地理系统要素间的相关分析与回归分析

——河北师范大学资环学院　　胡引翠

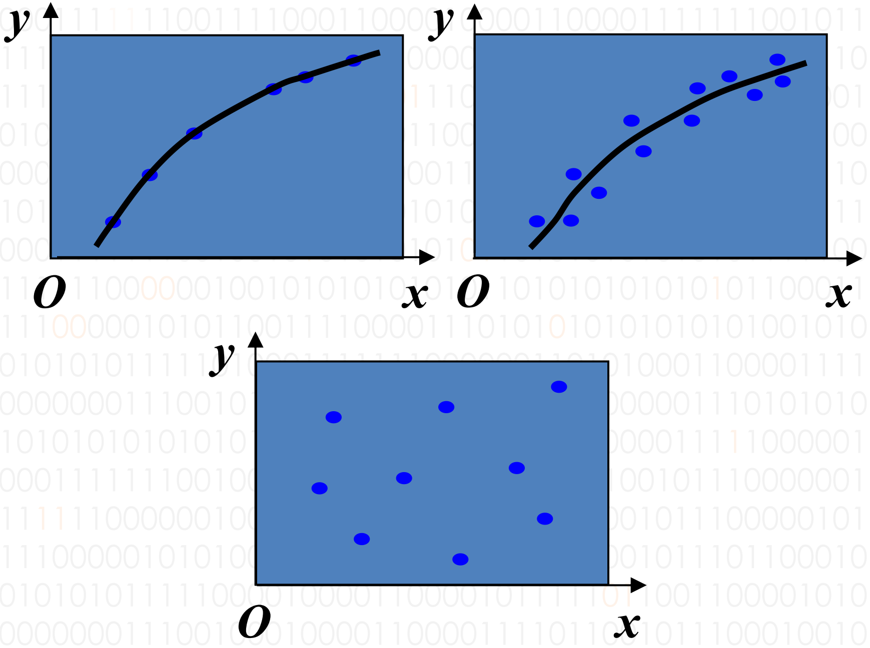
地理系统要素间的相关分析与回归分析

相关分析	回归分析
□地理相关的意义	□地理回归分析的意义和作用
□地理相关程度的度量方法	□一元地理回归模型的建立
□相关系数的显著性检验	□多元地理回归模型的建立

§1 地理要素间的相关分析

一、相关分析的意义

- 相关分析
- 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 地理要素间的相关分析

一、相关分析的意义

- 函数关系与相关关系的联系和区别：
  - 联系：
    - 在一定条件下是可以相互转化的。
  - 区别：
    - 研究分析方法不同。

§1 地理要素间的相关分析

二、地理要素间相关关系的种类

- 相关关系
  - 按变量多少
    - 两个变量
      - 直线相关
      - 曲线相关
    - 两个以上变量
      - 偏相关
      - 复相关
    - 单相关
  - 按相关方向
    - 正相关
    - 负相关
  - 按质量
    - 顺序相关
    - 品质相关

§1 地理要素间的相关分析

三、地理相关程度的测度方法

–简单线性相关程度的测度

–简单非线性相关程度的测度

–多要素相关与相关矩阵

§1 地理要素间的相关分析

三、地理相关程度的测度方法

–相关表

- 是一种显示变量之间相关关系的统计表。
- 将两个变量的对应值平行排列，且其中某一变量按其取值大小顺序排列，便可得到相关表。
- 某商店10名售货员的工龄和日工资的相关表：

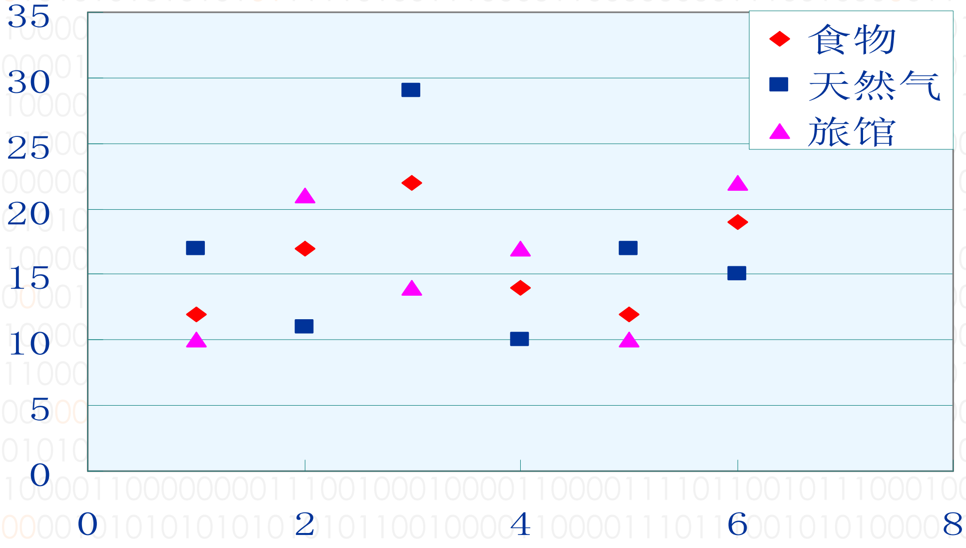
工龄（年）	4	4	5	6	7	8	8	9	9	10
日工资（元）	42	46	50	60	64	68	74	72	80	84

§1 地理要素间的相关分析

三、地理相关程度的测度方法

–相关图（散点图）

- 是将两个变量的对应值，在平面直角坐标系中用坐标点的形式描绘而成的图形。



§1 地理要素间的相关分析

三、地理相关程度的测度方法

- （一）简单线性相关程度的测度
  - 相关系数:一般相关系数r和顺序相关系数r_s

（1）一般常用的相关系数（r）

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

分子：两个变量的离均差积和；

分母：两变量离差平方和之积的平方根。

相关程度的测度方法-1、简单线性相关

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

$$=\frac{\sum x_iy_i-\frac{(\sum x_i)(\sum y_i)}{n}}{\sqrt{\left[\sum x_i^2-\frac{(\sum x_i)^2}{n}\right]\left[\sum y_i^2-\frac{(\sum y_i)^2}{n}\right]}}$$

$$=\frac{l_{xy}}{\sqrt{l_{xx}\cdot l_{yy}}}$$

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

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

$$l_{xy}=\sum x_iy_i-\frac{(\sum x_i)(\sum y_i)}{n}$$

一般常用的相关系数（r）

三、相关程度的测度方法-1、简单线性相关

- 相关系数r的性质：
  - $r \in [-1,1]$；
  - r>0时为正相关，r<0时为负相关。
  - 当|r|=1时，则r=1为完全正相关，r=-1时，为完全负相关。
  - 当r=0时，说明两变量之间完全无关。
  - 当|r|→1时，说明两变量之间关系密切；
    当|r|→0时，说明两变量之间相关程度差。

一般常用的相关系数（r）

三、相关程度的测度方法-1、简单线性相关

- 相关系数r的性质：
  - 通常认为：

$r=0$	完全不相关；
$0 < r \leqslant 0.3$	微弱相关；
$0.3 \leqslant r \leqslant 0.5$	低度相关；
$0.5 \leqslant r \leqslant 0.8$	显著相关；
$0.8 \leqslant r < 1$	高度相关；
$r=1$	完全相关。

一般常用的相关系数（r）

练一练

【例】某地区历年人均收入与商品销售额资料如下：

年份	人均收入(百元)x	商品销售额(百万元)y	xy	x^2	y^2
1998	24	11	264	576	121
1999	30	15	450	900	225
2000	32	14	448	1024	196
2001	34	16	544	1156	256
2002	38	20	760	1444	400
$\sum$	158	76	2466	5100	1198

要求计算x与y的相关系数，说明其相关方向和程度。

练一练

解：将计算表中的数值代入r计算公式得：

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

$$=\frac{5\times 2466-158\times 76}{\sqrt{5\times 5100-158^2}\sqrt{5\times 1198-76^2}}=\frac{12330-12008}{\sqrt{536\times 214}}=0.95$$

计算结果表明，人均收入与商品销售额之间存在高度的直线正相关关系。

三、相关程度的测度方法-1、简单线性相关

- （2）顺序（等级）相关系数（r_s）
  - 概念
  - 公式

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

设两个要素x和y有n对样本值，令T_1代表要素x的序号（或位次），T_2代表要素y的序号（或位次），则：

$$d_i^2 = (T_{1i} - T_{2i})^2$$

代表要素x和y的同一组样本位次差的平方。

练一练

【例】全国1999年31个省（市、区）的总人口（x）和国内生产总值（y）及其位次列于下表中。试计算x与y之间的顺序相关系数。

省（市、区）	总人口(x)及其位次		国内生产总值(y)及其位次		位次差的平方 $d_i^2=(T_{1i}-T_{2i})^2$
	人口数(万人)	位次T_1	产值(亿元)	位次T_2	
北京	1257	26	2174.46	15	121
天津	959	27	1450.06	23	16
河北	6614	6	4569.19	6	0
…	…	…	…	…	…
∑	124219	/	87671.13	/	962

练一练

解：将计算表中的数值代入r_s计算公式得:

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}$$

$$= 1 - \frac{6 \times 962}{31\ (31^2 - 1)} = 0.806$$

即：总人口（x）与国内生产总值（y）之间的等级相关系数为0.806。计算结果表明，二者存在高度的直线正相关关系。

三、相关程度的测度方法

2、简单非线性相关程度的测度

– 相关指数R

$$R = \sqrt{1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}}$$

$\hat{y}_i$ 称为回归值（或理论值），是y_i的预测值。

三、相关程度的测度方法-2、简单非线性相关

- 相关指数R的性质：
  - $R \in [0,1]$；
  - 当R=1时，两变量完全曲线相关；
  - 当R=0时，两变量完全曲线无关；
  - 当R→max时，相关程度密切，
    R→min时，相关程度差；
  - $R \geqslant |r|$；
  - $R_{xy} \neq R_{yx}$

三、相关程度的测度方法

3、多要素相关与相关矩阵

设有原始地理数据矩阵

$$
\begin{array}{cc|c}
 & & \begin{array}{cccc} 1 & 2 & \cdots & n \end{array} \quad \text{指标} \\
\text{要素} & \begin{array}{c} 1 \\ 2 \\ \vdots \\ m \end{array} & \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}
\end{array}
$$

三、相关程度的测度方法

要测度两两要素之间的相关程度，公式为：

$$r_{ij} = \frac{\sum (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum (x_{ik} - \bar{x}_i)^2 \cdot \sum (x_{jk} - \bar{x}_j)^2}}$$

得到相关系数矩阵：

$$\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mm} \end{bmatrix}$$

多要素相关矩阵的性质：

- 对角线上的元素均为1；
- 此矩阵为方阵；
- 沿对角线对称。

三、相关程度的测度方法

4、 相关系数的显著性检验

目的：判定相关系数是否有意义

三、相关程度的测度方法

4、相关系数的显著性检验

- 简单线性相关系数的显著性检验步骤
  - 计算出相关系数r。
  - 给定显著性水平α，按n-2查相关系数临界值（$r_α$）表，查出相应的临界值$r_α$。
  - 比较|r|与$r_α$的大小。
    - 当|r|≥$r_α$时，说明两变量在α水平上达到显著性，可进一步做回归分析；
    - 若|r|<$r_α$时，说明两变量在α水平上没有达到所要求的精度。

相关系数的检验

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。

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。

检验相关系数 $\rho = 0$ 的临界值（r_α）表

$$p\{|r| > r_\alpha\} = \alpha$$

f 称为自由度，其数值为 f=n-2，n 为样本数；

α 代表不同的置信水平；

f \ α	0.10	0.05	0.02	0.01	0.001
1	0.987 69	0.996 92	0.999 507	0.999 877	0.999 998
2	0.900 00	0.950 00	0.980 00	0.990 00	0.999 000
3	0.805 4	0.878 3	0.934 33	0.958 73	0.991 160
4	0.729 3	0.811 4	0.882 2	0.917 20	0.974 06
5	0.669 4	0.754 5	0.832 9	0.874 5	0.950 74
6	0.621 5	0.706 7	0.788 7	0.834 3	0.924 93
7	0.582 2	0.666 4	0.749 3	0.797 7	0.898 2
8	0.549 4	0.631 9	0.715 5	0.764 6	0.872 1
9	0.521 4	0.602 1	0.685 1	0.734 8	0.847 1
10	0.497 3	0.576 0	0.658 1	0.707 9	0.823 3
11	0.476 2	0.552 9	0.633 9	0.683 5	0.801 0
12	0.457 5	0.532 4	0.612 0	0.661 4	0.780 0

在表中，f称为自由度，其数值为$f=n-2$，n为样本数；上方的α代表不同的置信水平；表内的数值代表不同的置信水平下相关系数$\rho=0$的临界值，即r_α；公式$p\{|r|>r_\alpha\}=\alpha$的意思是当所计算的相关系数r的绝对值大于在α水平下的临界值r_α时，两要素不相关（即$\rho=0$）的可能性只有α。

相关分析实例1：

表3.1.1　伦敦的月平均气温与降水量

月份	1	2	3	4	5	6	7	8	9	10	11	12
月平均气温 $t/^{o}C$	3.8	4	5.8	8	11.3	14.4	16.5	16.2	13.8	10.8	6.7	4.7
降雨量 p/mm	77.7	51.2	60.1	54.1	55.4	56.8	45	55.3	67.5	73.3	76.6	79.6

资料来源：http://www.cwb.gov.tw/V4/climate/wta_station/wta20.htm

相关分析实例1：

(1)根据表3.1.1中的数据，我们可以利用公式（3.1.1），计算伦敦市月平均气温（t）与降水量(p)之间的相关系数

$$r_{tp}=\frac{\sum_{i=1}^{12}(t_i-\bar{t})(p_i-\bar{p})}{\sqrt{\sum_{i=1}^{12}(t_i-\bar{t})^2}\sqrt{\sum_{i=1}^{12}(p_i-\bar{p})^2}}=\frac{-300.91}{\sqrt{250.55}\sqrt{1\,508.34}}$$

$$=\frac{-300.91}{15.83\times 38.84}=-0.489\,5$$

(2)计算结果表明，伦敦市的月平均气温（t）与降水量(p)之间呈负相关，即异向相关。

相关分析实例1：

对伦敦市月平均气温（t）与降水量(p)之间的相关系数，f=12−2=10，在显著性水平$\alpha = 0.10$上，查表3. 1. 3，得知：$r_{0.10} = 0.4973$。因为$|r_{tp}| = 0.4895 < r_{\alpha} = 0.4973$，所以，伦敦市月平均气温（$t$）与降水量($p$)之间的相关性并不显著。

相关分析实例2：顺序相关系数的计算与检验

$$r'_{xy} = 1 - \frac{6\sum_{i=1}^{n} d_i^{\ 2}}{n(n^2 - 1)}$$

- 又称等级相关系数，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。

顺序相关系数检验的临界值

n	显著水平α		n	显著水平α	
	0.05	0.01		0.05	0.01
4	1.000	--	16	0.425	0.601
5	0.900	1.000	18	0.399	0.564
6	0.829	0.943	20	0.377	0.534
7	0.714	0.893	22	0.359	0.508
8	0.643	0.833	24	0.343	0.485
9	0.600	0.783	26	0.329	0.465
10	0.564	0.746	28	0.317	0.448
12	0.456	0.712	30	0.306	0.432
14	0.456	0.645	--	--	--

注：n代表样本个数，α代表不同的置信水平，也称显著水平，表中的数值为临界值 r_α。

第2节　回归分析

- 一元线性回归模型
- 多元线性回归模型
- 非线性回归模型

地理回归分析的意义和作用

- 相关分析揭示了地理要素之间相互关系的密切程度。若能在某些难测难控的要素与其他易测易控的要素之间建立一种近似的函数表达式，可以比较容易地通过那些易测易控要素的变化情况，了解那些难测难控要素的变化情况

- 回归分析方法，是研究要素之间具体的数量关系的强有力的工具，运用这种方法能够建立反映地理要素之间具体的数量关系的数学模型，即回归模型

- 回归分析
  - 就是对具有相互联系的要素，根据其联系的形态，选择一个合适的数学模式，用来近似地表达要素间平均变化关系。这个数学模式称为回归模型（回归方程）

- # 回归分析与相关分析的区别与联系
  - 研究对象和内容上:

    a.相关分析主要是研究要素（变量）之间是否存在关系和关系的密切程度，没有自变量与因变量之分

    b.回归分析主要是研究要素之间联系的形态、确定要素之间关系的方程式，即回归方程，可用于对未来进行预测，对某些要素进行控制。回归分析有自变量与因变量之分。回归分析尚有地理预测的性质

    从相关可以获得回归的一些重要信息，反之从回归也能获得相关的一些重要信息。故它们之间是紧密相连的两个概念

- 回归分析的主要内容
  - 从一组地理数据出发，确定这些要素（变量）间的定量数学表达式，即回归模型
  - 根据一个或几个要素（自变量）的值来预测或控制另一个要素（因变量）的取值
  - 从某一地理过程中的许多要素中，找出哪些要素（变量）是主要的，哪些要素是次要的，这些要素之间又有些什么关系
- 回归分析的分类
  - 一元地理回归模型和多元地理回归模型

一、一元线性回归模型

定义：假设有两个地理要素（变量）x 和y，x为自变量，y为因变量。则一元线性回归模型的基本结构形式为

$$y_\alpha = a + bx_\alpha + \varepsilon_\alpha \qquad (3.2.1)$$

式中：a和b为待定参数;$\alpha = 1,2, \cdots, n$为各组观测数据的下标;ε_a 为随机变量。

记 $\hat{a}$ 和 $\hat{b}$ 分别为参数a与b的拟合值，则一元线性回归模型为

$$\hat{y} = \hat{a} + \hat{b}x \tag{3.2.2}$$

(3.2.2）式代表x与y之间相关关系的拟合直线，称为回归直线； 是y的估计值,$\hat{y}$ 亦称回归值。

（一）参数*a*、*b*的最小二乘估计

① 参数a与b的最小二乘拟合原则要求y_i与$\hat{y}_i$的误差e_i的平方和达到最小，即

$$Q=\sum_{i=1}^{n}e_i^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}(y_i-a-bx_i)^2\to\min \quad （3.2.3）$$

- 根据取极值的必要条件，要使Q取最小值，必须使Q对a、b的一阶偏导数分别等于零，即

$$\begin{cases} \dfrac{\partial Q}{\partial a} = -2\sum_{i=1}^{n}(y_i - a - bx_i) = 0 \\ \dfrac{\partial Q}{\partial b} = -2\sum_{i=1}^{n}(y_i - a - bx_i)x_i = 0 \end{cases}$$

$$\begin{cases} \sum_{i=1}^{n}(y_i - a - bx_i) = \sum_{i=1}^{n} y_i - na - b\sum_{i=1}^{n} x_i = 0 \\ \sum_{i=1}^{n}(y_i - a - bx_i)x_i = \sum_{i=1}^{n} x_i y_i - a\sum_{i=1}^{n} x_i - b\sum_{i=1}^{n} x_i^2 = 0 \end{cases}$$

- 整理后可得 （3.2.4）

$$\begin{cases} na + (\sum x_i)b = \sum y_i \\ (\sum x_i)a + (\sum x_i^2)b = \sum x_i y_i \end{cases}$$

- 此方程通常称为正规方程组。又可写成矩阵形式

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

③ 解上述正规方程组（3.2.4）式，

得到参数a与b的拟合值

$$\hat{a} = \overline{y} - \hat{b}\,\overline{x}$$

$$\hat{b} = \frac{L_{xy}}{L_{xx}} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2} \qquad （3.2.6）$$

- 解此方程组，即可得到

$$a = \bar{y} - b\bar{x}$$

$$b = \frac{\sum x_i y_i - \frac{1}{n}\left(\sum x_i\right)\left(\sum y_i\right)}{\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2}$$

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

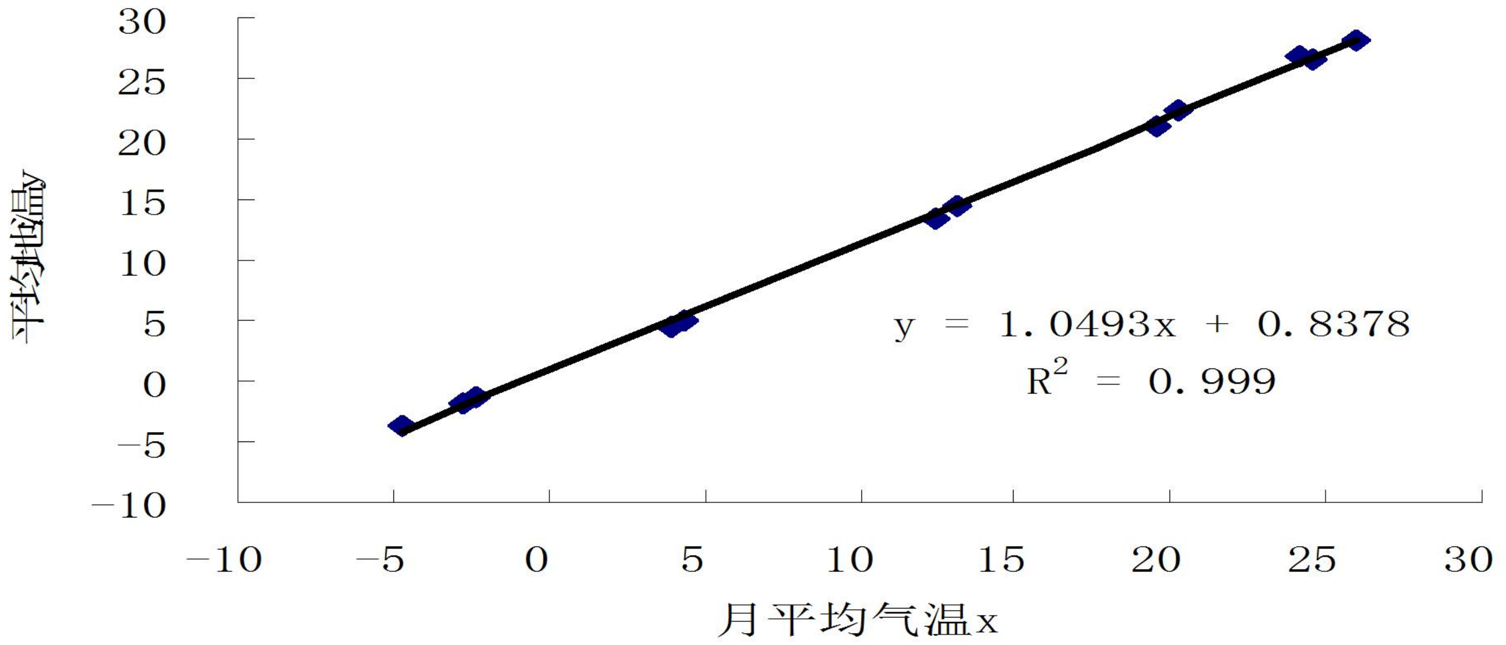
$$\begin{cases} b = \dfrac{l_{xy}}{l_{xx}} \\ r_{xy} = \dfrac{l_{xy}}{\sqrt{l_{xx} \cdot l_{yy}}} \end{cases} \Rightarrow b^2 = r^2 \frac{l_{yy}}{l_{xx}}$$

2、一元线性回归模型的具体建立方法与步骤

月份	气温(x)	地温(y)	xy	x^2	y^2
1	-4.7	-3.6	16.92	22.09	12.96
2	-2.3	-1.4	3.22	5.29	1.96
3	4.4	5.1	22.44	19.36	26.01
4	13.2	14.5	191.40	174.24	210.25
5	20.2	22.3	450.46	408.04	497.29
6	24.2	26.9	650.98	585.64	723.61
7	26.0	28.2	733.20	676.00	795.24
8	24.6	26.5	651.90	605.16	702.25
9	19.5	21.1	411.45	380.25	445.21
10	12.5	13.4	167.50	156.25	179.56
11	4.0	4.6	18.40	16.00	21.16
12	-2.8	-1.9	5.32	7.84	3.61
总和	138.8	155.7	3323.19	3056.16	3619.11
平均值	11.57	12.98			

将计算表中的相应数值代入参数计算公式得到a、b

$$\begin{cases} b = \dfrac{\sum(x_i y_i) - \dfrac{(\sum x_i)(\sum y_i)}{n}}{\sum x_i^2 - \dfrac{(\sum x_i)^2}{n}} \\ = \dfrac{3323.19 - \dfrac{138.8 \times 155.7}{12}}{3056.16 - \dfrac{138.8^2}{12}} = 1.049 \\ a = \overline{y} - b\overline{x} = 12.975 - 1.049 \times 11.57 = 0.84 \\ \hat{y} = a + bx = 0.84 + 1.05x \end{cases}$$



北京市月平均气温与5cm平均地温相关图

（二）一元线性地理回归模型的效果检验

- 回归模型的估计误差
- 回归模型的显著性检验

（1）回归模型的估计误差

$$S = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}$$

- 标准估计误差是一个非常重要的量，由于它的单位和y的单位相同，所以在实际地理问题中便于比较和检验，只要比较S与允许的偏差就行了，因此，它是检验回归效果的极其重要的标志，同时也是衡量地理预测精度的指标

（2）回归模型的显著性检验（F检验法）

- 在回归分析中，y的n次观测值y_1，y_2，…，y_n之间的差异，可以用观测值y_i与其算术平均值的离差平方和来表示，它称为总的离差平方和

$$S_{总} = l_{yy} = \sum (y_i - \overline{y})^2$$

（二）一元线性回归模型的显著性检验

① 方法：F 检验法。

② 总的离差平方和：在回归分析中，表示y的n次观测值之间的差异，记为

$$S_{总} = L_{yy} = \sum_{i=1}^{n}(y_i - \overline{y})^2$$

可以证明

$$S_{总} = L_{yy} = \sum_{i=1}^{n}(y_i - \overline{y})^2 \tag{3.2.8}$$

$$= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 = Q + U \tag{3.2.9}$$

在式（3.2.9）中，Q称为残差平方和，或剩余平方和

$$Q = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

而

$$U = \sum_{i=1}^{n} (\hat{y}_i - \overline{y}_i)^2 = \sum_{i=1}^{n} (a + bx_i - a - b\overline{x})^2$$

$$= b^2 \sum_{i=1}^{n} (x_i - \overline{x})^2 = b^2 L_{xx} = bL_{xy}$$

U称为回归平方和。

$$S_{总} = l_{yy} = \sum (y_i - \overline{y})^2$$
$$= \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \overline{y})^2$$
$$= Q + U$$

$$U = \sum (\hat{y}_i - \overline{y})^2 = b^2 \sum (x_i - \overline{x})^2 = b^2 l_{xx} = b l_{xy}$$
$$Q = \sum (y_i - \hat{y}_i)^2 = l_{yy} - b l_{xy}$$

- Q是所有观测点y_i离回归直线的残差平方和，它表示除x对y的线性影响以外的一切因素对y的变异影响，故称为剩余平方和（残差平方和）
- U反映了在y的总变差中由x与y的线性关系而引起y的变化部分，称为回归平方和。

- 一个回归效果的好坏取决于U和Q的大小，或者说取决于U在总平方和l_{yy}中所占的比例U/l_{yy}的大小，这个比值越大，回归效果越好；反之，则回归效果越不好

$$\frac{U}{l_{yy}} = \frac{b l_{xy}}{l_{yy}} = \frac{l_{xy}^2}{l_{xx} \bullet l_{yy}} = r^2$$

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

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

③ 统计量F

$$F = \frac{U}{\frac{Q}{n-2}} \qquad (3.2.10)$$

④ F越大，模型的效果越佳。统计量$F \sim F(1, n-2)$。在显著水平α下，若$F > F_{\alpha}$，则认为回归方程效果在此水平下显著。一般地，当$F < F_{0.10}(1, n-2)$时，则认为方程效果不明显。

方差分析表
（以北京市气温与地温关系为例）

变差来源	平方和	自由度	方差	F检验	显著性水平
回归（因素x）	$U=bl_{xy}$	1	$S_U^2=U/1$ $=1597.34$	$F=U/(Q/N-2)=10292.4>F_{1,10}^{0.01}=10.04$	$\alpha=0.01$**
剩余（随机因素）	$Q=l_{yy}-bl_{xy}$	$N-2=10$	$S_Q^2=Q/(N-2)$ $=0.157$		
总计	$l_{yy}=\sum y^2-(\sum y)^2/N$	$N-1=11$			

二、多元线性回归模型

- **回归模型的建立**

① 多元线性回归模型的结构形式为

$$y_a = \beta_0 + \beta_1 x_{1a} + \beta_2 x_{2a} + \cdots + \beta_k x_{ka} + \varepsilon_a \qquad (3.2.11)$$

式中：$\beta_0, \beta_1, \cdots, \beta_k$ 为待定参数；ε_a 为随机变量。

② 回归方程：

如果 $b_0, b_1, \cdots, b_k$ 分别为式（3.2.11）中 $\beta_0, \beta_1, \beta_2, \cdots, \beta_k$ 的拟和值，则回归方程为

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k \tag{3.2.12}$$

在（3.2.12）式中，b_0为常数，$b_1, b_2, \cdots b_k$称为偏回归系数。偏回归系数的意义是，当其他自变量都固定时，自变量x_i每变化一个单位而使因变量平均改变的数值。

③ 偏回归系数的推导过程：根据最小二乘法原理，

的估计值 $\beta_i(i=0,1,2,\cdots,k)$ $b_i(i=0,1,2,\cdots,k)$

应该使

$$Q=\sum_{a=1}^{n}(y_a-\hat{y}_a)^2==\sum_{a=1}^{n}[y_a-(b_0+b_1x_{1a}+b_2x_{2a}+\cdots+b_kx_{ka})]^2\rightarrow\min \tag{3.2.13}$$

由求极值的必要条件得

$$\begin{cases}\dfrac{\partial Q}{\partial b_0}=-2\sum_{a=1}^{n}(y_a-\hat{y}_a)=0\\ \dfrac{\partial Q}{\partial b_j}=-2\sum_{a=1}^{n}(y_a-\hat{y}_a)x_{ja}=0\quad(j=1,2,\cdots,k)\end{cases} \tag{3.2.14}$$

方程组（3. 2. 14）式经展开整理后得

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

（3.2.15)

方程组（3.2.15）式称为正规方程组。

引入矩阵

$$
X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ 1 & x_{13} & x_{23} & \cdots & x_{k3} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix}
$$

$$
A = X^T X = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{k1} & x_{k2} & x_{k3} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ 1 & x_{13} & x_{23} & \cdots & x_{k3} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix}
$$

$$= \begin{bmatrix} n & \sum_{a=1}^{n} x_{1a} & \sum_{a=1}^{n} x_{2a} & \cdots & \sum_{a=1}^{n} x_{ka} \\ \sum_{a=1}^{n} x_{1a} & \sum_{a=1}^{n} x_{1\alpha}^{2} & \sum_{a=1}^{n} x_{1a} x_{2a} & \cdots & \sum_{a=1}^{n} x_{1a} x_{ka} \\ \sum_{a=1}^{n} x_{2a} & \sum_{a=1}^{n} x_{1a} x_{2a} & \sum_{a=1}^{n} x_{2a}^{2} & \cdots & \sum_{a=1}^{n} x_{2a} x_{ka} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{a=1}^{n} x_{ka} & \sum_{a=1}^{n} x_{1a} x_{ka} & \sum_{a=1}^{n} x_{2a} x_{ka} & \cdots & \sum_{a=1}^{n} x_{ka}^{2} \end{bmatrix}$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$
B = X^T Y = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{k1} & x_{k2} & x_{k3} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{a=1}^{n} y_a \\ \sum_{a=1}^{n} x_{1a} y_a \\ \sum_{a=1}^{n} x_{2a} y_a \\ \vdots \\ \sum_{a=1}^{n} y_{ka} y_a \end{bmatrix}
$$

则正规方程组（3.2.15）式可以进一步写成矩阵形式

$$
Ab = B
$$

求解得 $b = A^{-1}B = (X^{T}X)^{-1}X^{T}Y$ （3.2.16）

引入记号

$$L_{ij} = L_{ji} = \sum_{a=1}^{n}(x_{ia} - \overline{x}_i)(x_{j\alpha} - \overline{x}_j)$$

$$(i, j = 1, 2, \cdots, k)$$

$$L_{iy} = \sum_{a=1}^{n}(x_{ia} - \overline{x}_i)(y_a - \overline{y})$$

$$(i = 1, 2, \cdots, k)$$

正规方程组也可以写成

$$\begin{cases} L_{11}b_1 + L_{12}b_2 + \cdots + L_{1k}b_k = L_{1y} \\ L_{21}b_1 + L_{22}b_2 + \cdots + L_{2k}b_k = L_{2y} \\ \quad \cdots\cdots\ \cdots\cdots \\ L_{k1}b_1 + L_{k2}b_2 + \cdots + L_{kk}b_k = L_{ky} \\ b_0 = \overline{y} - b_1\overline{x}_1 - b_2\overline{x}_2 - \cdots - b_k\overline{x}_k \end{cases} \tag{3.2.15''}$$

回归模型的显著性检验

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

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

③ 剩余平方和为

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

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。

三、非线性回归模型

- **非线性关系线性化的几种情况**

✓ 对于指数曲线 $y = de^{bx}$，令 $y' = \ln y$，$x' = x$ 可以将其转化为直线形式：$y' = a + bx'$，其中，$a = \ln d$；

✓ 对于对数曲线 $y = a + b\ln x$，令 $y' = y$，$x' = \ln x$，可以将其转化为直线形式：$y' = a + bx'$；

✓ 对于幂函数曲线 $y = dx^b$，令 $y' = \ln y$，$x' = \ln x$，可以将其转化为直线形式：$y' = a + bx'$ 其中，$a = \ln d$；

- ✓ 对于双曲线 $\frac{1}{y}=a+\frac{b}{x}$ ，令 $y'=\frac{1}{y}, x'=\frac{1}{x}$，转化为直线形式: $y'=a+bx'$ ；
- ✓ 对于S型曲线 $y=\frac{1}{a+be^{-x}}$，令 $y'=\frac{1}{y}, x'=e^{-x}$ ，可转化为直线形式: $y'=a+bx'$ ；
- ✓ 对于幂乘积 $y=dx_1^{\beta_1}\cdot x_2^{\beta_2}\cdots x_k^{\beta_k}$ ，只要令 $y'=\ln y, x_1'=\ln x_1, x_2'=\ln x_2, \cdots, x_k'=\ln x_k$ 就可以将其转化为线性形式， $y'=\beta_0+\beta_1 x_1'+\beta_2 x_2'+\cdots+\beta_k x_k'$

  其中，$\beta_0=\ln d$ ；

✓ 对于对数函数和

$$y = \beta_0 + \beta_1 \ln x_1 + \beta_2 \ln x_2 + \cdots + \beta_k \ln x_k$$

只要令 $y' = y, x_1' = \ln x_1, x_2' = \ln x_2, \cdots, x_k' = \ln x_k$ ，就可以将其化为线性形式

$$y' = \beta_0 + \beta_1 x_1' + \beta_2 x_2' + \cdots + \beta_k x_k'$$

例:表3.2.1给出了某地区林地景观斑块面积（area）与周长（perimeter）的数据。下面我们建立林地景观斑块面积A与周长P之间的非线性回归模型。

表3.2.1　某地区各个林地景观斑块面积（m²）与周长（m）

序号	面积A	周长P	序号	面积A	周长P
1	10 447.370	625.392	42	232 844.300	4 282.043
2	15 974.730	612.286	43	4 054.660	289.307
3	30 976.770	775.712	44	30 833.840	895.980
4	9 442.902	530.202	45	1 823.355	205.131
5	10 858.920	1 906.103	46	26 270.300	968.060
6	21 532.910	1 297.962	47	13 573.960	1 045.072
7	6 891.680	417.058	48	65 590.080	2 250.435
8	3 695.195	243.907	49	157 270.400	2 407.549
9	2 260.180	197.239	50	2 086.426	266.541
10	334.332	99.729	51	3 109.070	261.818
11	11 749.080	558.921	52	2 038.617	320.396
12	2 372.105	199.667	53	3 432.137	253.335
13	8 390.633	592.893	54	1 600.391	230.030
14	6 003.719	459.467	55	3 867.586	419.406

15	527 620.200	6 545.291	56	1 946.184	198.661
16	179 686.200	2 960.475	57	77.305	56.902
17	14 196.460	597.993	58	7 977.719	715.752
18	22 809.180	1 103.070	59	19 271.820	1 011.127
19	71 195.940	1 154.118	60	8 263.480	680.710
20	3 064.242	245.049	61	14 697.130	1 234.114
21	46 9416.700	8 226.009	62	4 519.867	326.317
22	5 738.953	498.656	63	13 157.660	1 172.916
23	8 359.465	415.151	64	6 617.270	609.801
24	6 205.016	414.790	65	4 064.137	437.355
25	6 0619.020	1 549.871	66	5 645.820	432.355
26	1 4517.740	791.943	67	6 993.355	503.784
27	31 020.100	1 700.965	68	4 304.281	267.951
28	26 447.160	1 246.977	69	6 336.383	347.136
29	7 985.926	918.312	70	2 651.414	292.235

30	3 638.766	399.725	71	2 656.824	298.473
31	58 5425.100	11 474.770	72	1 846.988	179.866
32	35 220.640	1 877.476	73	1 616.684	172.808
33	10 067.820	497.394	74	1 730.563	172.143
34	27 422.570	1 934.596	75	11 303.970	881.042
35	43 071.550	1 171.413	76	14 019.790	638.176
36	57 585.940	2 275.389	77	9 277.172	862.088
37	28 254.130	1 322.795	78	13 684.750	712.787
38	497 261.000	9 581.298	79	1 949.164	228.403
39	24 255.030	994.906	80	4 846.016	324.481
40	1 837.699	229.401	81	521 457.400	7 393.938
41	1 608.625	225.842	82	564 370.800	12 212.410

解:（1）作变量替换，令: $y = \ln A$,$x = \ln P$，将表3. 2. 1中的原始数据进行对数变换，变换后得到的各新变量对应的观测数据如表3. 2. 2所示。

表3. 2. 2　经对数变换后的数据

序号	y=ln$A$	x=LnP	序号	y=ln$A$	x=LnP
1	9.254 106	6.438 379	42	12.358 13	8.362 186
2	9.678 763	6.417 2	43	8.307 622	5.667 487
3	10.340 99	6.653 782	44	10.336 37	6.797 918
4	9.153 019	6.273 258	45	7.508 433	5.323 65
5	9.292 742	7.552 816	46	10.176 19	6.875 294
6	9.977 338	7.168 551	47	9.515 909	6.951 841
7	8.838 07	6.033 226	48	11.091 18	7.718 879
8	8.214 789	5.496 789	49	11.965 72	7.786 364
9	7.723 2	5.284 414	50	7.643 208	5.585 528
10	5.812 135	4.602 457	51	8.042 079	5.567 651
11	9.371 53	6.326 008	52	7.620 027	5.7695 58

12	7.771 533	5.296 653	53	8.140 938	5.534 711
13	9.034 871	6.385 013	54	7.378 003	5.438 211
14	8.700 134	6.130 066	55	8.260 386	6.038 839
15	13.176 13	8.786 501	56	7.573 626	5.291 597
16	12.098 97	7.993 105	57	4.347 755	4.041 328
17	9.560 748	6.393 579	58	8.984 408	6.573 334
18	10.034 92	7.005 852	59	9.866 399	6.918 821
19	11.173 19	7.051 092	60	9.019 601	6.523 136
20	8.027 556	5.501 457	61	9.595 408	7.118 109
21	13.059 25	9.0150 56	62	8.416 238	5.787 871
22	8.655 032	6.211 917	63	9.484 759	7.067 248
23	9.031 15	6.028 643	64	8.797 438	6.413 133
24	8.733 113	6.027 773	65	8.309 957	6.080 744
25	11.012 36	7.345 927	66	8.638 671	6.069 247
26	9.583 127	6.674 49	67	8.852 716	6.222 147

27	10.342 39	7.438 951	68	8.367 365	5.590 806
28	10.182 9	7.128 478	69	8.754 063	5.849 717
29	8.985 436	6.822 537	70	7.882 848	5.677 56
30	8.199 4	5.990 776	71	7.884 887	5.698 678
31	13.280 09	9.347 906	72	7.521 311	5.192 213
32	10.469 39	7.537 684	73	7.388 132	5.152 181
33	9.217 099	6.209 381	74	7.456 202	5.148 326
34	10.219 12	7.567 654	75	9.332 909	6.781 105
35	10.670 62	7.065 966	76	9.548 225	6.458 614
36	10.961 03	7.729 906	77	9.135 312	6.759 358
37	10.248 99	7.187 502	78	9.524 037	6.569 182
38	13.116 87	9.167 568	79	7.575 156	5.431 112
39	10.096 38	6.902 648	80	8.485 912	5.782 227
40	7.516 27	5.435 471	81	13.164 38	8.908 416
41	7.383 135	5.419 837	82	13.243 47	9.410 208

（2） 以x为横坐标、y为纵坐标，在平面直角坐标系中作出散点图。很明显，y与x呈线性关系。

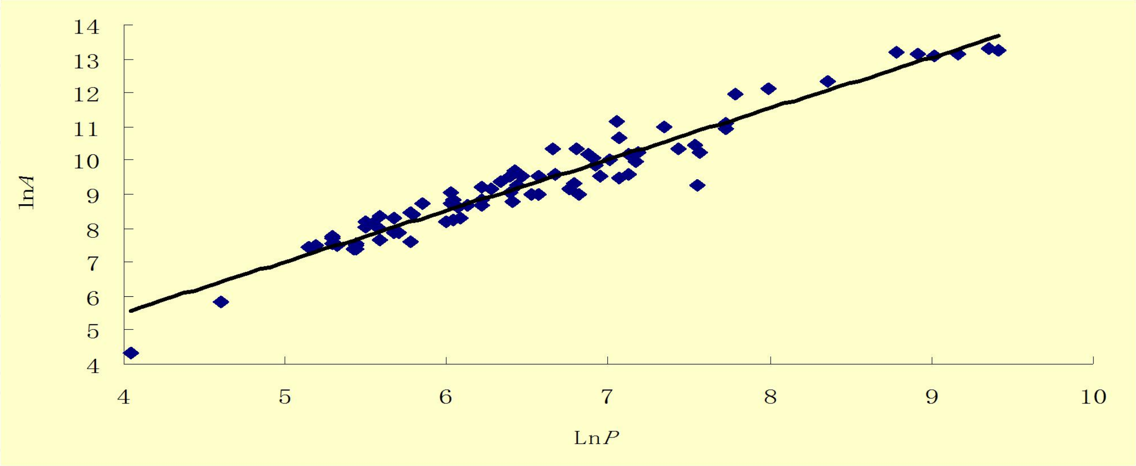


图3.2.2 林地景观斑块面积（A）与周长（P）之间的双对数关系

（3）根据所得表中的数据，运用建立线性回归模型的方法，建立y与x之间的线性回归模型，得到

$$y = 1.505\ x - 0.505\ 7 \tag{3.2.19}$$

对应于（3.2.19）式，x与y的相关系数高达 r_{xy}=0.966 5。

（4）将（3.2.19）还原成双对数曲线，即

$$\ln A = 1.505\ \ln P - 0.505\ 7 \tag{3.2.20}$$

一元非线性回归模型的效果检验

- 相关指数R^2
  - Q是剩余平方和，l_{yy}是总平方和
  - 相关指数的平方根也可称为相关系数
- 相关指数越大，表明选配的回归曲线效果越好，剩余标准差越小，其回归模型的预测精度就越高

$$R^2 = 1 - \frac{\sum (y-\hat{y})^2}{\sum (y-\overline{y})^2} = 1 - \frac{Q}{l_{yy}}$$

$$Q = \sum (y-\hat{y})^2$$

$$l_{yy} = (y-\overline{y})^2$$

§3 地理系统的空间趋势面分析

一、空间趋势面分析概述

趋势面分析

- 用数学的方法，以数学模型来模拟（或拟合）地理数据的空间分布及其区域性变化趋势的方法
- 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。

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。

	题名	作者	来源	发表时间	数据库	被引	下载	预览	分享
1	典范趋势面分析及其在山西省沙棘灌丛水平格局分析中的应用	米湘成; 上官铁梁; 张金屯; 张峰	生态学报	1999-11-30	期刊	32	309		
2	趋势面分析法在环境领域中应用的评述及展望	王江萍; 马民涛; 张菁	环境科学与管理	2009-01-15	期刊	3	338		
3	测井资料趋势面分析法标准化流程建立	王营营; 孙莉莉; 王志章	油气地球物理	2010-10-26	期刊	11	329		
4	趋势面分析在水深测量数据处理中的应用	胡光海; 周兴华	测绘工程	2004-09-30	期刊	28	355		
5	不同种源麻栎种子和苗木性状地理变异趋势面分析	刘志龙; 虞木奎; 马跃; 唐罗忠; 方升佐	生态学报	2011-11-23	期刊	8	174		
6	趋势面分析法在传染病地理分布研究中的应用	裘炯良; 郑剑宁; 周健; 孙智夫	中国热带医学	2004-09-28	期刊	10	159		
7	河南省食管癌死亡率地理分布的趋势面分析	娄清涛; 全培良; 陆建邦; 张欣峰; 孙喜斌	郑州大学学报(医学版)	2010-09-20	期刊	2	229		
8	趋势面分析法在地质灾害危险性区划中的应用——以山西省绛县为例	李晓聪	太原理工大学	2007-05-01	硕士	1	433		
9	趋势面分析法在瓦斯涌出量预测中的应用	张克树; 苏	焦作矿业学	1995-04-28	期刊	13	146		

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。

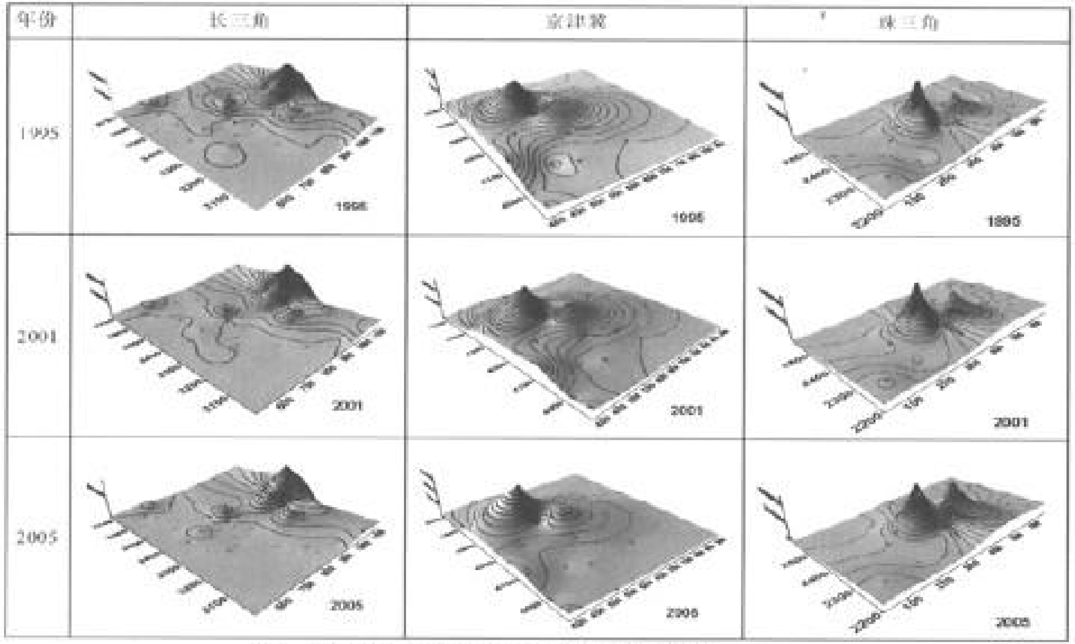
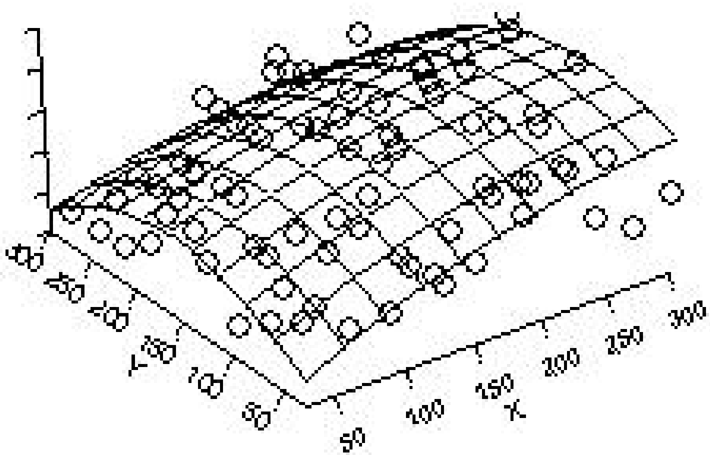


图1 中国三大城市群GDP空间趋势面图谱

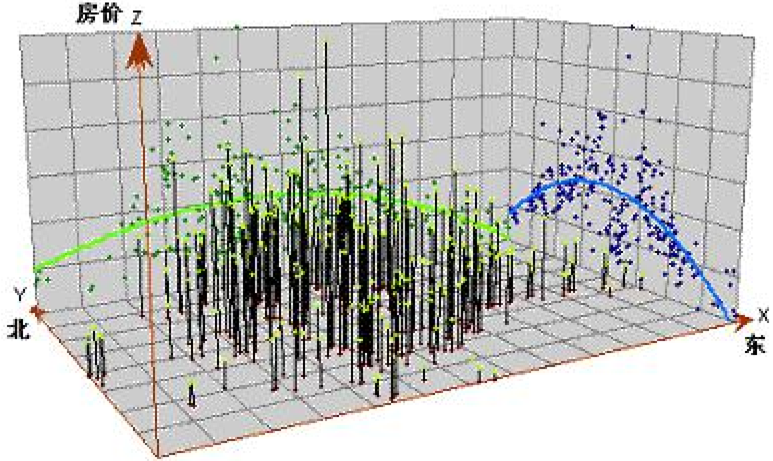
通常把实际的地理曲面分解为趋势面和剩余面两部分，前者反映地理要素的宏观分布规律，属于确定性因素作用的结果；而后者则对应于微观局域，是随机因素影响的结果。

Trend Surface, 2nd Degree Polynomial



趋势面分析的一个基本要求，就是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。



- 趋势面的性质与特点
  - 是一种光滑的数学曲面，它能集中地代表地理数据在大范围内的空间分布变化趋势
  - 与实际上的地理曲面不同，它只是实际曲面的一种近似值。
  - 实际曲面包括趋势面和剩余（或离差）曲面两部分，即
    - 实际曲面＝趋势面+剩余曲面

二、趋势面分析的数学模型

- （一）趋势面分析的数学原理
  - 设以$Z_i(x_i，y_i)$表示某一地理特征值在空间上的分布。其中$(x_i，y_i)$为平面上点的坐标。任一观测点Z_i可分解为两个部分即

$$Z_i(x_i, y_i) = T_i(x_i, y_i) + \varepsilon_i$$

$$\hat{z} = f(x, y) = T_i(x_i, y_i), Z_i(x_i, y_i) = \hat{z}_i(x_i, y_i) + \varepsilon_i$$

$$\varepsilon_i = z_i - \hat{z}_i$$

- 式中：εi即为剩余值（残差值）。
- 显然，当（xi，yi）在空间上变动时,式就刻画了地理要素的实际分布曲面、趋势面和剩余面之间的互动关系

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，即

$$Q = \sum_{i=1}^{n} \varepsilon^{2} = \sum_{i=1}^{n} [z_i(x_i, y_i) - \hat{z}_i(x_i, y_i)]^{2} \rightarrow \min$$

这就是在最小二乘法意义下的趋势面拟合。

用来计算趋势面的数学方程式有**多项式函数**和**傅立叶级数**，其中最为常用的是多项式函数形式。

- 多项式方程作为趋势面方程
  - 因为任何函数在一定范围内总可以用多项式来逼近，并可调整多项式的次数来满足趋势面分析的需要，一般来说，多项式的次数越高则趋势值越接近于观测值，而剩余值越小

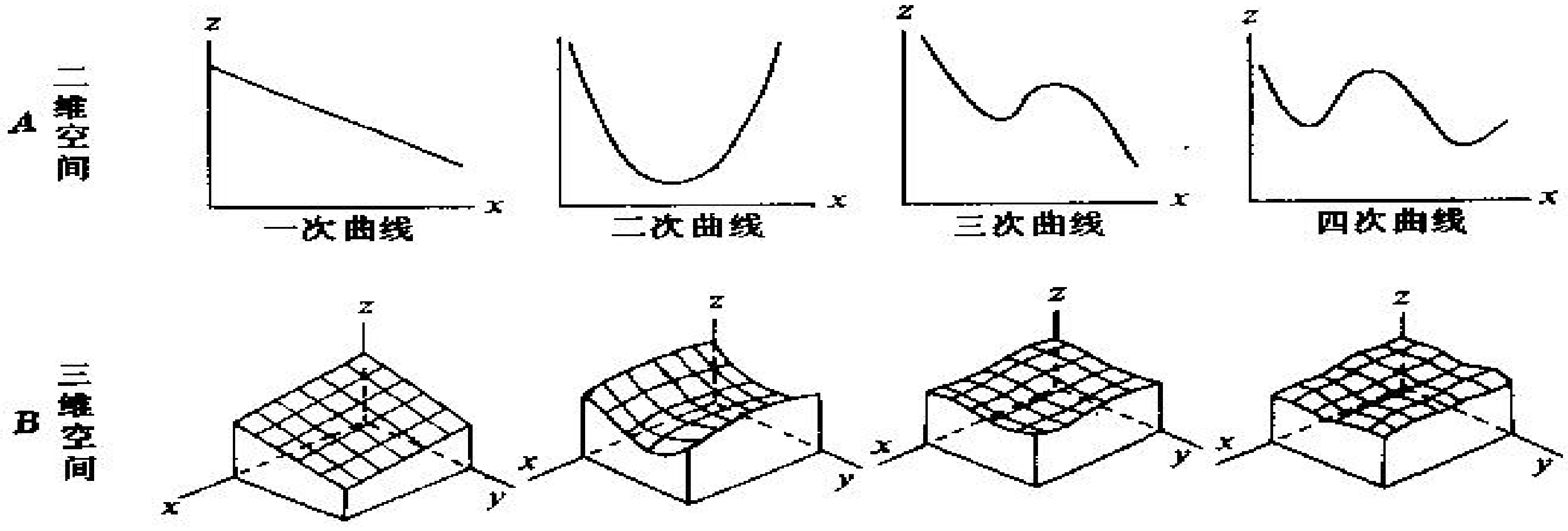
（二）多项式趋势面的数学模型

$$1.z = b_0 + b_1 x + b_2 y$$

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

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

$$4.z = b_0 + b_1 x + b_2 y + b_3 x^2 + b_4 xy + b_5 y^2$$
$$+ b_6 x^3 + b_7 x^2 y + b_8 xy^2 + b_9 x^3 +$$
$$b_{10} x^3 y + b_{11} xy^3 + b_{12} x^2 y^2 + b_{13} x^4 + b_{14} y^4$$



(a)一次倾斜面　(b)二次曲面　(c)三次曲面　(d)四次曲面

（三）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,…,n）确定多项式的系数a_0，a_1，…，a_p，使残差平方和最小。

– 趋势面参数的确定（最小二乘法）

使每一个观测值与趋势值的残差平方和为最小，即

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

按建立多元线性方程的方法，使Q对系数b_0，b_1，…，b_n求偏导，并令这些偏导数等于零，得趋势面的正规方程组，解正规方程组，即可求出系数，从而得到趋势面方程

- 过程

  ① 将多项式回归（非线性模型）模型转化为多元线性回归模型。

令

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

则

$$\hat{z} = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_p x_p$$

②其残差平方和为

$$Q = \sum_{i=1}^{n} [z_i - \hat{z}_i]^2 = \sum_{i=1}^{n} [z_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + \cdots + a_p x_{pi})]^2$$

（3.6.5）

③求Q对a_0，a_1，…，a_p的偏导数，并令其等于0，得正规方程组(式中 $a_0, a_1, \cdots, a_p$ 为p+1个未知量）

$$\begin{cases} na_0 + a_1\sum_{i=1}^{n} x_{1i} + \cdots + a_p\sum_{i=1}^{n} x_{pi} = \sum_{i=1}^{n} z_i \\ a_0\sum_{i=1}^{n} x_{1i} + a_1\sum_{i=1}^{n} x_{1i}x_{1i} + \cdots + a_p\sum_{i=1}^{n} x_{pi}x_{1i} = \sum_{i=1}^{n} x_{1i}z_i \\ \cdots\cdots\cdots\cdots \\ a_0\sum_{i=1}^{n} x_{pi} + a_1\sum_{i=1}^{n} x_{1i}x_{pi} + \cdots + a_p\sum_{i=1}^{n} x_{pi}x_{pi} = \sum_{i=1}^{n} x_{pi}z_i \end{cases} \tag{3.6.6}$$

④用矩阵形式表示

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p1} \\ 1 & x_{12} & x_{22} & \cdots & x_{p2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{pn} \end{bmatrix} \quad Z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \quad A = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{bmatrix}$$

则（3.6.6）式变为

$$X^{\mathrm{T}} XA = X^{T} Z \qquad \textbf{（3.6.7）}$$

⑤ 对于二元二次多项式有

$$z = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 xy + a_5 y^2$$

其正规方程组为

$$
\begin{bmatrix}
1 & 1 & \cdots & 1 \\
x_1 & x_2 & \cdots & x_n \\
y_1 & y_2 & \cdots & y_n \\
x_1^2 & x_2^2 & \cdots & x_n^2 \\
x_1 y_1 & x_2 y_2 & \cdots & x_n y_n \\
y_1^2 & y_2^2 & \cdots & y_n^2
\end{bmatrix}
\cdot
\begin{bmatrix}
1 & x_1 & y_1 & x_1^2 & x_1 y_1 & y_1^2 \\
1 & x_2 & y_2 & x_2^2 & x_2 y_2 & y_2^2 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & x_n & y_n & x_n^2 & x_n y_n & y_n^2
\end{bmatrix}
\cdot
\begin{bmatrix}
a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5
\end{bmatrix}
$$

$$
=
\begin{bmatrix}
1 & 1 & \cdots & 1 \\
x_1 & x_2 & \cdots & x_n \\
y_1 & y_2 & \cdots & y_n \\
x_1^2 & x_2^2 & \cdots & x_n^2 \\
x_1 y_1 & x_2 y_2 & \cdots & x_n y_n \\
y_1^2 & y_2^2 & \cdots & y_n^2
\end{bmatrix}
\cdot
\begin{bmatrix}
z_1 \\ z_2 \\ \vdots \\ z_n
\end{bmatrix}
$$

由式（3.6.7）求解，可得

$$A = (X^T X)^{-1} X^T Z \quad （3.6.8）$$

（三）趋势面拟合程度的检验

- 趋势面分析拟合程度与回归模型的效果直接相关，因此，对趋势面分析进行适度性检验是一个关系到趋势面能否在实际研究中加以应用的关键问题，也是趋势面分析中不可缺少的重要环节。
- 这可以通过以下检验来完成：

✓ 趋势面拟合适度的R^2检验

✓ 趋势面拟合适度的显著性F检验

✓ 趋势面适度的逐次检验

（一）趋势面拟合适度的R^2检验

趋势面与实际面的拟合度系数R^2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。即

$$SS_T = \sum_{i=1}^{n}(z_i - \hat{z}_i)^2 + \sum_{i=1}^{n}(\hat{z}_i - \bar{z})^2 = SS_D + SS_R$$

$SS_D=\sum_{i=1}^{n}(z_i-\hat{z}_i)^2$ 为剩余平方和，它表示随机因素对离差的影响，$SS_R=\sum_{i=1}^{n}(\hat{z}_i-\bar{z})^2$ 为回归平方和，它表示自变量对因变量的离差的总影响。

SS_R 越大（或 SS_D 越小）就表示因变量与自变量的关系越密切，回归的规律性越强、效果越好。

记 $$R^2=\frac{SS_R}{SS_T}=1-\frac{SS_D}{SS_T}$$

R^2 越大，趋势面的拟合度就越高。（3.6.9）

（二）趋势面拟合适度的显著性F 检验

趋势面适度的F检验，是对趋势面回归模型整体的显著性检验。

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。即

$$F=\frac{SS_R/p}{SS_D/n-p-1} \quad （3.6.10）$$

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

- 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。

表3.6.1　多项式趋势面由K次增高至（K+1）次 的回归显著性检验

离差来源	平方和	自由度	均方差	F检验
（K+1）次回归	$SS_R^{(K+1)}$	p	$MS_R^{(K+1)}=SS_R^{(K+1)}/p$	$MS_R^{(K+1)}/MS_D^{(K+1)}$
（K+1）次剩余	$SS_D^{(K+1)}$	$n-p-1$	$MS_D^{(K+1)}$ $=SS_D^{(K+1)}/(n-p-1)$	
K次回归	$SS_R^{(K)}$	q	$MS_R^{(K)}=SS_R^{(K)}/q$	$MS_R^{(K)}/MS_D^{(K)}$
K次剩余	$SS_D^{(K)}$	$n-q-1$	$MS_D^{(K)}=$ $SS_D^{(K)}/(n-q-1)$	
由K次增高至（K+1）次的回归	$SS_R^{(I)}=$ $SS_R^{(K+1)}-SS_R^K$	$p-q$	$MS_R^{(I)}=$ $SS_R^{(I)}/(p-q)$	$MS_R^{(I)}/MS_D^{(K+1)}$
总离差	SS_T			

需要注意的是，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

某流域1月份降水量与各观测点的坐标位置数据如表3.6.2所示。下面，我们以降水量为因变量z，地理位置的横坐标和纵坐标分别为自变量x、y，进行趋势面分析，并对趋势面方程进行适度F检验。

表3. 6. 2　流域降水量及观测点的地理位置数据

序号	降水量Z/mm	横坐标 $x/10^4$ m	纵坐标 $y/10^4$ m
1	27.6	0	1
2	38.4	1.1	0.6
3	24	1.8	0
4	24.7	2.95	0
5	32	3.4	0.2
6	55.5	1.8	1.7
7	40.4	0.7	1.3
8	37.5	0.2	2
9	31	0.85	3.35
10	31.7	1.65	3.15
11	53	2.65	3.1
12	44.9	3.65	2.55

解题步骤

- 建立趋势面模型

（1）首先采用二次多项式进行趋势面拟合，用最小二乘法求得拟合方程为

$$z = 5.998 + 17.438x + 29.787y - 3.558x^2 + 0.357xy - 8.070y^2$$

$$R^2 = 0.839, F = 6.236$$

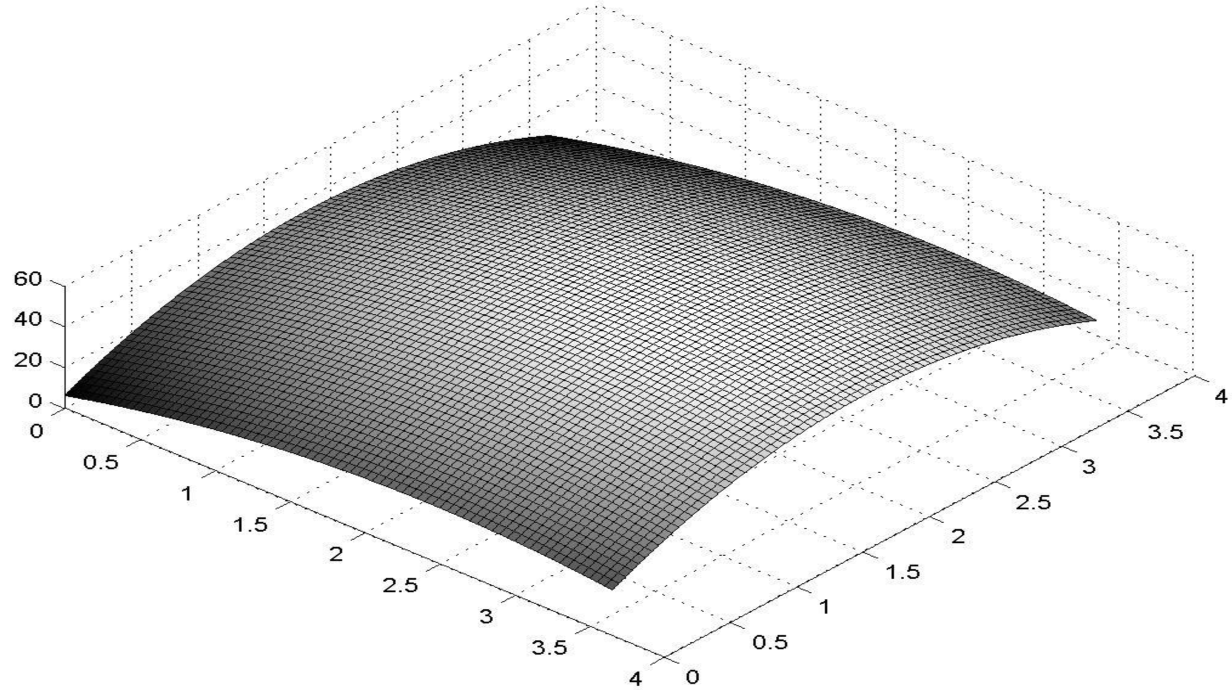


图3.6.1　某流域降水量的二次多项式趋势面

(2)再采用三次趋势面进行拟合，用最小二乘法求得拟合方程为

$$z = -48.810 + 37.557x + 130.130y + 8.389x^2 - 33.166xy$$

$$-62.740y^2 - 4.133x^3 + 6.138x^2y + 2.566xy^2 + 9.785y^3$$

$$R^2 = 0.965, F = 6.054$$

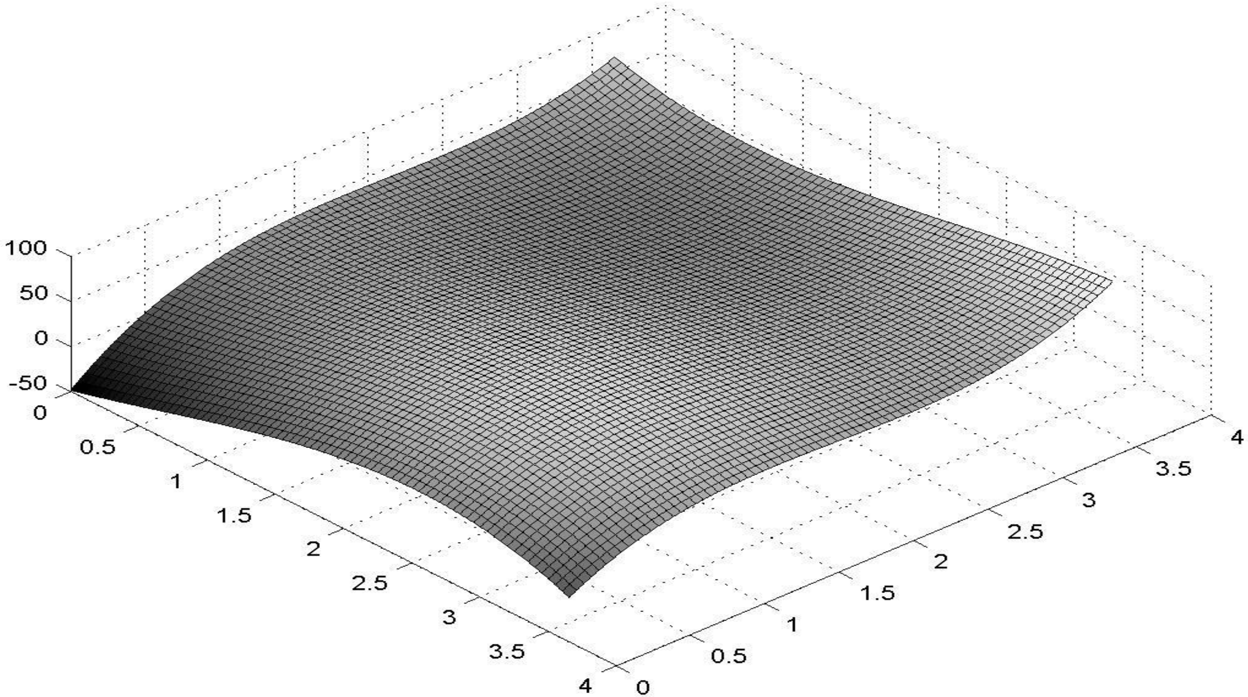


图3.6.2　某流域降水量的三次多项式趋势面

■模型检验

(1)趋势面拟合适度的R^2检验: 根据R^2检验方法计算，结果表明，二次趋势面的判定系数为$R_2^2=0.839$，三次趋势面的判定系数为$R_3^2=0.965$，可见二次趋势面回归模型和三次趋势面回归模型的显著性都较高，而且三次趋势面较二次趋势面具有更高的拟合程度。

(2) 趋势面适度的显著性F检验：根据F检验方法计算，结果表明，二次趋势面和三次趋势面的F值分别为F_2=6.236和F_3=6.054。在置信水平 α=0.05下，查F分布表得 $F_{2\alpha} = F_{0.05}(5,6) = 4.53$ $F_{3\alpha}=F_{0.05}(9,2) = 19.4$。显然 $F_2 > F_{2\alpha}$，$F_3 < F_{3\alpha}$， 故二次趋势面的回归方程显著而三次趋势面不显著。因此，F检验的结果表明，用二次趋势面进行拟合比较合理。

(3)趋势面适度的逐次检验：

趋势面比较：在二次和三次趋势面检验中，对两个阶次趋势面模型的适度进行比较，相应的方差分析计算结果见表3.6.3。

表3. 6. 3　二次和三次趋势面回归模型的逐次检验方差分析表

离差来源	平方和	自由度	均方差	F 检验
三次回归	1 129.789	9	125.532	
三次剩余	41.474	12−9−1	20.737	6.054
二次回归	982.244	5	196.449	
二次剩余	189.018	12−5−1	31.503	6.236
由二次增高至三次的回归	147.545	4	36.886	1.779

分析：从二次趋势面增加到三次趋势面，$F_{3\to2}$=1.779。在置信度水平 α=0.05下，查F分布表得$F_{0.05}(4,2)$=6.94，由于$F_{3\to2} < F_{0.05}(4,2)$=6.94，故将趋势面拟合次数由二次增高至三次，对回归方程并无新贡献，因而选取二次趋势面比较合适。这也进一步验证了趋势面拟合适度的显著性F检验的结论。

趋势面的具体计算方法与步骤

- 原始数据列表
- 等间隔选取纵横坐标网，将原始数据点入坐标
- 按多元线性回归分析方法求出趋势面的正规方程组，解出参数
- 从趋势值等值线图中，获得地理要素的区域性变化规律
- 用F分布对趋势面进行拟合程度检验