

计量地理学复习纲要及相关概念

第1章 绪论 (基本不考)

计量地理学: 又称地理学或理论地理学, 是用数学方法和计算机研究地理现象及地理要素的科学, 是应用地理学的分支, 是数学与地理学相交叉的学科。

第2章 地理数据及其采集与预处理 (考察概念、计算题步骤。所有公式不用记)

章节概要:

地理数据的类型

地理数据的基本特征

地理数据的采集与处理

地理数据的统计处理

地理数据分布的集中化与均衡度指数

第1节 地理数据的类型

1、地理数据的类型 √：空间数据、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。

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。

3 空间数据的表达: 点、线、面。

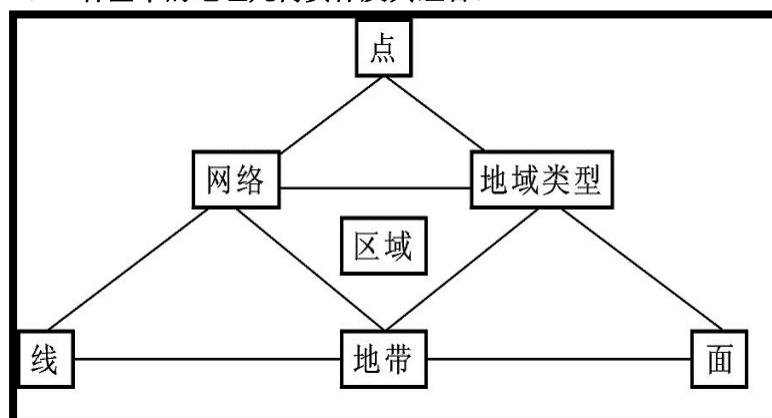
①点: 由一个独立的坐标点 (x, y) 定位, 是空间上不可再分的几何实体。

②线: 由若干个 (至少两个, 理论上是无穷个) 坐标点 (x_i, y_i) ($i=1, 2, \dots$) 定义, 有一定的长度和走向, 表示线状地物或点实体之间的联系。

③面: 表示在空间上连续分布的地理景观或区域。

④点、线、面之间的拓扑关系。

4、三种基本的地理几何实体及其组合:



5、属性数据的类型: 数量标志数据、品质标志数据

6、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。

7、品质标志数据:

①有序数据: 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 两个数据表示地理事物、地理现象或地理事件的是非判断问题。

③名义尺度数据: 用数字表示地理实体、地理要素、地理现象或地理事件的状态类型。

第3节 地理数据的采集与处理

地理数据的采集

地理数据处理

1、地理数据的渠道来源 √

- ①来自于观测、测量部门的有关专业数据。
- ②来自于统计年鉴、统计公报中的有关自然资源及社会经济发展数据。
- ③来自于有关单位或个人的不定期的典型调查数据、抽样调查数据。
- ④来自于政府公报、政府文件中的有关数据。
- ⑤来自于档案、图书等文献资料中的有关数据。
- ⑥来自于互联网（internet）的有关共享数据。
- ⑦来自遥感数据。主要包括各种航空遥感数据和卫星遥感数据。
- ⑧其他来源的有关数据。

2、采集地理数据过程中需要注意的问题

- ①数据的完备性和可靠性。
- ②在数据采集过程中，最大限度地减小数据的误差。
- ③在数据采集完毕后，进行检验，辨别真伪，通过数据筛选，去粗取精、去伪存真。

第4节 地理数据的统计处理

统计整理

几种常用的统计指标与参数

应用实例：中国大陆省份人均GDP的变异系数

1、统计整理的基本步骤：

一、求极差 $R = X_{\max} - X_{\min}$

二、确定组数 n : 组数的多少是根据样本容量的多少确定的。

一般说，样本容量大，组数可多一些，反之就少。 $n=1+3.32\lg N$

三、计算组距 h : 即一组两头分界点距离 $h=R/n$

四、计算组限 y_i : 即各组分界点的数值。

第一组下限 $y_1=X_{\min}$ $y_1=X_{\min}-1/2h$

第 i 组上限 $y_i=y_1+i*h$

五、统计频数 m : 某组频数就是落在该组内的数据个数

六、计算频率 f_i 和累积频率 F_i

$f_i=n_i/m$ F_i =从第一组至该组的频率之和。

七、分析数据分布的特征

2、描述地理数据一般水平的指标：

①平均值：反映了地理数据一般水平。

②中位数：从一个侧面反映地理数据的一般水平

③众数：出现频率最多的数，从一个侧面反映地理数据的一般水平

3、描述地理数据分布离散程度的指标 √

①极差：指所有数据中最大值与最小值之差

②离差：指每一个地理数据与平均值的差

③离差平方和：它从总体上衡量一组地理数据与平均值的离散程度

④变异系数：表示地理数据的相对变化（波动）程度

⑤方差：从平均概况衡量一组地理数据与平均值的离散程度

⑥标准差：方差的平方根

4、描述地理数据分布特征的参数

偏度系数：测度地理数据分布的不对称性情况，刻画以平均值为中心的偏向情况

峰度系数：测度了地理数据在均值附近的集中程度

第4章 地理学中的经典统计分析方法 (数学方法概念, 区别联系, 解决问题)

章节概要：

相关分析

回归分析

时间序列分析

系统聚类分析

主成分分析

趋势面分析方法

马尔可夫预测方法

第1节 相关分析 (不考计算)

1、两要素之间相关程度的测定

2、多要素间相关程度的测定

一、两要素之间相关程度的测定

1、相关系数的计算与检验

2、秩相关系数的计算与检验

(一) 相关系数的计算与检验

1、相关分析：相关分析的任务，是揭示地理要素之间相互关系的密切程度。而地理要素之间相互关系密切程度的测定，主要是通过对相关系数的计算与检验来完成的。

2、相关系数：是根据要素之间的样本值计算出来，它随着样本数的多少或取样方式的不同而不同，因此它只是要素之间的样本相关系数，只有通过检验，才能知道它的可信度。

3、相关系数的检验：是通过在给定的置信水平下，查相关系数检验的临界值表来实现的。

(二) 秩相关系数的计算与检验

秩相关系数：又称等级相关系数，或顺序相关系数，是将两要素的样本值按数据的大小顺序排列位次，以各要素样本值的位次代替实际数据而求得的一种统计量。

二、多要素间相关程度的测定

偏相关系数的计算与检验

复相关系数的计算与检验

(一) 偏相关系数的计算与检验

1、偏相关系数的计算

定义：在多要素所构成的地理系统中，先不考虑其他要素的影响，而单独研究两个要素之间的相互关系的密切程度，这称为偏相关。用以度量偏相关程度的统计量，称为偏相关系数。

2、偏相关系数的性质

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，

即 $R_{1 \cdot 23} \geq |r_{12 \cdot 3}|$ 。

3、偏相关系数的显著性检验→ t 检验法

(二) 复相关系数的计算与检验

1、**复相关系数**: 反映几个要素与某一个要素之间的复相关程度。

2、**复相关系数的性质**

① 复相关系数介于 0 到 1 之间

② 复相关系数越大，则表明要素（变量）之间的相关程度越密切。复相关系数为 1，表示完全相关；复相关系数为 0，表示完全无关。

③ 复相关系数必大于或至少等于单相关系数的绝对值。

3、**复相关系数的显著性检验** → F 检验法。

第 2 节 回归分析 (考察公式的推导，证明题，文字解释清晰，不考计算)

一元线性回归模型

多元线性回归模型

非线性回归模型

回归分析: 建立反映地理要素之间具体数量关系的数学模型，揭示要素之间的相关程度，是研究要素之间具体数量关系的一种强有力的工具。

一、一元线性回归模型

(一) 参数 a、b 的最小二乘估计

(二) 一元线性回归模型的显著性检验 F 检验法。

二、多元线性回归模型

回归模型的建立

回归模型的显著性检验

三、非线性回归模型 (非→线)

第 3 节 时间序列分析 (概念 √、平滑处理 √、数据预处理 √)

时间序列分析的基本原理

趋势拟合方法

季节变动预测

时间序列: 也叫时间数列或动态数列，是要素（变量）的数据按照时间顺序变动排列而形成的一种数列，反映了要素（变量）随时间变化的过程。

地理过程的时间序列分析: 通过地理要素（变量）随时间变化的历史过程，揭示其发展变化规律，并对其未来状态进行预测。

一、时间序列分析的基本原理

(一) 时间序列的组合成份

①**长期趋势 (T)** 是指时间序列随时间的变化而逐渐增加或减少的长期变化的趋势。

②**季节变动 (S)** 是指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。

③**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 (business cycle movement)

④**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。

(二) 时间序列的组合模型 1、加法模型 2、乘法模型

二、趋势拟合方法 1、平滑法 2、趋势线法 3、自回归模型

(一) 平滑法

1、移动平均法

2、滑动平均法

3、指数平滑法

①一次指数平滑

②高次指数平滑法

(二) 趋势线法

- 1、直线型趋势线
- 2、指数型趋势线
- 3、抛物线型趋势线

(三) 自回归模型

- 1、自相关性判断
- 2、自回归模型的建立

常见的线性自回归模型:

- ① 一阶线性自回归预测模型
- ② 二阶线性自回归预测模型
- ③ p 阶线性自回归模型

三、季节性预测法 (书 P115)

基本步骤

第5节 系统聚类分析 (概念+计算 (①原始数→标准化 ②以标准化))

聚类要素的数据处理

距离的计算

直接聚类法

最短距离聚类法 ✓

最远距离聚类法 ✓

系统聚类法计算类之间距离的统一公式

系统聚类分析实例

一、聚类要素的数据处理

常用的聚类要素的数据处理方法

- ① **总和标准化:** 分别求出各聚类要素所对应的数据的总和, 以各要素的数据除以该要素的数据的总和
- ② **标准差标准化** (由这种标准化方法所得到的新数据, 各要素的平均值为 0, 标准差为 1)
- ③ **极大值标准化** (经过这种标准化所得的新数据, 各要素的极大值为 1, 其余各数值小于 1)
- ④ **极差的标准化** (经过这种标准化所得的新数据, 各要素的极大值为 1, 极小值为 0, 其余的数值均在 0 与 1 之间。)

二、距离的计算

常见的距离:

- ① 绝对值距离
- ② 欧氏距离
- ③ 明科夫斯基距离
- ④ 切比雪夫距离

三、直接聚类法

原理: 先把各个分类对象单独视为一类, 然后根据距离最小的原则, 依次选出一对分类对象, 并成新类。如果其中一个分类对象已归于一类, 则把另一个也归入该类; 如果一对分类对象正好属于已归的两类, 则把这两类并为一类。每一次归并, 都划去该对象所在的列与列序相同的行。经过 $m-1$ 次就可以把全部分类对象归为一类, 这样就可以根据归并的先后顺序作出聚类谱系图。

四、最短距离聚类法

最短距离聚类法：是在原来的 $m \times m$ 距离矩阵的非对角元素中找出 $d_{pq} = \min\{d_{ij}\}$ ，把分类

$$d_{rk} = \min\{d_{pk}, d_{qk}\} \quad (k \neq p, q)$$

对象 G_p 和 G_q 归并为一新类 G_r ，然后按计算公式计算原来各类与新类之间的距离，这样就得到一个新的 $(m-1)$ 阶的距离矩阵；再从新的距离矩阵中选出最小者 d_{ij} ，把 G_i 和 G_j 归并成新类；再计算各类与新类的距离，这样一直下去，直至各分类对象被归为一类为止。

五、最远距离聚类法

最远距离聚类法与最短距离聚类法的区别在于计算原来的类与新类距离时采用的公式不同。

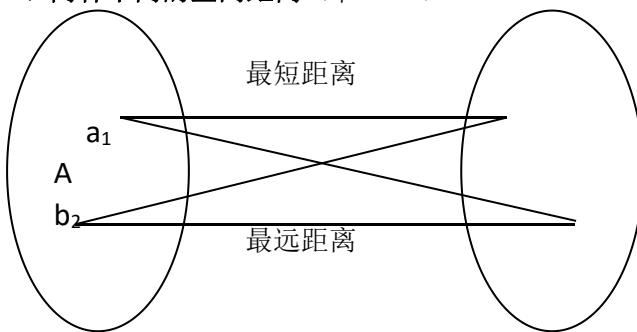
$$d_{rk} = \max\{d_{pk}, d_{qk}\} \quad (k \neq p, q)$$

六、计算类之间距离的统一公式

1、最短距离和最远距离，可以用一个公式表示：

$$d_{kr}^2 = \alpha_p d_{pk}^2 + \alpha_q d_{qk}^2 + \gamma |d_{pk}^2 - d_{qk}^2|$$

2、两种不同的空间距离（书 P133）



3、系统聚类其他方法的公式

$$d_{kr}^2 = \alpha_p d_{kp}^2 + \alpha_q d_{kq}^2 + \beta d_{pq}^2 + \gamma |d_{kp}^2 - d_{kq}^2|$$

七、实例分析

第6节 主成分分析（不考计算，考察步骤，概念，结果分析）

主成分分析的基本原理

主成分分析的计算步骤

主成分分析方法应用实例

主成分分析：是把原来多个变量划为少数几个综合指标的一种统计分析方法。从数学角度来看，这是一种降维处理技术。

一、主成分分析的基本原理

二、主成分分析的计算步骤

（一）计算相关系数矩阵

（二）计算特征值与特征向量

（三）计算特征值与特征向量

第7节 趋势面分析方法（原理、概念、三个模型、计算不一定会考察）

趋势面分析的一般原理

趋势面模型的适度检验

趋势面分析应用实例

一、趋势面分析的一般原理

1、趋势面分析：是利用数学曲面模拟地理系统要素在空间上的分布及变化趋势的一种数学方法。它实质上是通过回归分析原理，运用最小二乘法拟合一个二维非线性函数，模拟地理要素在空间上的分布规律，展示地理要素在地域空间上的变化趋势。

2、趋势面分析方法：常常被用来模拟资源、环境、人口及经济要素在空间上的分布规律，它在空间分析方面具有重要的应用价值。

3、趋势面：是一种抽象的数学曲面，它抽象并过滤掉了一些局域随机因素的影响，使地理要素的空间分布规律明显化。

4、实际的地理曲面：

①趋势面：反映地理要素的宏观分布规律，属于确定性因素作用的结果；

②剩余面：对应于微观局域，是随机因素影响的结果。

5、趋势面分析基本要求：所选择的趋势面模型应该是剩余值最小，而趋势值最大，这样拟合度精度才能达到足够的准确性。空间趋势面分析，正是从地理要素分布的实际数据中分解出趋势值和剩余值，从而揭示地理要素空间分布的趋势与规律。

(一) 建立趋势面模型

1、趋势面分析的核心

从实际观测值出发推算趋势面，一般采用回归分析方法，使得残差平方和趋于最小

2、计算趋势面的数学方程式：

①多项式函数（最为常用，因为任何一个函数都可以在一个适当的范围内用多项式来逼近，而且调整多项式的次数，可使所求的回归方程适合实际问题的需要。）

②傅立叶级数

2、多项式趋势面的形式

$$\textcircled{1} \text{ 一次趋势面模型 } z = a_0 + a_1x + a_2y$$

$$\textcircled{2} \text{ 二次趋势面模型 } z = a_0 + a_1x + a_2y + a_3x^2 + a_4xy + a_5y^2$$

$$\textcircled{3} \text{ 三次趋势面模型 } z = a_0 + a_1x + a_2y + a_3x^2 + a_4xy + a_5y^2 + a_6x^3 + a_7x^2y + a_8xy^2 + a_9y^3$$

(二) 估计趋势面模型的参数

实质：根据观测值 z_i, x_i, y_i ($i=1, 2, \dots, n$) 确定多项式的系数 a_0, a_1, \dots, a_p ，使残差平方和最小。

过程：（书 P140）

二、趋势面模型的适度检验

1、检验方法（书 P142）

①趋势面拟合适度的 R² 检验

②趋势面拟合适度的显著性 F 检验

③趋势面适度的逐次检验

(一) 趋势面拟合适度的 R² 检验

趋势面与实际面的拟合度系数 R² 是测定回归模型拟合优度的重要指标。

一般用变量 z 的总离差平方和中回归平方和所占的比重表示回归模型的拟合优度。

总离差平方和等于回归平方和与剩余平方和之和

(二) 趋势面拟合度的显著性 F 检验

1、趋势面适度的 F 检验: 对趋势面回归模型整体的显著性检验。

2 方法: 是利用变量 z 的总离差平方和中剩余平方和与回归平方和的比值, 确定变量 z 与自变量 x、y 之间的回归关系是否显著

3 结果分析: 在显著性水平 α 下, 查 F 分布表得 F_α , 若计算的 F 值大于临界值 F_α , 则认为趋势面方程显著; 反之则不显著。

(三) 趋势面适度的逐次检验

1、方法:

(1)求出较高次多项式方程的回归平方和与较低次多项式方程的回归平方和之差;

(2)将此差除以回归平方和的自由度之差, 得出由于多项式次数增高所产生的回归均方差;

(3)将此均方差除以较高次多项式的剩余均方差, 得出相继两个阶次趋势面模型的适度性比较检验值 F。

若所得的 F 值是显著的, 则较高次多项式对回归作出了新贡献, 若 F 值不显著, 则较高次多项式对于回归并无新贡献。相应的方差分析表见表 3.6.1。

2、注意: 在实际应用中, 往往用次数低的趋势面逼近变化比较小的地理要素数据, 用次数高的趋势面逼近起伏变化比较复杂的地理要素数据。次数低的趋势面使用起来比较方便, 但具体到某点拟合较差; 次数较高的趋势面只在观测点附近效果较好, 而在外推和内插时则效果较差。

三、趋势面分析应用实例 (书 P143)

1、解题步骤

2、模型检验

第 8 节 马尔可夫预测方法 (概念、思想、计算)

几个基本概念

马尔可夫预测法

马尔可夫 (Markov) 预测方法, 一种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。

一、几个基本概念

1、状态: 指某一事件在某个时刻 (或时期) 出现的某种结果。

2、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。

3、马尔可夫过程: 在事件的发展过程中, 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, 则这样的状态转移过程就称为马尔可夫过程。

4、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

5、条件概率: 由状态 E_i 转为状态 E_j 的状态转移概率 $P(E_i \rightarrow E_j)$ 就是条件概率

$$P(E_j / E_i) \text{ 即 } P(E_i \rightarrow E_j) = P(E_j / E_i) = P_{ij}$$

6、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 E_1, E_2, \dots, E_n 。

记为从状态 E_i 转变为状态 E_j 的状态转移概率 $P(E_i \rightarrow E_j)$, 则矩阵

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ \vdots & \vdots & & \vdots \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{bmatrix}$$

称为状态转移概率矩阵。

7、概率矩阵

$$\begin{cases} 0 \leq P_{ij} \leq 1 & (i, j = 1, 2, \dots, n) \\ \sum_{j=1}^n P_{ij} = 1 & (i = 1, 2, \dots, n) \end{cases}$$

将满足上式条件的任何矩阵都称为随机矩阵，或概率矩阵。

(不难证明，如果 P 为概率矩阵，则对于任何整数 $m > 0$ ，矩阵都是概率矩阵。)

8、标准概率矩阵：如果 P 为概率矩阵，而且存在整数 $m > 0$ ，使得概率矩阵 P^m 中诸元素皆非零，则称 P 为标准概率矩阵。

9、平衡向量：如果 P 为标准概率矩阵，则存在非零向量 $\alpha = [x_1, x_2, \dots, x_n]$ ，而且 x_i 满足

$0 \leq x_i \leq 1, \sum_{i=1}^n x_i = 1$ 使得 $\alpha P = \alpha$ 这样的向量 α 称为平衡向量，或终极向量。这就是说，标准概率矩阵一定存在平衡向量。

10、状态转移概率矩阵的计算

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：

$P_{ij} (i, j = 1, 2, \dots, n)$ 为了求出每一个，一般采用频率近似概率的思想进行计算。

二、马尔可夫预测方法

1、状态概率 $\pi_j(k)$ ：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 E_j 的概率。且 $\sum_{j=1}^n \pi_j(k) = 1$

2、第 k 个时刻 (时期) 的状态概率预测 (书 P149)

3、终极状态概率预测

①定义：经过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，即

$$\pi = [\lim_{k \rightarrow \infty} \pi_1(k), \lim_{k \rightarrow \infty} \pi_2(k), \dots, \lim_{k \rightarrow \infty} \pi_n(k)] = \lim_{k \rightarrow \infty} \pi(k) \quad \pi = \pi P$$

② 终极状态概率应满足的条件 $0 \leq \pi_i \leq 1 \quad (i = 1, 2, \dots, n) \quad \sum_{i=1}^n \pi_i = 1$

③ 例题

4、关键：在地理事件的预测中，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。

5、马尔可夫预测的基本方法: 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。

6、马尔可夫预测方法的基本要求(基本要求): 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。

第6章 线性规划与多目标规划 (概念、建模、图解法、考察可能性 90%)

章节概要

线性规划及其单纯形求解方法

线性规划的对偶理论

运输问题的求解方法: 表上作业法

线性规划是运筹学中发展较快、应用较广和比较成熟的一个分支。

在地理学领域, 线性规划, 作为传统的计量地理学方法之一, 是解决有关规划、决策和系统优化问题的重要手段。

第1节 线性规划及其单纯形求解方法

线性规划的数学模型

线性规划的标准形式及方法

线性规划的解及其性质

线性规划问题的求解方法——单纯形法 (不考)

应用实例: 农场种植计划模型

一、线性规划的数学模型

(一) 线性规划模型之实例

1、线性规划研究的两类问题:

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。

2、实例:

①运输问题

②资源利用问题

⑤合理下料问题

(二) 线性规划的数学模型

1、线性规划问题的特征:

①每一个问题都用一组未知变量 (x_1, x_2, \dots, x_n) 表示某一规划方案, 其一组定值代表一个具体的方案, 而且通常要求这些未知变量的取值是非负的。

②每一个问题的组成部分: 一是目标函数, 按照研究问题的不同, 常常要求目标函数取最大或最小值; 二是约束条件, 它定义了一种求解范围, 使问题的解必须在这一范围之内。

③每一个问题的目标函数和约束条件都是线性的。

2、线性规划例子

二、线性规划的标准形式及方法

(一) 线性规划的标准形式

(二) 化为标准形式的方法

1、目标函数化为标准形式的方法

2、约束方程化为标准形式的方法

三、线性规划的解及其性质

(一) 线性规划的解

1、可行解与最优解

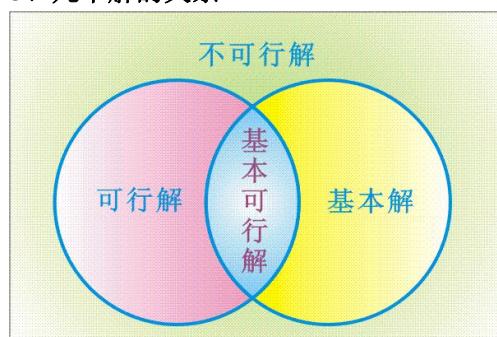
可行解：满足约束条件（即满足线性约束和非负约束）的一组变量为可行解。

可行域：所有可行解组成的集合称为可行域。

最优解：使目标函数最大或最小化的可行解称为最优解。

2、基本解与基本可行解

3、几个解的关系



(二) 线性规划解的性质

1、凸集和顶点

①**凸集：**若连接 n 维点集 S 中的任意两点 $X(1)$ 和 $X(2)$ 之间的线段仍在 S 中，则 S 为凸集。

②**顶点：**若凸集 S 中的点 $X(0)$ 不能成为 S 中任何线段的内点，则称 $X(0)$ 为 S 的顶点或极点。

2、线性规划解的性质

①线性规划问题的可行解集（可行域）为凸集。

②可行解集 S 中的点 X 是顶点的充要条件是基本可行解。

③若可行解集有界，则线性规划问题的最优值一定可以在其顶点上达到。

因此线性规划的最优解只需从其可行解集的有限个顶点中去寻找。

四、线性规划问题的求解方法——单纯形法

(一) 单纯形表

1、最优解的判定

2、最优解的判定定理

(二) 单纯形法的计算步骤

五、应用实例：农场所种植计划模型

第 10 章 地理网络分析 (概念 基本思想)

章节概要：

地理网络的图论描述

最短路径与选址问题 ✓

最大流与最小费用流 ✓

网络分析是运筹学的一个重要分支，它主要运用图论方法研究各类网络的结构及其优化问题。

第 1 节 地理网络的图论描述

地理网络的图论描述

地理网络的测度

1、通俗意义上的“图”，主要是指各种各样的地图、遥感影像图，或者是由各种符号、文字代表的示意图，或者是由各种地理数据绘制而成的曲线图、直方图等等。

2、图论中的“图”，是一个数学概念，这种“图”能从数学本质上揭示地理实体与地理事物空间分布格局，地理要素之间的相互联系以及它们在地域空间上的运动形式、地理事件发生的先后顺序等。

一、地理网络的图论描述

(一) 图的定义

(1) 图：设 V 是一个由 n 个点 v_i ($i=1, 2, \dots, n$) 所组成的集合，即 $V=\{v_1, v_2, \dots, v_n\}$ ，
 E 是一个由 m 条线 e_i ($i=1, 2, \dots, m$) 所组成的集合，即 $E=\{e_1, e_2, \dots, e_m\}$ ，而且 E 中任意一条线，都是以 V 中的点为端点；任意两条线除了端点外没有其他的公共点。

那么，把 V 与 E 结合在一起就构成了一个图 G ，记作 $G=(V, E)$ 。

(2) 顶点： V 中的每一个点 v_i ($i=1, 2, \dots, n$) 称为图 G 的顶点。

(3) 边： E 中每一条线称为图 G 的边（或弧）；若一条边 e 连接 u, v 两个顶点，则记为 $e=(u, v)$ 。

(4) 在图 $G=(V, E)$ 中， V 不允许是空集，但 E 可以是空集。

(5) 图的基本要素：①点集（或称顶点集）；②边集（或称弧集）。

(6) 在现实地理系统中

图：对于地理位置、地理实体、地理区域以及它们之间的相互联系，可以经过一定的简化与抽象，将它们描述为图论意义上的地理网络，即图。

点：地理位置、地理实体、地理区域，譬如：山顶、河流汇聚点、车站、码头、村庄、城镇等——点。

点与点的连线：它们之间的相互联系，譬如，构造线、河流、交通线、供电与通讯线路、人口流、物质流、资金流、信息流、技术流等——点与点的连线。

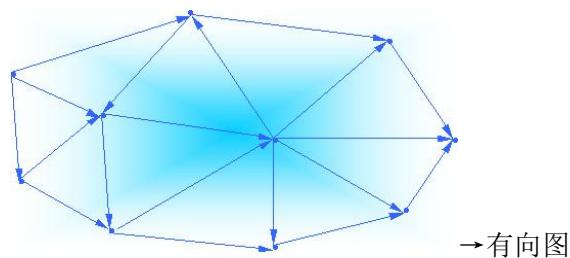
一个由基本流域单元组成的复杂的流域地貌系统，如果舍弃各种复杂的地貌形态，各条河流——线，河流分岔或汇聚处——点，流域地貌系统——水系的基本结局（树）。

(二) 图的一些相关概念

(1) 无向图与有向图：

1、**无向图**——图的每条边都没有给定方向，即 $(u, v) = (v, u)$ ；

2、**有向图**——图的每条边都给定了方向，即 $(u, v) \neq (v, u)$ 。



一般将有向图的边集记为 A ，无向图的边集记为 E 。这样， $G=(V, A)$ 就表示有向图，而 $G=(V, E)$ 则表示无向图。

(2) 赋权图：如果图 $G=(V, E)$ 中的每一条边 (v_i, v_j) 都相应地赋有一个数值 w_{ij} ，则称 G 为赋权图，其中 w_{ij} 称为边 (v_i, v_j) 的权值。

除了可以给图的边赋权外，也可以给图的顶点赋权。这就是说，对于图 G 中的每一顶点 v_j ，也可以赋予一个载荷 $a(v_j)$ 。

(3) 关联边：若 $e=(u, v)$ ，则称 u 和 v 是边 e 的端点， e 是 u 和 v 的关联边。

(4) 环：若 e 的两个端点相同，即 $u=v$ ，则称为环。

- (5) **多重边**: 若连接两个端点的边多于一条以上，则称为多重边。
- (6) **多重图**: 含有多重边的图，称为多重图。
- (7) **简单图**: 无环、无多重边的图，称为简单图。
- (8) **点与次**:
- 1、**次**: 以点 v 为端点的边的个数称为点 v 的次，记为 $d(v)$ 。
 - 2、**悬挂点**: 次等于 1 的点称为悬挂点；
 - 3、**悬挂边**: 与悬挂点关联的边称为悬挂边。
 - 4、**孤立点**: 次为零的点称为孤立点。
 - 5、**奇点**: 次为奇数的点称为奇点；
 - 6、**偶点**: 次为偶数的点称为偶点。
- (9) **连通图**: 在图 G 中，若任何两点之间至少存在一条路（对于有向图，则不考虑边的方向），则称 G 为连通图，否则称为不连通图。
- (10) **路（链）**: 若图 $G = (V, E)$ 中，若顶点与边交替出现的序列（对于有向图来说，要求排在每一条边之前和之后的顶点分别是这条边的起点和终点） $P = \{v_{i1}, e_{i1}, v_{i2}, e_{i2}, \dots, e_{ik-1}, v_{ik}\}$ 满足 $e_{it} = (v_{it}, v_{t+1}) \quad (t=1, 2, \dots, k-1)$ ，则称 P 为一条从 v_{i1} 到 v_{ik} 的路（或链），简记为 $P = \{v_{i1}, v_{i2}, \dots, v_{ik}\}$ 。
- (11) **回路**: 若一条路的起点与终点相同，即 $v_{i1} = v_{ik}$ ，则称它为回路。
- (12) **树**: 不含回路的连通的无向图称为树。
- (13) **基础图**: 从一个有向图 $D = (V, A)$ 中去掉所有边上的箭头所得到的无向图，就称为 D 的基础图，记之为 $G(D)$ 。
- (14) **截**: 如果从图中移去边的一个集合将增加亚图的数目时，被移去的边的集合就称为截。
- (15) **子图**: 设 $G = (V, E)$ 是一个无向图， V_1 与 E_1 分别是 V 与 E 的子集，即 $V_1 \subseteq V$ ， $E_1 \subseteq E$ 。如果对于任意 $e_i \in E_1$ ，其两个端点都属于 V_1 ，则称 $G_1 = (V_1, E_1)$ 是图 G 的一个子图。
- (16) **支撑子图**: 设 $G_1 = (V_1, E_1)$ 是图 $G = (V, E)$ 的一个子图，如果 $V_1 = V$ ，则称 G_1 是 G 的支撑子图。
- (17) **支撑树**: 设 $G = (V, E)$ 是一个无向图，如果 $T = (V_1, E_1)$ 是 G 的支撑子图，并且 T 是树，则称 T 是 G 的一个支撑树。
- (18) **树的重量**: 一个树的所有边的权值之和称为该树的重量。
- (19) **最小支撑树**: 在一个图的所有支撑树中，重量最小的那个叫做该图的最小支撑树。
- ## 二、地理网络的测度
- 1、**地理网络的拓扑分类** (书 P280 图 10.1.5)
- 2、**平面图**: 被规定为：各连线之间不能交叉，而且每一条连线除顶点以外，不能再有其他的公共点（牛文元，1987）。
- ### (一) 关联矩阵与邻接矩阵
- 1、**关联矩阵**: 测度网络图中顶点与边的关联关系。
- 2、**邻接矩阵**: 测度网络图中各顶点之间的连通性程度。
- ### (二) 有关测度指标
- 1、**任意网络图共同的基础指标**:
- ① 连线（边或弧）数目 m ;
 - ② 结点（顶点）数目 n ;
 - ③ 网络中亚图的数目 p 。
- (由它们可以产生如下几个更为一般性的测度指标: β 指数、回路数 k 、 α 指数、 γ 指数。)

2、 β 指数：也称为线点率，是网络内每一个结点的平均连线数目

3、 α 指数： α 指数指实际回路数与网络内可能存在的最大回路数之间的比率

网络内可能存在的最大回路数目为连线的最大可能数目减去最低限度连接的连线数目

4、回路数 k

①回路是一种闭合路径,它的始点同时也是终点。

若网络内存在回路，则连线的数目就必须超过 $n-p$ (最低限度连接网络的连接数目)。

②回路数 k 为实际连线数目减去最低限度连接的连线数目

5、 γ 指数：指网络内连线的实际数目与连线可能存在的最大数目之间的比率。

γ 指数是测度网络连通性的一种指标，其数值变化范围为[0, 1]。

$\gamma = 0$, 表示网络内无连线，只有孤立点存在；

$\gamma = 1$, 则表示网络内每一个结点都存在与其他所有结点相连的连线。

第2节 最短路径与选址问题 ✓ (考察计算、标号法基本思想、原理)

最短路径问题

选址问题

一、最短路径问题

(一) 最短路径的含义

1、“纯距离”意义上的最短路径

例如，需要运送一批物资从一个城市到另一个城市，选择什么样的运输路线距离最短？

2、“经济距离”意义上的最短路径

例如，某公司在 10 大港口 C_1, C_2, \dots, C_{10} 设有货栈，从 C_i 到 C_j 之间的直接航运价格，是由市场动态决定的。如果两个港口之间无直接通航路线，则通过第三个港口转运。那么，各个港口之间最廉价的货运线路是什么？

3、“时间”意义上的最短路径

例如，某家经营公司有一批货物急需从一个城市运往另一个城市，那么，在由公路、铁路、河流航运、航空运输等 4 种运输方式和各个运输线路所构成的交通网络中，究竟选择怎样的运输路线最节省时间？

(二) 最短路径的算法

1、标号法：1959 年 E.W.Dijkstar 提出的标号法是最短路径问题最好的求解方法。

2、标号法优点：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。

3、标号法的基本思想

①设 G 是一个赋权有向图，即对于图中的每一条边，都赋予了一个权值。在图 G 中指定两个顶点，确定为起点和终点，不妨设 v_1 为起点， v_k 为终点。

②首先从 v_1 开始，给每一个顶点标一个数，称为标号。这些标号，又进一步区分为 T 标号和 P 标号两种类型。其中，每一个顶点的 T 标号表示从起点 v_1 到该点的最短路径长度的上界，这种标号为临时标号； P 标号表示从 v_1 到该点的最短路径长度，这种标号为固定标号。

③在最短路径计算过程中，对于已经得到 P 标号的顶点，不再改变其标号；对于凡是没有标上 P 标号的顶点，先给它一个 T 标号；算法的每一步就是把顶点的 T 标号逐步修改，将其变为 P 标号。

④那么，最多经过 $k-1$ 步，就可以求得到从起点 v_1 到每一个顶点的最短路径及其长度。

4、标号法具体计算步骤 (书 P284)

二、选址问题

1、选址问题的数学模型取决于两个方面的条件：

- ①可供选址的范围、条件
- ②怎样判定选址的质量

2 选址问题的分类

根据选址的质量判据，可分为①中心点选址问题 ②中位点选址问题

(一) 中心点选址问题

1、中心点选址问题的质量判据：使最佳选址位置所在的顶点的最大服务距离为最小。

2 中心点选址问题适用范围：医院、消防站点等一类服务设施的布局问题。

3、中心点选址问题的数学描述 (书 P286)

(二) 中位点选址问题

1、中位点选址问题的质量判据：使最佳选址位置所在的顶点到网络图中其他各个顶点的最短路径距离的总和（或者以各个顶点的载荷加权求和）达到最小。

2、中位点选址问题的数学描述 (书 P287)

第3节 最大流与最小费用流

最大流问题及其求解方法

最小费用流及其求解方法

一、最大流问题及其求解方法

(一) 最大流问题

最大流问题 (考察计算)

设有向网络 $N(V, A)$ ，在发点 v_s 有一批货，要通过网络上的弧运输到收点 v_t 去，受运输条件限制，每条弧 a_{ij} 在单位时间内通过的车辆数不能超过 c_{ij} 辆，分析：如何组织运输才能使从 v_s 到 v_t 在单位时间内通过的车辆达到最多？

上面描述的这类问题，称为最大流问题。

应用：交通运输、供水、油管供油、邮电通讯，生产安排，管理优化等实际问题上。

(二) 求最大流的方法：弧标号法

(三) 求解最小费用流问题

第12章 灰色系统方法

灰色系统：既含有已知信息，又含有未知和非确定信息的系统，称为灰色系统。

第1节 灰色关联分析方法 (计算 分析)

灰色关联分析：从其思想方法上来看，属于几何处理的范畴，其实质是对反映各因素变化特征的数据序列所进行的几何比较。用于度量因素之间关联程度的灰色关联度，就是通过对因素之间的关联曲线的比较而得的。

一、灰色关联度的计算步骤如下：

1、求各序列的初值像。令 $X_i' = X_i / x_i(1) = (x_i'(1), x_i'(2), \dots, x_i'(n))$

$$I=0, 1, 2, \dots, m$$

2、求差序列。(设以 X_0 为系统特征序列)

记 $\Delta_i(k) = |x_0(k) - x_i(k)|$, $\Delta_i = (\Delta_i(1), \Delta_i(2), \dots, \Delta_i(n))$, $I=1, 2, \dots, m$

3、求两极最大差与最小差。记

$$M = \max_{i=1}^m \max_{k=1}^n \Delta_i(k), m = \min_{i=1}^m \min_{k=1}^n \Delta_i(k)$$

4、求关联系数 $r_{0i}(k) = (m + M) / (\Delta_i(k) + M)$, $k=1, 2, \dots, n; I=1, 2, \dots, m$

$$r_{0i} = \frac{1}{n} \sum_{k=1}^n r_{0i}(k); i = 1, 2, \dots, m$$

5、计算关联度

例 某城市工业、农业、运输业、商业各部门的行为数据如下：

工业： $X_1 = (X_{1(1)}, X_{1(2)}, X_{1(3)}, X_{1(4)}) = (45.8, 43.4, 42.3, 41.9)$

农业： $X_2 = (X_{2(1)}, X_{2(2)}, X_{2(3)}, X_{2(4)}) = (39.1, 41.6, 43.9, 44.9)$

运输业： $X_3 = (X_{3(1)}, X_{3(2)}, X_{3(3)}, X_{3(4)}) = (3.4, 3.3, 3.5, 3.5)$

商业： $X_4 = (X_{4(1)}, X_{4(2)}, X_{4(3)}, X_{4(4)}) = (6.7, 6.8, 5.4, 4.7)$

分别以 X_1, X_2 为系统特征序列，计算灰色关联度。

一、先以 X_1 为系统特征序列求关联度

第一步：求初值像

$$X'_i = X_i / x_i(1) = (x'_i(1), x'_i(2), \dots, x'_i(n))$$

$$X'_1 = (1, 0.9475, 0.9235, 0.9138)$$

$$X'_2 = (1, 1.063, 1.1227, 1.1483)$$

$$X'_3 = (1, 0.97, 1.0294, 1.0294)$$

$$X'_4 = (1, 1.0149, 0.805, 0.7015)$$

第二步：求差序列

$$\Delta_i(k) = |x'_i(k) - x_i(k)|; i=2,3,4$$

$$\Delta_2 = (0, 0.1155, 0.1992, 0.2335)$$

$$\Delta_3 = (0, 0.0225, 0.1059, 0.1146)$$

$$\Delta_4 = (0, 0.0674, 0.1185, 0.2123)$$

第三步：求两极差

$$M = \max_{i=1}^4 \max_{k=1}^n \Delta_i(k) = 0.2335, m = \min_{i=1}^4 \min_{k=1}^n \Delta_i(k) = 0$$

第四步：求关联系数

取 $\epsilon = 0.5$, 有

$$r_{1i}(k) = 0.11675 / (\Delta_i(k) + 0.11675); i=2,3,4; k=1,2,3,4$$

从而

$$r_{12}(1)=1, r_{12}(2)=0.503, r_{12}(3)=0.3695, r_{12}(4)=0.3333$$

$$r_{13}(1)=1, r_{13}(2)=0.8384, r_{13}(3)=0.5244, r_{13}(4)=0.504$$

$$r_{14}(1)=1, r_{14}(2)=0.634, r_{14}(3)=0.4963, r_{14}(4)=0.352$$

第五步：求灰色关联度

$$r_{12}=0.551, r_{13}=0.717, r_{14}=0.621$$

二、对于以 X_2 为系统特征的情况

$$\Delta_i(k) = |x'_i(k) - x_i(k)|; i=1,3,4$$

$$\Delta_1 = (0, 0.1155, 0.1992, 0.2335)$$

$$\Delta_3 = (0, 0.093, 0.0933, 0.1189)$$

$$\Delta_4 = (0, 0.0481, 0.3177, 0.4483)$$

$$M = \max_{i=1}^4 \max_{k=1}^n \Delta_i(k) = 0.4483, m = \min_{i=1}^4 \min_{k=1}^n \Delta_i(k) = 0$$

取 $\epsilon = 0.5$, 有

$$r_{2i}(k) = 0.22415 / (\Delta_i(k) + 0.22415); i=1,3,4; k=1,2,3,4$$

从而有

$$r_{21}(1)=1, r_{21}(2)=0.66, r_{21}(3)=0.53, r_{21}(4)=0.489$$

$$r_{23}(1)=1, r_{23}(2)=0.706, r_{23}(3)=0.706, r_{23}(4)=0.653$$

$$r_{24}(1)=1, r_{24}(2)=0.823, r_{24}(3)=0.415, r_{24}(4)=0.333$$

于是 $r_{21}=0.670$

$$r_{23}=0.766$$

$$r_{24}=0.643$$

第 2 节 灰色预测方法

基于灰色建模理论的灰色预测法，按照预测问题的特征，可分为**五种基本类型**：

- ①数列预测、
- ②变灾预测、
- ③季节变灾预测、
- ④拓扑预测
- ⑤系统综合预测

书本介绍：

- 1、数列预测（考察填空计算）
- 2、变灾预测（已建模，考察灾害预测）

（平滑处理 ✓ 构建矩阵 ✓ 求时间相应函数，系数 ✓ 形模后预测 ✓ ）

PS：计算过程越详细越好