{\varepsilon}\|^2 = \frac{1}{2}\|\mathbf{y} - \mathbf{x}^T\boldsymbol{\beta}\|^2$$

① 回归平方和U与剩余平方和Q:

$$S_{总} = L_{yy} = U + Q$$

② 回归平方和

$$U = \sum_{a=1}^{n}(\hat{y}_a - \bar{y})^2 = \sum_{a=1}^{n} b_i L_{iy}$$

③ 剩余平方和为

$$Q = \sum_{a=1}^{n}(y_a - \hat{y}_a)^2 = L_{yy} - U \quad x_1, x_2, \cdots, x_k$$

④ F统计量为

$$F = \frac{U/k}{Q/(n-k-1)}$$

计算出来F之后，可以查F分布表对模型进行显著性检验

2.4 一元非线性回归模

通过变换坐标轴（取 ln）将线性模型转化为幂指对模型（换元为线性问题）

$y = ax^b \qquad y = ae^{bx} \qquad y = a + b\ln x \qquad y = dx_1^{\beta_1} \cdot x_2^{\beta_2} \cdots x_k^{\beta_k}$

$\dfrac{1}{y} = a + \dfrac{b}{x} \qquad y = \dfrac{1}{a + be^{-x}}$, 令$y' = \dfrac{1}{y}, x' = e^{-x} \qquad y = \beta_0 + \beta_1 \ln x_1 + \beta_2 \ln x_2 + \cdots + \beta_k \ln x_k$

相关指数（与剩余平方和相似）$R^2 = 1 - \dfrac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} = 1 - \dfrac{Q}{l_{yy}}$

根号得到相关系数 R

2.5 趋势面分析

趋势面时一种代表地理数据在大范围内得空间变化趋势的光滑曲面，是实际曲面的近似。

即：**实际曲面=趋势面+剩余曲面**

某观测点的观测值=确定性函数值+随机性函数值=趋势值+剩余值

$z_i(x_i, y_i) = T_i(x_i, y_i) + e_i = \hat{z}_i + e_i \qquad \hat{z}_i = T_i(x_i, y_i)$(趋势值) $\quad e_i = z_i - \hat{z}_i$(残差)

$$Q = \sum_{i=1}^{n}(z_i - \hat{z}_i)^2 \Rightarrow \min$$

$$z = b_0 + b_1 x + b_2 y$$

$$z = b_0 + b_1 x + b_2 y + b_3 x^2 + b_4 xy + b_5 y^2$$

$$z = b_0 + b_1 x + b_2 y + b_3 x^2 + b_4 xy + b_5 y^2 + b_6 x^3 + b_7 x^2 y + b_8 xy^2 + b_9 y^3$$

将多项式回归（非线性模型）模型转化为多元线性回归模型。

$$x_1 = x, x_2 = y, x_3 = x^2, x_4 = xy, x_5 = y^2, \cdots\cdots$$

F 检验 ：U 为回归平方和; Q 为残差平方和（剩余平方和）;
p 为多项式的项数（不包括 b0）; n 为使用的资料数目。

$$F = \frac{U / p}{Q / (n - p - 1)}$$

拟合程度指数：$I = \left[1 - \frac{\sum (z_i - \hat{z}_i)^2}{\sum (z_i - \bar{z}_i)^2}\right] \times 100\%$　I≥75%时，拟合程度较为满意

线性规划（LP）

1 某项任务确定后，如何统筹安排，以最少的人力、物力和财力去完成该项任务；
2 面对一定数量的人力、物力和财力资源，如何安排使用，使得完成的任务最多。
它们都属于最优规划的范畴

线性规划的标准形式： 1）目标函数为 max 2）约束条件为等式 3）决策变量为非负约束

在约束条件

$a_{11} x_1 + a_{12} x_2 + \cdots + a_{1n} x_n = b_1$
$a_{21} x_1 + a_{22} x_2 + \cdots + a_{2n} x_n = b_2$
……
$a_{m1} x_1 + a_{m2} x_2 + \cdots + a_{mn} x_n = b_m$
$x_j \geqslant 0$（j = 1，2，…，n）

$$\sum_{j=1}^{n} a_{ij} x_j = b_i (i = 1, 2, \cdots, m)$$

$$\begin{cases} \max Z = CX \\ AX = b \\ X \geq 0 \end{cases}$$

下，求一组未知变量 x_j（j = 1，2，…，n）的值，使 $Z = \sum_{j=1}^{n} c_j x_j \to \max$

若第 k 个约束方程为不等式，则引入**松弛变量**改写成等式。

可行解与最优解

满足约束条件（即满足线性约束和非负约束）的一组变量为**可行解**。所有可行解组成的集合称为**可行域**。使目标函数最大或最小化的可行解称为**最优解**。

如果 B 是 A 中的一个阶的非奇异子阵，则称 B 为该线性规划问题的一个**基**。
则称非奇异阵为**基向量**，与基向量相对应的变量为**基变量**，而其余的变量为**非基变量**。
如果 $X_B = [x_1, x_2, \cdots, x_m]^T$ 是方程组的解，则 $X = [x_1, x_2, \cdots, x_m, 0, 0, \cdots, 0]^T$ 就是方程组 *Z= CX* 的一个解，它称之为对应于基 *B* 的**基本解**，简称**基解**。

1.化为标准形式
2.做单纯形表
3.化可行基、化检验行为为基变量系数为零的形式，判别所有非基变量检验系数是否都为负
4.若非则换基变量，优先将 Z 表达式中含有的变量选为基变量
直到检验行中：基变量系数为零的，非基变量系数为负

时间序列

时间序列分析的基本原理

两种形式：**增长和下降**　　四种序列：**线性、对数、指数、波动**

时间序列概念：是要素（变量）的数据按照**时间顺序**排列而形成的一种数列，反映**变量时间变化的发展过程**。地理过程的时间序列分析，就是通过分析地理要素（变量）随时间变化历史过程，揭示其发展**变化规律**，并对其未来状态进行**预测**。

长期趋势（T），是时间序列随时间的变化而逐渐增加或减少的**长期变化**之趋势。
季节变动（S），是时间序列在**一年中或固定时间**内，呈现出的**固定规则**的变动。
循环变动（C），是指沿着**趋势线如钟摆般地循环变动**，又称景气循环变动。
不规则变动（I），是指在时间序列中由于**随机因素**影响所引起的变动。

时间序列的分解模型

- 时间序列 y 可以表示为以上四个因素的函数，即：yt=f(Tt,St,Ct,It)
- 常用的模型有：加法模型：Y=T+S+C+I 乘法模型：Y=T×S×C×I

分解步骤：

（1）运用**移动平均法剔除长期趋势和周期变化**，得到序列 *TC*。然后再用按**月（季）平均法**求出季节指数 *S*。
（2）做散点图，选择适合的**曲线模型拟合序列的长期趋势**，得到长期趋势 T。
（3）计算周期因素 C。用序列 TC 除以 T 即可得到周期变动因素 C。
（4）将时间序列的 T、S、C 分解出来后，剩余的即为不规则变动，即：I=Y/TSC

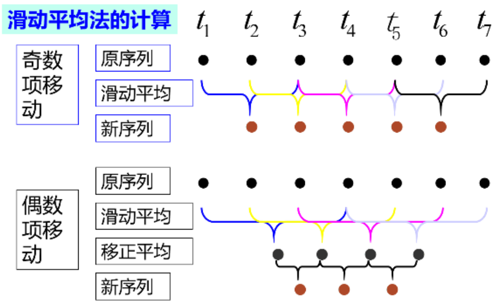
长期趋势分析：平滑法、趋势线法、自回归分析法

时间序列分析的平滑法主要有三类　：

移动平均法：移动平均=最近 N 期之和/N，N 为移动时距

当数据的**随机因素较大时，宜选用较大的 *N*，**这样有利于较大限度地平滑由随机性所带来的严重偏差；反之，当数据的**随机因素较小时，宜选用较小的 *N***，这有利于跟踪数据的变化，并且预测值滞后的期数也少。

滑动平均法：单侧滑动时距 l (2l+1) 点滑动平均



使用滑动平均法应注意的问题：

- 滑动平均法可以平滑修匀序列；
- 对于**季节性序列，要采用 4 项或 12 项滑动平均**，方可平滑掉其季节波动；
- **一般的滑动平均方法使原序列首尾各去除了若干项，因此不能用于外推预测**；
- 当序列没有明显的长期趋势、季节变动和循环变动时，可以用移动平均法进行预测，但要进行特别的计算处理。

指数平滑法：时间序列的态势具有**稳定性或规则性**，所以时间序列可被合理地顺势推延；他认为**最近的过去态势，在某种程度上会持续的未来**，所以将较大的权数放在最近的资料。

任一期的指数平滑值都是本期实际观察值与前一期指数平滑值的加权平均。

一次指数平滑法：利用前一期的预测值，订正得到预测通式（平滑系数α平稳取小波动取大）变化趋势较稳定，以直接用第一个数据作为初始值；变动趋势有起伏**波动**时，则应以 n 个数据的**平均值**为初始值，以减少初始值对平滑值的影响。

高次指数平滑：平滑后平滑

$$\hat{y}_{t+1} = \sum_{j=0}^{t-1} \alpha(1-\alpha)^j y_{t-j} = \alpha y_t + (1-\alpha)\hat{y}_t$$

趋势线拟合法：用某种趋势线（直线或曲线）来对原序列的**长期趋势进行拟合**。其主要作用是进行外推预测。

直线趋势方程、曲线趋势方程（指数函数、二次函数）

1 判断趋势类型 2 计算待定参数 3 利用方程预测

1 判断趋势类型

散点图法

差分法

当数据的一阶差分趋近于一常数时，可以配合直线方程

当数据的二阶差分趋近于一常数时，可以配合二次曲线方程。

当数据的环比趋近于一常数时，可配合指数曲线方程。yi/yi+1=const

2 计算待定参数

线性回归（最小二乘）

3 利用方程预测

自回归模型：时间序列的自相关，是指序列前后期数值之间的相关关系，对这种相关关系程度的测定便是自相关系数

K 阶自相关系数：y1,y2,···,yt-1,···,yn-1 与 yk+1,yk+2,···,yk+t, ···,yk+n 的相关系数

$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varepsilon_t$ p 阶线性自回归模型 $y_t = \varphi_0 + \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + \varepsilon_t$

季节变动分析：是对包含**季节波动**的**时间序列**进行**预测**的方法

经济现象由于自然条件、生产条件和生活习惯等因素的影响，随着季节的转变而呈现的**周期性**变动。这种周期通常为 1 年。

规律性。每年重复出现，其表现为逐年同月(或季)有相同的变化方向和大致相同的变化幅度。

季节变动：一年之内因**纯季节原因**造成的序列的波动，以及**与季节无关的类似**的变动。

要点：

首先，利用**统计**方法计算出预测目标的季节指数，以测定季节变动的规律性；

然后，在已知季节的平均值的条件下，预测未来某个月（季）的预测值。

1）对原时间序列求移动平均，以消除季节变动和不规则变动，保留长期趋势；

2）将原序列 y 除以其对应的趋势方程值（或平滑值），分离出季节变动（含不规则变动），即 SI=TSCI/TS,TS=平滑值（**S=各季平均/历年季度总平均数**）

3）将月度（或季度）的季节指标加总，以由计算误差导致的值去除理论加总值，得到一个**校正系数**，并以**该校正系数乘以季节性指标**从而获得调整后季节性指标。

4）求预测模型，若求下一年度的预测值，延长趋势线即可；若求各月（季）的预测值，需以**趋势值乘以各月份（季度）的季节性指标**。

主成分分析：主成分分析是采取一种数学降维的方法，找出几个综合变量来代替原来众多的变量，使这些综合变量能尽可能地代表原来变量的信息量，而且彼此之间互不相关。

（一）每一成分，归一化处理、计算相关系数矩阵

（二）计算特征值与特征向量 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$

① 解特征方程 $|\lambda I - R| = 0$，常用雅可比法（Jacobi）求出特征值，并使其按大小顺序排列

② 分别求出对应于特征值λi 的特征向量 ei(i=1,2,…,p)，要求||ei||=1，即 $\sum_{j=1}^{p} e_{ij}^2 = 1$，其中 eij 表示向量 ei 的第 *j* 个分量。

③ 计算主成分贡献率及累计贡献率,一般取累计贡献率达 85%~95%的特征值所对应的前 *m*（*m*≤*p*）个主成分。

④ 计算主成分载荷 $p(z_i, x_j) = \sqrt{\lambda_i} e_{ij} (i, j = 1,2, \cdots, p)$，表示与原成分的正负相关性

⑤ 计算主成分得分矩阵 Z

马尔可夫预测方法:预测事件发生的概率的方法。它是基于马尔可夫链，根据事件的目前状况预测其将来各个时刻（或时期）变动状况的一种预测方法。马尔可夫预测法是对地理事件进行预测的基本方法，它是地理预测中常用的重要方法之一.

状态：指某一事件在某个时刻（或时期）出现的某种结果。如：商品销售“畅销”、“一般”、“滞销”等状态。农业收成预测中，有“丰收”、“平收”、“歉收”等状态。

状态转移过程：事件的发展，从一种状态转变为另一种状态，称为状态转移。

马尔可夫过程：事件发展过程中，若**每次状态的转移都仅与前一时刻的状态有关，而与过去的状态无关**，或者说状态转移过程是**无后效性**的，这样的状态转移过程称为马尔可夫过程。

状态转移概率：在事件的发展变化过程中，从某一种状态出发，下一时刻转移到其他状态的可能性，称为状态转移概率。由状态 Ei 转为状态 Ej 的状态转移概率 P(Ei→Ej)就是条件概率 P(Ei|Ej),即 **P(Ei→Ej)=P(Ei|Ej)**

状态转移概率矩阵：假定某一个事件的发展过程有 n 个可能的状态，即 E1，E2，…，En。记为从状态 Ei 转变为状态 Ej 的状态转移概率，则矩阵 P=..称为状态转移概率矩阵。

概率矩阵:满足每行和为 1，取值 0-1 的矩阵

计算状态转移概率矩阵 *P*，就是求从每个状态转移到其他任何一个状态的状态转移概率 Pij。为了求出每一个 Pij，一般采用**频率近似概率**的思想进行计算。

在 15 个从 E_1 出发（转移出去）的状态中:有 3 个是从 E_1 转移到 E_1 的（即 1→2，24→25，34→35）；有 7 个是从 E_1 转移到 E_2 的（即 2→3，9→10，12→13，15→16，29→30，35→36，39→40）；有 5 个是从 E_1 转移到 E_3 的（即 6→7，17→18，20→21，25→26，31→32）。

状态概率：表示事件在初始（k＝0）状态为已知的条件下，经过 k 次状态转移后，在第 k 个时刻（时期）处于状态 Ej 的概率。且 $\sum_{j=1}^{n} \pi_j(k) = 1$，根据马尔可夫过程的无后效性及 Bayes 条件概率公式，有 $\pi_j(k) = \sum_{i=1}^{n} \pi_i(k-1) P_{ij} \quad (j = 1,2, \cdots, n)$

记行向量π(k)=[π1(k),π2(k),…,πn(k)]，则可以得到逐次计算状态概率的递推公式$\pi(k)=\pi(0)P^k$

π(0)为初始状态概率向量

第 k 个时刻（时期）的状态概率预测：求π(k)

终极状态概率预测：无穷多次状态转移后所得到的状态概率称为终极状态概率π(k)，k→∞

π=πP,求解即可

地理事件的预测中，被预测对象所经历的过程中各个阶段（或时点）的状态和状态之间的转移概率是最为关键的。马尔可夫预测的基本方法就是利用状态之间的转移概率矩阵预测事件发生的状态及其发展变化趋势。马尔可夫预测方法的基本要求是**状态转移概率矩阵必须具有一定的稳定性**。因此，必须具有**足够的统计数据**，才能保证预测的精度与准确性。换句话说，马尔可夫预测模型必须建立在大量的统计数据的基础之上。这一点也是运用马尔可夫预测方法预测地理事件的一个最为基本的条件。

AHP 决策分析（层次分析法）

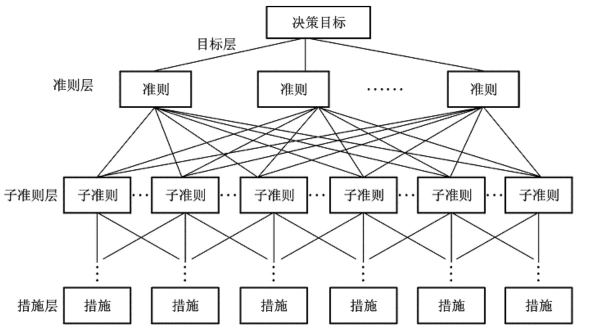
一种**定性与定量**相结合的决策分析方法。它常常被运用于**多目标、多准则、多要素、多层次**的**非结构化**的复杂决策问题，特别是战略决策问题的研究，具有十分广泛的实用性。对复杂问题的决策思维过程**模型化、数量化**的过程。通过这种方法，可以将复杂问题分解为**若干层次和若干因素**，在各因素之间进行简单的比较和计算，就可以得出不同方案重要性程度的权重，从而为决策方案的选择提供依据。

1 明确问题:弄清问题的范围，所包含的因素，各因素之间的关系等，以便尽量掌握充分信息。

2 建立层次结构模型

在这一个步骤中，要求将问题所含的要素进行分组，把每一组作为一个层次，并将它们按照：**最高层（目标层）—若干中间层（准则层）—最低层（措施层）**的次序排列起来。

这种层次结构模型常用结构图来表示，图中要标明上下层元素之间的关系。



如果某一个元素与下一层的所有元素均有联系，则称这个元素与下一层次存在有**完全层次的关系**。

如果某一个元素只与下一层的部分元素有联系，则称这个元素与下一层次存在有**不完全层次**的关系。

层次之间可以建立子层次，子层次从属于主层次中的某一个元素，它的元素与下一层的元素有联系，但**不形成独立层次**。

阶数	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49	1.52	1.54	1.56	1.58	1.59

3 构造判断矩阵

b_{ij} 表示对于 A_k 而言，元素 B_i 对 B_j 的相对重要性程度的判断值。

一般判断矩阵的数值是根据数据资料、专家意见和分析者的认识，加以平衡后给出的。

如果判断矩阵存在关系 $b_{ij}=b_{ik}/b_{jk}$（i，j，k＝1，2，3，…，n），则称它具有**完全一致性**。

考察 AHP 决策分析方法得出的结果是否基本合理，需要对判断矩阵进行**一致性检验**。

4 层次单排序

①目的：确定本层次与上层次中的某元素有联系的各元素重要性次序的权重值。

②任务：计算判断矩阵的特征根和特征 $BW=\lambda_{\max}W$

③检验判断矩阵的一致性:通过前面的分析，我们知道，如果判断矩阵 B 具有完全一致性时,λmax＝n。但是，在一般情况下是不可能的。为了检验判断矩阵的一致性，需要计算它的**一致性指标** $CI=\dfrac{\lambda_{\max}-n}{n-1}$ **CI＝0** 时，判断矩阵具有完全一致性；反之，**CI 愈大，一致性越差**。为了检验判断矩阵是否具有令人满意的一致性，需要将 CI 与平均**随机一致性指标 RI** 进行比较。一般而言，1 或 2 阶的判断矩阵总是具有完全一致性的。对于 2 阶以上的判断矩阵，其一致性指标 CI 与同阶的平均随机一致性指标 RI 之比，称为**判断矩阵的随机一致性比例**，记为 CR。一般地，当 $CR=\dfrac{CI}{RI}<0.10$时，就认为判断矩阵具满意的一致性，否则要调整矩阵。

5.层次总排序

①定义：利用同一层次中所有层次单排序的结果，就可以计算针对上一层次而言，本层次所有元素的重要性权重值，这就称为层次总排序。

②层次总排序需要从上到下逐层顺序进行。对于最高层而言，其层次单排序的结果也就是总排序的结果。

层次 A / 层次 B	A_1	A_2	…	A_m	B 层次的总排序
	a_1	a_2	…	a_m	
B_1	b_1^1	b_1^2	…	b_1^m	$\sum_{j=1}^{m} a_j b_1^j$
B_2	b_2^1	b_2^2	…	b_1^m	$\sum_{j=1}^{m} a_j b_2^j$
⋮	⋮	⋮		⋮	⋮
B_n	b_n^1	b_n^2	…	b_n^m	$\sum_{j=1}^{m} a_j b_n^j$

一致性检验 CI= $\sum_{j=1}^{m} a_j CI_j$ RI= $\sum_{j=1}^{m} a_j RI_j$

近似算法求解判断矩阵的**最大特征根**及其所对应的特征向量

1.方根法:

1.计算判断矩阵每一行元素的乘积 $M_i = \prod_{j=1}^{n} b_{ij} (i = 1,2,\cdots,n)$

2.计算 Mi 的 *n* 次方根 $\overline{W}_i = \sqrt[n]{M_i} (i = 1,2,\cdots,n)$

3.将向量 W = [W1,W2,⋯,Wn]归一化(除以平均),则 W 即为所求的特征向量。

4.计算最大特征根 $\lambda_{\max} = \sum_{i=1}^{n} \frac{(AW)_i}{nW_i}$，(AW)i 表示向量 *AW* 的第 *i* 个分量。

2.和积法:

1.将判断矩阵每一列归一化 $\bar{b}_{ij} = b_{ij} \Big/ \sum_{k=1}^{n} b_{kj} (i = 1,2,\cdots,n)$

2.对按列归一化的判断矩阵，再按行求和 $\overline{W}_i = \sum_{j=1}^{n} \bar{b}_{ij} (i = 1,2,\cdots,n)$

1.思路简单明了，它将决策者的**思维过程条理化、数量化**，便于计算，容易被人们所接受；
2.所需要的**定量化数据较少**，但对**问题的本质**，**问题所涉及的因素**及其内在关系分析得比较透彻、清楚。

1.存在着较大的随意性。譬如，对于同样一个决策问题，如果在**互不干扰**、**互不影响**的条件下，让不同的人同样都采用 AHP 决策分析方法进行研究，则他们所建立的**层次结构模型**、**所构造的判断矩阵**很可能是各不相同的，分析所得出的结论也可能各有差异。
多**部门、多领域的专家共同会商、集体决定**

随机型决策分析方法:是处理随机型决策问题的分析技术。由于许多地理问题与地理数据具有**随机性**特征，所以许多地理决策问题属于随机型决策问题。因此，随机型决策分析方法是地理学中必不可少的方法。

一般来说，凡是根据**预定的目标**做出的**任何行动决定**，都可以称之为**决策**。

决策问题:在实际生产或生活问题中，对于一个需要处理的事件，面临**几种客观条件**，又有**几种可供选择的方案**，这就构成了一个决策问题。

自然状态:在决策问题中，决策者所面临的**每一种客观条件**就称之为一个自然状态，简称**状态或条件**，有时也称为**状态变量**。

行动方案:在决策问题中，那些可供选择的方案就称之为行动方案，简称**方案或策略**，有时也称为**方案变量或决策变量**。

状态概率:指在决策问题中，每一种自然状态出现的概率。

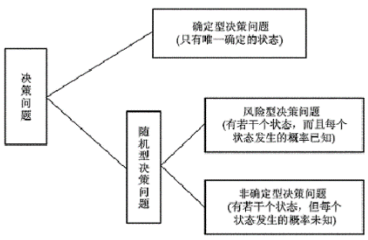
益损值:指每一种行动方案在各种**自然状态**（客观条件约束）下所获得的**报酬**或者需要付出的**损失**（成本、代价）。

最佳决策方案:就是依照某种决策准则，使决策目标取**最优值**（譬如，收益最大值或者成本最小值）的那个（些）行动方案。

（1）确定型决策问题:指决策者已经**完全确切地知道**将发生什么样的自然状态，从而可以在既定的状态下选择**最佳行动方案**。也就是说，对于确定型决策问题而言，只存在一个**唯一确定的自然状态**。对于确定型决策问题，在实际工作中，决策者所面临的方案数目可能是很大的，最佳决策方案的选择往往需要采用**各种规划方法**（如**线性规划**、**目标规划**等）才能实现。

（2）随机型决策问题:指决策者所面临的各种自然状态（客观条件）将是随机出现的。 随机型决策问题，必须具备以下几个条件：

① 存在着决策者希望达到的**明确目标**；
② 存在着不依决策者的主观意志转移的两个以上的**自然状态**
③ 存在着**两个以上**的可供选择的**行动方案**；
④ 不同行动方案在不同自然状态下的**益损值**可以计算出来。



随机型决策问题可进一步分为**风险型决策问题**和**非确定型决策**问题。

风险型决策问题：每一种**自然状态发生的概率**是已知的或者可以预先估计的。

非确定型决策问题：各种自然状态发生的概率也是未知的和无法预先估计的。

风险型决策方法

①**最大可能法**：选择一个**概率最大的自然状态**，把它看成是将要发生的**唯一确定的状态**，而把其他概率较小的自然状态忽略。即通过**比较**各行动方案在那个**最大概率**的自然状态下的益损值进行决策。("将大概率事件看成必然事件，小概率事件看成不可能事件"的假设条件下)

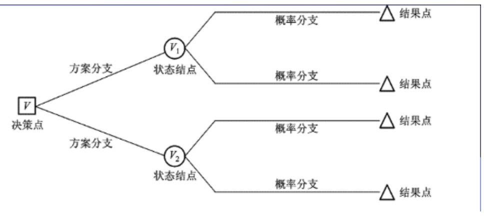
②期望值决策法：计算**各方案的期望益损值**，并以它为依据，选择**平均收益最大**或者**平均损失最小的方案**作为最佳决策方案。对于一个离散型的随机变量 *X*，随机变量 *x* 的期望值代表了它在**概率意义下的平均值**。矩阵运算

（1）把每一个**行动方案**看成是一个**随机变量**，而它在不同自然状态下的**益损值**就是该随机变量的取值；(2) 把每一个行动方案在不同的自然状态下的**益损值**与其对应的**状态概率相乘**，再相加，计算该行动方案在概率意义下的**平均益损值**；(3) 选择平均收益最大或平均损失最小的行动方案作为最佳决策方案。

③树型决策法：树型决策法,是研究风险型决策问题经常采取的决策方法。

决策树,是树型决策法的基本结构模型,它由**决策点（□）、方案分枝、状态结点（○）、概率分枝**和**结果点（△）**等要素构成。

决策原则：树型决策法的决策依据是各个方案的**期望益损值**，决策的原则一般是选择期望收益值最大或期望损失（成本或代价）值最小的方案作为最佳决策方案。



逻辑顺序：树根→树杆→树枝，最后向树梢逐渐展开。各个方案的期望值的计算过程恰好与分析问题的逻辑顺序相反，它一般是从每一个树梢开始，经树枝、树杆、逐渐向树根进行。

(1)**单级风险型决策：**在整个决策过程中，只需做出一次决策方案的选择，就可完成决策任务。

(2)**多级风险型决策：**在整个决策过程中，需要做出多次决策方案的选择，才能完成决策任务。

步骤:（1）画出决策树。把一个具体的决策问题，由决策点逐渐展开为方案分支、状态结点，以及概率分支、结果点等。**（2）计算期望益损值。**在决策树中，由树梢开始，经树枝、树杆、逐渐向树根，依次计算各个方案的期望益损值。 **（3）剪枝。**将各个方案的期望益损值分别标注在其对应的状态结点上，进行比较优选，将优胜者填入决策点，用"||"号剪掉舍弃方案，保留被选取的最优方案。

④灵敏度分析法：对于风险型决策问题，其各个方案的期望益损值是在对状态概率预测的基础上求得的。由于**状态概率的预测会受到许多不可控因素的影响**，因而基于状态概率预测结果的期望益损值也不可能同实际完全一致，会产生一定的误差。这样，就必须对可能产生的**数据变动是否会影响最佳决策方案**的选择进行分析，这就是**灵敏度分析**。

(1)以**最大期望效益值**为准则确定最佳方案(2)**灵敏度分析**:最佳概率变化时，求两方案期望效益值相等的转移概率 P，当 P 不同取值时取不同方案

⑤效用分析法：面对同一决策问题，不同的决策者对相同的利益和损失的反应不同。即便是对于相同的决策者，在不同的时期和情况下，这种反应也不相同。这就是决策者的主观价值概念，即**效用值**概念。

（1）**画出效用曲线：**<u>以益损值为横坐标，以效用值为纵坐标。</u>规定：益损值的最大效用值为 1，益损值的最小效用值为 0（2）**按效用值进行决策：**① 找出每一行动方案在不同状态下的益损值的效用值；② 计算各个行动方案的期望效用值；③ 选择**期望效用值最大**的方案。保守型决策者、风险型决策者、中间型决策者的效用曲线。

非确定型决策方法：对于非确定型决策问题，不但状态的发生是随机的，而且各**状态发生的概率也是未知**的和无法事先确定的。对于这类问题的决策，主要取决于决策者的素质、经验和决策风格等，没有一个完全固定的模式可循，对于同一个决策问题，不同的决策者可能会采用不同的处理方法。

①**乐观法**:**最大最大准则法**，其决策原则是“**大中取大**”。乐观法的特点是，决策者持最乐观的态度，决策时不放弃任何一个获得最好结果的机会，愿意以承担一定风险的代价去获得最大的利益。（1）计算每一个方案在各状态下的最大收益值（2）计算各方案在各状态下的最大效益值的最大值

②**悲观法**:**最大最小准则**法或瓦尔德（Wold Becisia）准则法，其决策原则是**“小中取大”**。特点是决策者持最**悲观**的态度，他总是把**事情估计**得**很不利**。(1)计算每一个方案在各状态下的最小效益值 (2)计算各方案在各状态下的最小效益值的最大值

③**折衷法:**特点是既不非常乐观，也不非常悲观，而是通过一个**系数** $\boldsymbol{\alpha}$（$0 \leq \alpha \leq 1$）表示决策者对客观条件估计的乐观程度。(1)计算每一个方案在各状态下的最大效益值和最小效益值 (2)计算每一个方案的折衷效益值 $V=\alpha V_{max}+(1-\alpha)V_{min}$ (3)计算各方案的折衷效益值的最大值

④**等可能性法:** (1)假设各个状态发生的概率相等，即 $P_1 = P_2 = \cdots = P_n = \cdots$；(2)计算各个方案的期望益损值，通过比较各个方案的期望益损值，选择最佳决策方案。

⑤**后悔值法:最小最大后悔值法**。对于一个实际的非确定型决策问题，当某一状态出现后，就能很容易地知道哪个方案的效益最大或损失最小。如果决策者在决策后感到后悔，遗憾当时没有选准效益最大或损失最小的方案。为了避免事后遗憾太大，可以采用后悔值法进行决策。**后悔值**指某状态下的最大效益值与各方案的效益值之差。

（1）计算每一个状态下各方案的**最大效益值**（2）对于**每一个状态下**的各方案，计算其**后悔值**取其**最大后悔值**（4）计算各方案的**最大后悔值的最小值**