# 第三章 相关分析与回归分析

- 相关分析方法
- 回归分析方法

# §3.1 相关分析方法

相关分析法的任务是，揭示地理要素之间相互关系的密切程度。

# 1. 两要素之间相关程度的测定

- 相关系数的计算与检验
- 秩相关系数的计算与检验

# 相关系数的计算与检验

■相关系数的计算

定义：

$$r_{xy}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \qquad (3.1.1)$$

$\bar{x}$ 和 $\bar{y}$ 为两要素的平均值。即

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i \qquad \bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$$

相关系数 $r_{xy}$ ，是表示该两个要素之间的相关程度的统计指标，有如下性质：

（1）$r_{xy}$的分布范围，介于[-1，1]区间；

（2）$r_{xy}>0$，表示正相关，即两要素同向关；$r_{xy}\langle 0$，表示负相关，即两要素异向相关；

（3）的绝对值越接近于1，表示两个要素之间的关系越密切；越接近于0，表示两个要素的关系越不密切。

如果记：

令

$$L_{xy}=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\sum_{i=1}^{n}x_iy_i-\frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)\left(\sum_{i=1}^{n}y_i\right)$$

$$L_{xx}=\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}x_i^{\ 2}-\frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)^2$$

$$L_{yy}=\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}y_i^{\ 2}-\frac{1}{n}\left(\sum_{i=1}^{n}y_i\right)^2$$

公式4.1.1可简化为：

$$r_{xy}=\frac{L_{xy}}{\sqrt{L_{xx}L_{yy}}} \qquad (3.1.2)$$

显然，由公式（3.1.1）或（3.1.2）容易知道：

（1）$r_{xx}=1$，$r_{yy}=1$，即每一个要素与它本身的相关程度最大。

（2）$r_{xy}=r_{yx}$，即要素 $x$ 对要素 $y$ 的相关程度相等。

例如，根据表4.1.1中的数据，我们可以代入公(3.1.1)，计算伦敦市月平均气温与之间的相关系数：

$$r_{tp}=\frac{\sum_{i=1}^{12}(t_i-\bar{t})(p_i-\bar{p})}{\sqrt{\sum_{i=1}^{12}(t_i-\bar{t})^2}\sqrt{\sum_{i=1}^{n12}(p_i-\bar{p})^2}}=\frac{-300.91}{\sqrt{250.55}\sqrt{1508.34}}=-0.4895$$

表3.1.1 伦敦市的月平均气温与降水量

| 月份 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 平均气温 t/℃ | 3.8 | 4.0 | 5.8 | 8.0 | 11.3 | 14.4 | 16.5 | 16.2 | 13.8 | 10.8 | 6.7 | 4.7 |
| 降水量 p/mm | 77.7 | 51.2 | 60.1 | 54.1 | 55.4 | 56.8 | 45.0 | 55.3 | 67.5 | 73.3 | 76.6 | 79.6 |

计算结果表明，伦敦市的月平均气温与降水量之间呈负相关，即异向相关。

## ▲相关系数的检验：

相关系数的检验是在给定的置信水平下，查相关系数检验的临界值表来完成的，前人已经制出了相关系数检验表（表略）。

我们对上面计算出来的伦敦市月平均气温($t$)与降水量($p$)的相关系数进行显著性检验，$f=12-10=2$，在显著水平$\alpha=0.10$上查表可知：$r_{0.10}=0.4973$。因为$|r_{tp}|=0.4895\langle r_{\alpha}=0.4973$，所以，伦敦市月平均气温与降水量之间的相关性不显著。

# 秩相关系数的计算与检验

- 定义：秩相关系数，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。
- $R_1$代表要素x的序号（或位次），$R_2$代表要素y的序号（或位次），$d_i^2=(R_{1i}-R_{2i})^2$ 代表要素x和y的同一组样本位次差的平方，那么要素X与Y之间的秩相关系数被定义为：

$$r'_{xy}=1-\frac{6\sum_{i=1}^{n}d_i^2}{n(n^2-1)} \qquad (3.1.3)$$

# 2 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

# 偏相关系数的计算与检验

## 偏相关系数的计算

① 定义。在多要素所构成的地理系统中，先不考虑其它要素的影响，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 计算。3个要素的偏相关系数

$$r_{12\cdot3}=\frac{r_{12}-r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} \tag{3.1.5}$$

$$r_{13\cdot2}=\frac{r_{13}-r_{12}r_{23}}{\sqrt{(1-r_{12}^2)(1-r_{23}^2)}} \tag{3.1.6}$$

$$r_{23\cdot1}=\frac{r_{23}-r_{12}r_{13}}{\sqrt{(1-r_{12}^2)(1-r_{13}^2)}} \qquad (3.1.7)$$

四个要素的偏相关系数

$$r_{12\cdot34}=\frac{r_{12\cdot3}-r_{14\cdot3}r_{24\cdot3}}{\sqrt{(1-r_{14\cdot3}^2)(1-r_{24\cdot3}^2)}} \qquad (3.1.8)$$

$$r_{13\cdot24}=\frac{r_{13\cdot2}-r_{14\cdot2}r_{34\cdot2}}{\sqrt{(1-r_{14\cdot2}^2)(1-r_{34\cdot2}^2)}} \qquad (3.1.9)$$

$$r_{14\cdot23}=\frac{r_{14\cdot2}-r_{13\cdot2}r_{43\cdot2}}{\sqrt{(1-r_{13\cdot2}^2)(1-r_{43\cdot2}^2)}} \qquad (3.1.10)$$

$$r_{23\cdot14}=\frac{r_{23\cdot1}-r_{24\cdot1}r_{34\cdot1}}{\sqrt{(1-r_{24\cdot1}^2)(1-r_{34\cdot1}^2)}} \qquad (3.1.11)$$

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

偏相关系数的显著性检验，一般采用t-检验法。其统计量计算公式为：

$$t = \frac{r_{12\circ 34\cdots m}}{\sqrt{1 - r^2_{12\circ 34\cdots m}}}\sqrt{n - m - 1}$$

■ 例3：对于某四个地理要素$x_1$，$x_2$，$x_3$，$x_4$的23个样本数据，经过计算得到了如下的单相关系数矩阵：

$$R=\begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} \\ r_{21} & r_{22} & r_{23} & r_{24} \\ r_{31} & r_{32} & r_{33} & r_{34} \\ r_{41} & r_{42} & r_{43} & r_{44} \end{bmatrix}=\begin{bmatrix} 1 & 0.416 & 0.346 & 0.579 \\ 0.416 & 1 & -0.592 & 0.950 \\ -0.346 & -0.592 & 1 & -0.469 \\ 0.579 & 0.950 & -0.469 & 1 \end{bmatrix}$$

解题步骤:

① 利用一级偏向关系数公式计算一级偏向关系数，如表3.1.5 所示：

| $r_{12\cdot3}$ | $r_{13\cdot2}$ | $r_{14\cdot2}$ | $r_{14\cdot3}$ | $r_{23\cdot1}$ | $r_{24\cdot1}$ | $r_{24\cdot3}$ | $r_{24\cdot1}$ | $r_{34\cdot2}$ |
|---|---|---|---|---|---|---|---|---|
| 0.821 | 0.808 | 0.647 | 0.895 | −0.863 | 0.956 | 0.945 | −0.875 | 0.371 |

② 利用 二级偏相关系数公式计算二级偏相关系数，见下表。

| $r_{12\cdot34}$ | $r_{13\cdot24}$ | $r_{14\cdot23}$ | $r_{23\cdot14}$ | $r_{24\cdot13}$ | $r_{34\cdot12}$ |
|---|---|---|---|---|---|
| -0.170 | 0.802 | 0.635 | -0.187 | 0.821 | -0.337 |

③ 检验

$$t = \frac{0.821}{\sqrt{1-0.821^2}}\sqrt{23-3-1} = 6.268$$

查t分布表，在自由度为23-3-1=19时，

$t_{0.001}$=3.883，显然 $t>t_\alpha$，这表明在置信度水平 $\alpha$ = 0.001上，偏相关系数$r_{24\cdot13}$是显著的。

# 复相关系数的计算与检验

**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 。

计算：

当有两个自变量时，$R_{y\circ12}=\sqrt{1-(1-r^2_{y1})(1-r^2_{y2\circ1})}$ （3.1.15）

当有三个自变量时，

$$R_{y\circ123}=\sqrt{1-(1-r^2_{y1})(1-r^2_{y2\circ1})(1-r^2_{y3\circ12})}$$ （3.1.16）

当有k个自变量时，

$$R_{y\circ12\ldots k}=\sqrt{1-(1-r^2_{y1})(1-r^2_{y2\circ1})\cdots[1-r^2_{yk\circ12\ldots(k-1)})]}$$ （3.1.17）

性质

① 复相关系数介于0到1之间，即

$$0 \le R_{y \circ 12 \cdots k} \le 1$$

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。

## 显著性检验:

F-检验法。其统计量计算公式为

$$F = \frac{R^2_{y \circ 12 \cdots k}}{1 - R^2_{y \circ 12 \cdots k}} \times \frac{n - k - 1}{k}$$

其中，n为样本数，k为自变量个数

# §3.2 回归分析方法

- 一元线性回归模型
- 多元线性回归模型
- 非线性回归模型的建立方法

# 1.一元线性回归模型

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

$$y_\alpha = a + bx_a + \varepsilon_a \qquad (3.2.1)$$

或

$$\hat{y} = \hat{a} + \hat{b}x \qquad (3.2.2)$$

（记 $\hat{a}$ 和 $\hat{b}$ 分别为参数a与b的拟合值）

参数a、b的最小二乘估计

## (1) 参数a,b的最小二乘估计

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：

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

根据取极值的必要条件，有 ：

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

即：

$$\begin{cases}\sum_{i=1}^{n}(y_i-a-bx_i)=0\\ \sum_{i=1}^{n}(y_i-a-bx_i)x_i=0\end{cases}\qquad (3.2.4)$$

解上述正规方程组（3.2.4）式，得到参数a与b的**拟合值：**

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

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

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

## (2)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：

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

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

（3.2.9）式中，称为误差平方和，或剩余平方和，而

$$U=\sum_{i=1}^{n}(\hat{y}_i-\bar{y}_i)^2=\sum_{i=1}^{n}(a+bx_i-a-b\bar{x})^2=b^2\sum_{i=1}^{n}(x_i-\bar{x})^2=b^2L_{xx}=bL_{xy}$$

称为回归平方和。

③ 统计量F

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

显然，F越大，模型的效果越佳。实际上，统计量F服从自由度$f_1$=1和$f_2$= 2的F分布，即F～F(1,n-2）。在显著水平a下，若$F>F_{\alpha}(1,n-2)$，则认为回归方程效果在此水平下显著的。一般地，当$F<F_{0.10}(1,n-2)$时，则认为方程效果不明显。

# 2. 多元回归模型

## (1) 回归模型的建立

① 多元线性回归模型的结构形式：

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

② 回归方程：

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

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，当其它自变量都固定时，自变量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

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

由求极值的必要条件得

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

方程组（3.2.14）式经展开整理后得

$$
\begin{cases}
nb_0+(\sum_{a=1}^{n}x_{1a})b_1+(\sum_{a=1}^{n}x_{2a})b_2+\cdots+(\sum_{a=1}^{n}x_{ka})b_k=\sum_{a=1}^{n}y_a \\
(\sum_{a=1}^{n}x_{1a})b_0+(\sum_{a=1}^{n}x_{1\alpha}^2)b_1+(\sum_{a=1}^{n}x_{1a}x_{2a})b_2+\cdots+(\sum_{a=1}^{n}x_{1a}x_{ka})b_k=\sum_{a=1}^{n}x_{1a}y_a \\
(\sum_{a=1}^{n}x_{2a})b_0+(\sum_{a=1}^{n}x_{1a}x_{2a})b_1+\sum_{a=1}^{n}(x_{2a}^2)b_2+\cdots+(\sum_{a=1}^{n}x_{2a}x_{ka})b_k=\sum_{a=1}^{n}x_{2a}y_a \\
\cdots\cdots\cdots\cdots \\
(\sum_{a=1}^{n}x_{ka})b_0+(\sum_{a=1}^{n}x_{1a}x_{ka})b_1+(\sum_{a=1}^{n}x_{2a}x_{ka})b_2+\ldots.+(\sum_{a=1}^{n}x_{ka}^2)b_k=\sum_{a=1}^{n}x_{ka}y_a
\end{cases}
\quad (3.2.15)
$$

方程组（3.2.15）式称为正规方程组。

引入矩阵：

$$b=\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}, \quad Y=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X=\begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ 1 & x_{13} & x_{23} & \cdots & x_{k3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix}$$

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

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

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

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：

$$b = A^{-1}B = (X^TY)^{-1}X^TY$$

引入记号：

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

$$= \sum_{a=1}^{n} x_{ia}x_{ja} - \frac{1}{n}(\sum_{a=1}^{n} x_{ia})(\sum_{a=1}^{n} x_{ja}) \qquad (i, j = 1,2,\cdots,k)$$

正规方程组也可以写成:

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

## (2) 多元线性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：

$$Q = \sum_{a=1}^{n} (y_a - \hat{y}_a)^2 = L_{yy} - U$$

④ F统计量为：

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

当统计量F计算出来之后，就可以查F分布表对模型进行显著性检验。

# 3. 非线性回归模型

（1）非线性关系线性化的几种情况:

① 对于指数曲线 $y = de^{bx}$，可以将其转化为直线形式：$y' = a + bx'$，其中a=lnd。

② 对于对数曲线 $y=a+b\ln x$，可以将其转化为直线形式：$y' = a + bx'$

③ 对于幂函数曲线 $y = dx^{b}$，可以将其转化为直线形式：$y' = a + bx'$，其中a=lnd。

④ 对于双曲线 $\frac{1}{y}=a+\frac{b}{x}$，转化为直线形式：$y'=a+bx'$

⑤ 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：$y'=a+bx'$。

⑥ 对于幂函数乘积 $y=dx_1^{\beta_1}\cdot x_2^{\beta_2}\cdots x_k^{\beta_k}$ 转化为直线形式：

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

⑦ 对于对数函数和：

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

转化为线性形式：

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

# 第3节 相关分析与回归分析实例分析

1.相关分析应用实例

2.多元线性回归模型的建立

# 1.相关分析应用实例

下面我们选取且末县1985年-2004年序列资料作为分析样本（表3.3.1），从中选取9个解释变量，$X_1$-总人口（人），$X_2$-人均耕地面积（亩），$X_3$-播种面积（亩），$X_4$-GDP（万元），$X_5$ -人均GDP（元），$X_6$-社会消费品零售总额（万元），$X_7$-固定资产投资总额（万元），$X_8$-农业总产值（万元），$X_9$-粮食总产量（t），Y-耕地面积。

| 年份 | 总人口 | 人均耕地（亩） | 总播面积（亩） | GDP 总值（万元） | 人均 GDP（元） | 社会消费品零售总额（万元） | 固定资产投资总额（万元） | 农业总产值（万元） | 粮食总产量（吨） |
|---|---|---|---|---|---|---|---|---|---|
| 1985 | 37471 | 3.41 | 117000 | 3278.43 | 874.93 | 1064.30 | 484.00 | 1410.06 | 33559 |
| 1986 | 38389 | 3.32 | 117300 | 3707.91 | 965.88 | 1192.00 | 588.00 | 1533.58 | 33882 |
| 1987 | 39154 | 3.14 | 112200 | 3759.71 | 960.29 | 1502.00 | 923.00 | 1677.68 | 20510 |
| 1988 | 40838 | 3.01 | 122100 | 4293.87 | 1051.44 | 1985.00 | 2965.00 | 1893.98 | 21434 |
| 1989 | 42539 | 2.95 | 129700 | 4495.92 | 1056.89 | 1947.00 | 170.00 | 2072.32 | 22285 |
| 1990 | 43942 | 2.87 | 129700 | 5146.70 | 1171.25 | 1985.00 | 613.00 | 5162.44 | 22928 |
| 1991 | 44230 | 2.85 | 136500 | 5389.30 | 1018.47 | 2296.00 | 243.00 | 5295.39 | 23289 |
| 1992 | 45727 | 2.76 | 143100 | 5322.20 | 1163.91 | 2297.00 | 438.00 | 5700.42 | 24088 |
| 1993 | 46349 | 2.72 | 133500 | 6537.93 | 1410.59 | 2423.00 | 593.00 | 5915.71 | 24378 |
| 1994 | 48151 | 2.62 | 137700 | 7761.43 | 1611.89 | 3323.00 | 2011.00 | 6778.69 | 21130 |
| 1995 | 48330 | 2.61 | 139500 | 8072.89 | 1670.37 | 4046.00 | 4404.00 | 6847.75 | 23380 |
| 1996 | 48499 | 2.60 | 145950 | 10155.37 | 2093.93 | 4803.00 | 5003.00 | 7901.70 | 24420 |
| 1997 | 49963 | 2.52 | 145050 | 12014.90 | 2384.43 | 5454.00 | 7221.90 | 8926.79 | 25455 |
| 1998 | 51701 | 2.54 | 164681 | 13690.59 | 2648.03 | 5608.00 | 16596.00 | 10704.20 | 26166 |
| 1999 | 52651 | 2.69 | 167507 | 13047.92 | 2481.00 | 5949.40 | 5216.00 | 10901.02 | 23841 |
| 2000 | 52168 | 2.71 | 173800 | 23492.00 | 4480.00 | 6237.30 | 12585.00 | 10826.03 | 25897 |
| 2001 | 54056 | 2.63 | 171150 | 25976.00 | 4732.00 | 6884.00 | 16693.00 | 11864.37 | 26824 |
| 2002 | 55591 | 2.57 | 176700 | 28686.00 | 5169.00 | 7724.03 | 17894.00 | 12901.86 | 31023 |
| 2003 | 56600 | 2.69 | 182400 | 32737.00 | 5783.92 | 8883.00 | 13040.00 | 29205.00 | 22336 |
| 2004 | 58100 | 2.66 | 187500 | 36059.00 | 6206.00 | 15874.00 | 13532.00 | 31235.00 | 19082 |

表3.3.1 且末县1985-2004年部分指标序列数据

以公式（3.1.1）为计算基础，可以得出相关系数矩阵

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $X_8$ | $X_9$ | y |
|---|---|---|---|---|---|---|---|---|---|---|
| $X_1$ | 1 | | | | | | | | | |
| $X_2$ | -0.8283 | 1 | | | | | | | | |
| $X_3$ | 0.9678 | -0.7033 | 1 | | | | | | | |
| $X_4$ | 0.8902 | -0.5202 | 0.9253 | 1 | | | | | | |
| $X_5$ | 0.8887 | -0.5220 | 0.9247 | 0.9989 | 1 | | | | | |
| $X_6$ | 0.8770 | -0.5618 | 0.8817 | 0.9223 | 0.9126 | 1 | | | | |
| $X_7$ | 0.8306 | -0.5823 | 0.8652 | 0.8734 | 0.8833 | 0.7670 | 1 | | | |
| $X_8$ | 0.8495 | -0.5109 | 0.8578 | 0.9075 | 0.8941 | 0.9333 | 0.6884 | 1 | | |
| $X_9$ | -0.2179 | 0.3686 | -0.1158 | -0.0774 | -0.0602 | -0.243 | 0.102 | -0.2876 | 1 | |
| y | 0.8187 | -0.3599 | 0.8950 | 0.9468 | 0.9413 | 0.8815 | 0.7611 | 0.8985 | -0.0393 | 1 |

表**3.3.2** 相关系数矩阵

# 相关系数矩阵结果分析

- 从表3.2.2中可以看，影响耕地数量变化的9个因子之间存在不同程度的相关性。它们的相关性可以如下概括：
- （1）耕地面积与播种面积，GDP，人均GDP，社会消费平零售总额之间（y与$x_3$，$x_4$，$x_5$，$x_6$）存在较大的正相关性，说明经济水平的变化影响耕地数量的变化，耕地数量的变化也会把各种经济指标向正发展方向推动。
- （2）播种面积和GDP，人均GDP，农业总产值之间($x_3$与$x_4$，$x_5$,$x_8$之间)存在较大的正相关关系。这说明在当前的发展水平条件下，农业生产还是提高人们生活水平，促进经济发展的基础。
- （3）$x_4$与$x_5$,$x_6$,$x_8$之间，$x_5$与$x_6$之间也存在较大的正相关关系，这些都是表示社会经济发展发展水平的不同指标之间相有着一定的关性。

# 2.多元线性回归模型的建立

- 以且末县1985年-2004年的序列数据为分析 样本，以上所述的9个解释变量为自变量，耕地面积为因变量，以1.1.2中所讲的建立回归模型的步骤为基本原理，建立多元线性回归模型如下：

$$y = -98676.0843 + 2.4037x_1 + 38754.3977x_2 + 0.0858x_3 + 0.2161x_4 - 0.5463x_5$$
$$-0.2292x_6 - 0.164x_7 - 0.0433x_8 - 0.1527x_9$$

# 多元线性回归模型的检验

- 对多元回归模型进行显著性检验时，通常用F检验法。当F≥F0.01(k，n-k-1)成立时认为线性回归是高度显著，用F**来表示。模型中F=1199.2227，远大于F0.01（9，10）=5.26，即F**=1199.2227≥F0.01（9，10）=5.26所以该方程是高度显著的。

# 多元线性回归模型的检验

| | 平方和 | 自由度 | 均方 | F检验与 $\overline{R}^2$ |
|---|---|---|---|---|
| 回归 | U=1.855e+009 | k=9 | U/K=2.061e+008 | |
| 残差 | Q=1718423.1247 | n-k-1=10 | $S^2=\dfrac{Q}{n-k-1}=$ 171842.3125 | $F=\dfrac{U}{KS^2}$=1199.2227<br>F**≥$F_{0.01}$(9，10)<br>=5.26 |
| 总和 | U+Q=1.856e+009 | n-1=19 | | $\overline{R}^2$=0.9982 |

表3.3.3 方差与模型分析表

# 回归系数的显著性检验

- t检验是多元回归分析中对各个回归系数的检验，目的在于检验当其他变量不变时，该回归系数对应的自变量是否对因变量有显著影响。当 $|t| \geq t_{\alpha/2}(n-k)$ 成立时说明自变量Xi对因变量y的影响是显著的。
- 因为 $t_{0.01/2(10)}=3.1693$，$t_{0.05/2(10)}=2.2281$。所以自变量 $x_1$，$x_2$，$x_3$，$x_6$, $x_7$，$x_9$ 回归的t值都通过显著水平为5%的统计检验。变量 $x_4$，$x_5$，$x_8$ 没有通过t检验，这可能与我们对变量的赋值有关(见表3.3.4)。

# 回归系数的显著性检验

| | 回归系数 | 标准误 | t |
|---|---|---|---|
| 常数项 | -98676.0843 | 10586.2146 | -9.3212 |
| $x_1$ | 2.4037 | 0.1824 | 13.1753 |
| $x_2$ | 38754.3977 | 1899.2404 | 20.4052 |
| $x_3$ | 0.0858 | 0.0251 | 3.4149 |
| $x_4$ | 0.2161 | 0.2757 | 0.7837 |
| $x_5$ | -0.5463 | 1.5345 | -0.3560 |
| $x_6$ | -0.2292 | 0.0935 | -2.4527 |
| $x_7$ | -0.1640 | 0.0441 | -3.7195 |
| $x_8$ | -0.0433 | 0.0458 | -0.9471 |
| $x_9$ | -0.1527 | 0.0352 | -4.3367 |

表3.3.4 回归系数分析