第三章 回归分析

处理变量与变量之间的统计相关关系

星系 氢含量、色指数、光度
太阳 耀斑、黑子、太阳射电辐射流量

统计相关关系 不完全确定

观测误差 ↑ ↓ 深入了解

函数关系 完全确定

实质：概率统计 + 最小二乘

§ 一元线性回归

一 一元线性回归模型及参数估计

$$y_k = \beta_0 + \beta x_k + \varepsilon_k$$ 一元线性回归模型

$$E(y_k) = \beta_0 + \beta x_k \qquad \varepsilon_k \sim N(0,\sigma^2)$$

$$D(y_k) = \sigma^2$$ 正态误差回归模型

寻找β_0, β的好的估计值，得到最能描述y和x关系的回归直线

$$\hat{y}_k = b_0 + b x_k$$

利用最小二乘法给出b_0, b的计算公式

$$Q = \sum (y_k - \hat{y}_k)^2 = \sum (y_k - b_0 - b x_k)^2 = \min$$

$$\frac{\partial Q}{\partial b_0} = 0 \quad \rightarrow \quad b_0 = \frac{1}{n}\left(\sum y_k - b\sum x_k\right) = \bar{y} - b\bar{x}$$

$$\frac{\partial Q}{\partial b} = 0 \quad \rightarrow \quad b = \frac{\sum (x_k - \bar{x})(y_k - \bar{y})}{\sum (x_k - \bar{x})^2} = \frac{l_{xy}}{l_{xx}}$$

回归分析

$$E(b_0)=\beta_0$$

$$E(b)=\beta$$

$$D(b_0)=\sigma^2[\frac{1}{n}+\frac{\overline{x}^2}{\sum(x_k-\overline{x})^2}]$$

$$D(b)=\sigma^2[\frac{1}{\sum(x_k-\overline{x})^2}]$$

二　回归方程的显著性检验

$$\begin{aligned}\sum(y_k-\overline{y})^2&=\sum(y_k-\hat{y}_k+\hat{y}_k-\overline{y})^2\\&=\sum(y_k-\hat{y}_k)^2+\sum(\hat{y}_k-\overline{y})^2+2\sum(y_k-\hat{y}_k)(\hat{y}_k-\overline{y})\\&=\sum(y_k-\hat{y}_k)^2+\sum(\hat{y}_k-\overline{y})^2\\&=Q+U\end{aligned}$$

Q: 残差平方和　剩余平方和

U: 回归平方和　自变量变化引起

回归分析

1. 相关系数的检验

$$r=\sqrt{U/l_{yy}}=\sqrt{l_{xy}^{\ 2}/l_{xx}l_{yy}}=l_{xy}\Big/\sqrt{l_{xx}l_{yy}} \quad \Rightarrow \quad 0\le r\le 1$$

$|r|$大　y与x线性相关密切

$|r|$小　y与x线性相关较弱

$r=1$　y与x完全线性相关

$r=0$　y与x毫无线性关系

$r>0$　$b>0$　正相关

$r<0$　$b<0$　负相关

$r>r_\alpha$　r在α水平上显著

2. *F*检验（方差分析）

$$l_{yy}/\sigma^2 \sim \chi^2(n-1)$$

$$U/\sigma^2 \sim \chi^2(1)$$

$$Q/\sigma^2 \sim \chi^2(n-2)$$

$$\frac{U(n-2)}{Q} \sim F(1, n-2)$$

$F > F_\alpha(1, n-2)$ 拒绝域 回归方程显著

相关系数显著性检验 $\Leftrightarrow$ 回归方程的F检验

即 $r > r_\alpha \Leftrightarrow F > F_\alpha(1, n-2)$

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

$$F = \frac{U(n-2)}{Q} = \frac{(n-2)r^2}{1-r^2}$$

$$r = \sqrt{\frac{F}{(n-2)+F}}$$

$$r_\alpha = \sqrt{\frac{F_\alpha(1, n-2)}{(n-2)+F_\alpha(1, n-2)}}$$

三 回归系数和回归值的精度估计

β_0、β的区间估计

1. β的置信区间

1) σ已知

$$E(b)=\beta \quad D(b)=\sigma^2/l_{xx}$$

$$\Downarrow$$

$$b \sim N(\beta,\sigma^2/l_{xx})$$

$$\frac{b-\beta}{\sigma}\sqrt{l_{xx}} \sim N(0,1)$$

$$P(-u_\alpha < \frac{b-\beta}{\sigma}\sqrt{l_{xx}} < u_\alpha) = 1-\alpha$$

β的区间估计 $(b-\mu_\alpha\sigma/\sqrt{l_{xx}}, b+\mu_\alpha\sigma/\sqrt{l_{xx}})$

2) σ未知

$$S_y^2 = \hat{\sigma}^2 = Q/(n-2)$$

$$\frac{b-\beta}{S_y/\sqrt{l_{xx}}} \sim t(n-2)$$

$$\frac{b-\beta}{\sigma}\sqrt{l_{xx}} \sim N(0,1) \qquad \frac{Q}{\sigma^2} \sim \chi^2(n-2)$$

$$\Downarrow$$

$$\frac{b-\beta}{\sigma}\sqrt{l_{xx}} \Big/ \sqrt{\frac{Q/\sigma^2}{n-2}} \sim t(n-2)$$

而 $S_y^2 = Q/n-2$

有 $\dfrac{b-\beta}{S_y/\sqrt{l_{xx}}} \sim t(n-2)$

$$P\left(-t_\alpha(n-2) < \frac{b-\beta}{S_y/\sqrt{l_{xx}}} < t_\alpha(n-2)\right) = 1-\alpha$$

β的区间估计 $(b - t_\alpha S_y/\sqrt{l_{xx}}, b + t_\alpha S_y/\sqrt{l_{xx}})$

3. 回归值的置信区间

定义残差 $\delta_i = y_i - \hat{y}_i$

则

$$E(\delta_i) = E(\beta_0 + \beta x_i + \varepsilon_i - b_0 - bx_i) = 0$$

$$\begin{aligned}
D(\delta_i) &= D(y_i - b_0 - bx_i) \\
&= D[y_i - \bar{y} - b(x_i - \bar{x})] \\
&= D\left[y_i - \bar{y} - \sum_k \frac{(x_k - \bar{x})(x_i - \bar{x})}{\sum_j (x_j - \bar{x})^2} y_k \right] \\
&= D\left\{ y_i - \sum_k \left[\frac{1}{n} + \frac{(x_k - \bar{x})(x_i - \bar{x})}{\sum_j (x_j - \bar{x})^2} \right] y_k \right\} \\
&= \left[1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_j (x_j - \bar{x})^2} \right] \sigma^2
\end{aligned}$$

回归分析

$$\delta \sim N\left(0, \sigma\sqrt{1+\frac{1}{n}+\frac{(x_i-\bar{x})^2}{\sum_j (x_j-\bar{x})^2}}\right)$$

$$P(-\delta_n < y-\hat{y} < \delta_n) = 1-\alpha$$

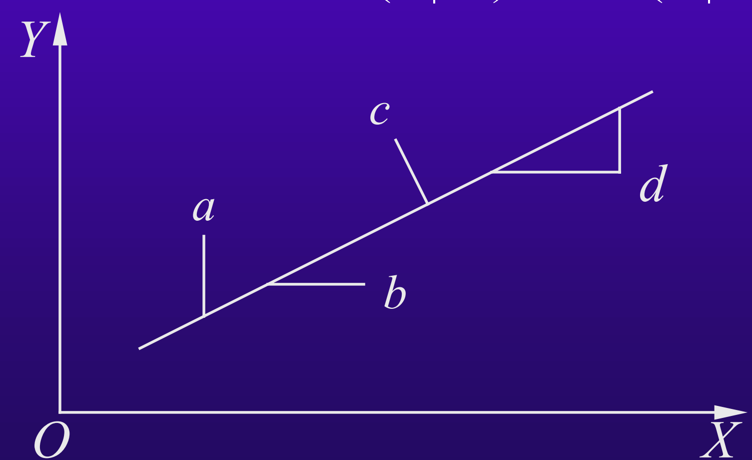
y的区间估计　$(y-\delta_n, \hat{y}+\delta_n)$

$$\delta_N = u_\alpha \sigma \sqrt{1+\frac{1}{n}+\frac{(x_i-\bar{x})^2}{\sum_j (x_j-\bar{x})^2}}$$

四 五种一元线性回归及其在天文上的应用

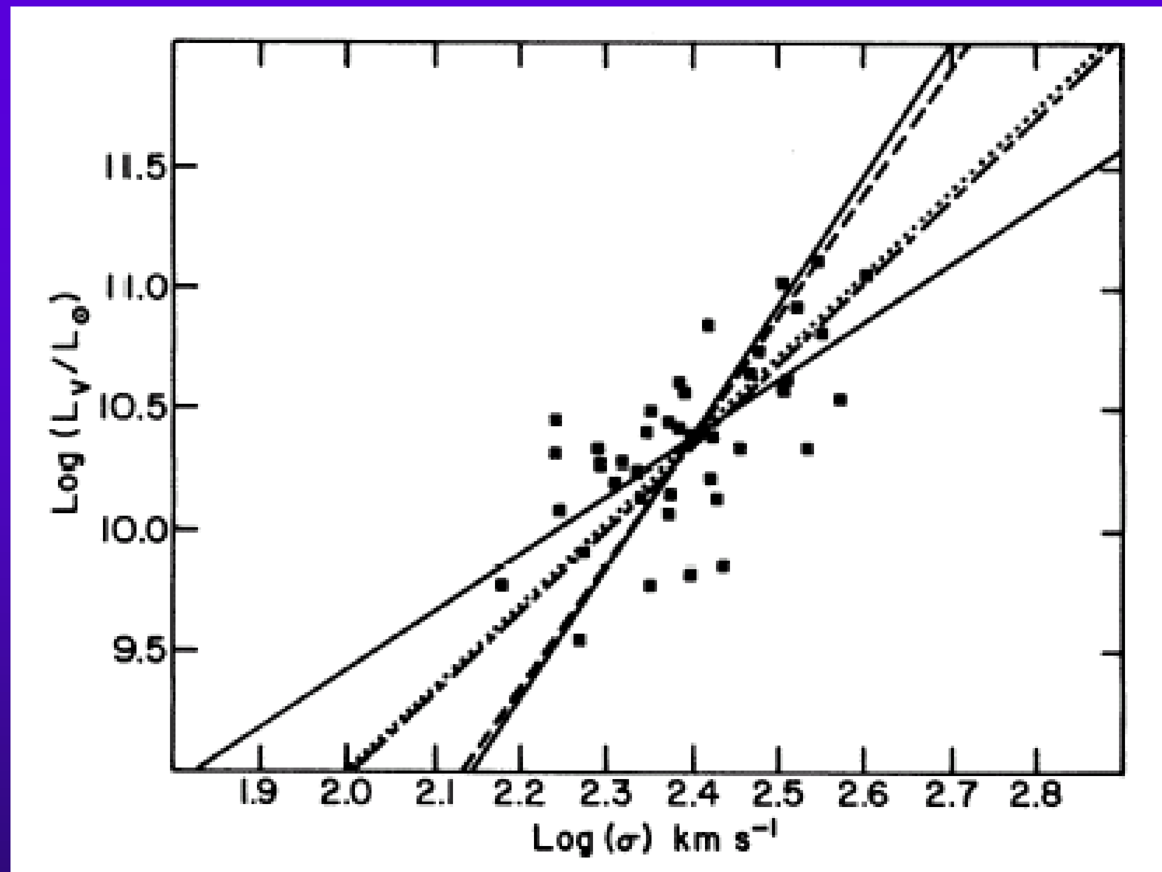
1. 五种线性回归方法

1) OLS($Y\,|\,X$): 观测点和回归直线上同一 x 的 y 的差;
2) 逆回归OLS($X\,|\,Y$): 观测点和回归直线上同一 y 的 x 的差;
3) 正交回归线OR : 观测点到回归线的垂直距离;
4) 简化主轴回归RMA : 观测点对回归线在垂直、水平两个方向测量的距离;
5) OLS平分线: OLS($Y\,|\,X$)和OLS($X\,|\,Y$)的平分线。



回归分析

应用五种回归方法测椭圆星系速度弥散σ和光学光度之间的关系$L\sim\sigma^n$



图：L和σ的对数散点图及它们的五种回归线：1. OLS($Y\,|\,X$)
2. OLS($X\,|\,Y$) 3. OLS平分线(点虚线) 4. OR(虚线) 5. RMA(点线)

§ 曲线回归分析

一 曲线回归类型的确定

1. 散点图

   利用观测数据的散点图，对比已知函数形式的各种曲线，选择最为接近的曲线作为回归函数

2. 多项式

   $$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_m x^m + \varepsilon$$

二 曲线回归参数的确定

$$\mathrm{I}\quad \begin{cases} y = \beta_0 + \beta\,\mathrm{e}^x \\ y = \beta_0 + \beta \ln x \\ y = \beta_0 + \beta x^l \end{cases} \quad \Rightarrow \quad y = \beta_0 + \beta x' \quad \begin{array}{l} x' = \mathrm{e}^x \\ x' = \ln x \\ x' = x^l \end{array}$$

回归分析

$$\mathrm{II}\quad \begin{cases} y=\dfrac{1}{\beta_0+\beta\,\mathrm{e}^{x}} \\ y=\beta_0\,\mathrm{e}^{\beta x} \\ y=\beta_0 x^{\beta} \end{cases} \quad\Rightarrow\quad y'=\beta_0{}'+\beta x' \qquad \begin{array}{l} y'=1/y \\ \begin{cases}\beta_0{}'=\ln\beta_0 \\ y'=\ln y\end{cases} \\ \begin{cases}\beta_0{}'=\ln\beta_0 \\ x'=\ln x \\ y'=\ln y\end{cases} \end{array}$$

$$\mathrm{III}\quad y=\mathrm{e}^{\beta_1 x}+\mathrm{e}^{\beta_2 x}$$

I、II进行变换，转化为线性回归；III泰勒级数展开，变为线性。

三　曲线回归的有效性检验

相关指数 $$R=\sqrt{1-\frac{\sum(y_i-\hat{y}_i)^2}{\sum(y_i-\overline{y})^2}}$$

标准剩余差 $$S_y=\sqrt{\frac{\sum(y_i-\hat{y}_i)^2}{n-2}}$$

§ 多元线性回归

一 模型参数估计

$$y = \beta_0 + \sum_{i=1}^{m} \beta_i x_i + \varepsilon \qquad \varepsilon \sim N(0, \sigma^2)$$

$$Y = \boldsymbol{\beta} X + \boldsymbol{\varepsilon}$$ 回归超平面

$$\hat{y} = b_0 + \sum_i b_i x_i$$ 回归方程

$$E(\hat{y}) = \beta_0 + \sum_i \beta_i x_i$$

$$Q = \sum_{k=1}^{n} (y_k - \hat{y}_k)^2 = \min$$

$$\begin{cases} \partial Q / \partial b_0 = 0 \\ \partial Q / \partial b_j = 0 \quad j = 1, 2, \cdots, m \end{cases}$$

回归分析

$$l_{ij} = \sum_{k=1}^{n} (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)$$

$$l_{iy} = \sum_{k=1}^{n} (x_{ki} - \bar{x}_i)(y_k - \bar{y})$$

$$\begin{cases} l_{11}b_1 + l_{12}b_2 + \cdots + l_{1m}b_m = l_{1y} \\ l_{21}b_1 + l_{22}b_2 + \cdots + l_{2m}b_m = l_{2y} \\ \cdots\cdots \\ l_{m1}b_1 + l_{m2}b_2 + \cdots + l_{mm}b_m = l_{my} \end{cases}$$

表示为矩阵形式 $\mathbf{LB} = \mathbf{L}_y$

$$\mathbf{C} = \mathbf{L}^{-1}$$

$$\mathbf{B} = \mathbf{L}^{-1}\mathbf{L}_y = \mathbf{C}\mathbf{L}_y$$

$$b_i = \sum_{j=1}^{m} c_{ij} l_{jy} \quad i = 1,2,\cdots,m$$

$$b_0 = \bar{y} - \sum_j b_j \bar{x}_j$$

标准化模型：对原数据进行标准化变换，而对变换后的数据建立的回归模型。 WHY？量纲不同

标准化变换 $y'_k = \dfrac{y_k - \bar{y}}{\sigma_y} \quad x'_{ki} = \dfrac{x_{ki} - \bar{x}_i}{\sigma_i}$

其中 $\sigma_y = \sqrt{l_{yy}} \quad \sigma_i = \sqrt{l_{ii}}$

$$\Downarrow$$

$$\begin{cases} \sigma'_y = 1 \\ \bar{y}' = 0 \end{cases} \quad \begin{cases} \sigma'_i = 1 \\ \bar{x}'_i = 0 \end{cases}$$

新的正规方程组

$$\begin{cases} l'_{11}b'_1 + l'_{12}b'_2 + \cdots + l'_{1m}b'_m = l'_{1y} \\ l'_{21}b'_1 + l'_{22}b'_2 + \cdots + l'_{2m}b'_m = l'_{2y} \\ \cdots\cdots \\ l'_{m1}b'_1 + l'_{m2}b'_2 + \cdots + l'_{mm}b'_m = l'_{my} \end{cases}$$

其中 $l'_{ij} = \sum_{k=1}^{n}(x'_{ki} - \bar{x}'_i)(x'_{kj} - \bar{x}'_j) = \sum_{k=1}^{n} x'_{ki}x'_{kj} = \dfrac{l_{ij}}{\sqrt{l_{ii}l_{jj}}}$

回归分析

记 $r_{ij}=l'_{ij}$　　正规方程组可表示为矩阵形式

$$\mathbf{R}=(r_{ij})$$

$$\mathbf{R}\mathbf{B}'=\mathbf{R}_y$$

$$\mathbf{C}'=(\mathbf{R})^{-1}$$

$$\mathbf{B}'=\mathbf{C}'\mathbf{R}_y$$

$$b'_i=\sum_{j=1}^{m}c'_{ij}r_{jy}\qquad b'_0=0$$

标准回归方程　$\hat{y}'_k=\sum_{i=1}^{m}b'_i x'_{ki}$

即　$$\frac{\hat{y}_k-\bar{y}_k}{\sigma_y}=\sum_{i=1}^{m}b'_i\frac{x_{ki}-\bar{x}_i}{\sigma_i}=\sum_{i=1}^{m}b'_i\frac{x_{ki}}{\sigma_i}-\sum_{i=1}^{m}b'_i\frac{\bar{x}_i}{\sigma_i}$$

$$\hat{y}_k=\bar{y}_k+\sum_{i=1}^{m}b'_i\frac{\sigma_y}{\sigma_i}x_{ki}-\sum_{i=1}^{m}b'_i\frac{\sigma_y}{\sigma_i}\bar{x}_i$$

$$\therefore\ b_i=b'_i\frac{\sigma_y}{\sigma_i}\qquad b_0=\bar{y}_k-\sum_{i=1}^{m}b'_i\frac{\sigma_y}{\sigma_i}\bar{x}_i$$

二　多元回归效果检验

1.　回归方程的显著检验

原假设　$H_0:\quad \beta_i = 0 \quad i = 1 \sim m$

$$l_{yy} = \sum (y_k - \bar{y})^2 = \sum (y_k - \hat{y}_k)^2 + \sum (\hat{y}_k - \bar{y})^2 = Q + U$$

$$U/\sigma^2 \sim \chi^2(m) \quad Q/\sigma^2 \sim \chi^2(n-m-1)$$

$$F = \frac{U/m}{Q/(n-m-1)} \sim F(m, n-m-1)$$

$$F > F_\alpha \quad \text{拒绝}H_0;\quad F < F_\alpha \quad \text{接受}H_0$$

回归平方和、剩余平方和的计算

$$U = \sum_{i=1}^{m} b_i l_{iy} \quad Q = l_{yy} - \sum_{i=1}^{m} b_i l_{iy}$$

剩余标准差　$S_y = \sqrt{Q/(n-m-1)}$

复相关系数　$R = \sqrt{U/l_{yy}}$

2. 回归系数的显著性检验

偏回归平方和 $p_j = U^m - U_j^{m-1}$

U_j^{m-1}：去除x_j后的回归平方和

原假设 $H_0: \quad \beta_j = 0$

$$p_j / \sigma^2 \sim \chi^2(1) \qquad Q / \sigma^2 \sim \chi^2(n-m-1)$$

$$F = \frac{p_j}{Q/(n-m-1)} \sim F(1, n-m-1)$$

$F > F_\alpha$ 拒绝H_0； $F < F_\alpha$ 接受H_0

剔除x_k后回归系数

$$\begin{cases} b_j^* = b_j - \dfrac{c_{kj}}{c_{kk}} b_k & j \neq k \\ b_0^* = \bar{y} - \sum\limits_j b_j^* \bar{x}_j & \end{cases}$$

$$p_j = \frac{b_j^2}{c_{jj}} \qquad j = 1 \sim m$$

回归分析

证：

$$p_k = U^m - U_k^{m-1} = \sum_{i=1}^{m} b_i l_{iy} - \sum_{\substack{i=1\\ i\neq k}}^{m} b_i^* l_{iy}$$

$$= \sum_{i=1}^{m} b_i l_{iy} - \sum_{\substack{i=1\\ i\neq k}}^{m} (b_i - \frac{c_{ki}}{c_{kk}} b_k) l_{iy}$$

$$= b_k l_{ky} + \frac{b_k}{c_{kk}} \sum_{\substack{i=1\\ i\neq k}}^{m} c_{ki} l_{iy}$$

$$= \frac{b_k}{c_{kk}} c_{kk} l_{ky} + \frac{b_k}{c_{kk}} \sum_{\substack{i=1\\ i\neq k}}^{m} c_{ki} l_{iy}$$

$$= \frac{b_k}{c_{kk}} \sum_{i=1}^{m} c_{ki} l_{iy}$$

$$= \frac{b_k^2}{c_{kk}}$$

3. 回归系数的精度估计

$$E(b_i)=\beta_i$$

$$D(b_i)=c_{ii}\sigma^2$$

$$b_i\sim N(\beta_i,c_{ii}\sigma^2)$$

$$\frac{b_i-\beta_i}{\sqrt{c_{ii}}\sigma}\sim N(0,1)$$

三 残差检验

1. 残差图分析

残差$v_k=y_k-\hat{y}_k$ 作$v_k\sim x_{ki}$图

$$\begin{cases}\text{模型不适（图3.7b）}\\ \text{方差不等（图3.7c）}\\ \text{缺自变量（图3.7d）}\end{cases}$$

2. 残差的统计检验

1) 等方差检验　将观测值分成两段，分别拟合利用 F 检验方差是否相等

2) 正态检验　a. 残差直方图　b. χ^2拟合检验

3) 随机检验——游程检验

符号序列　$+$　$-$

n_1　n_2

游程总数***R***

$$+ + - - - + - - + + + - - + \qquad R=7$$

概率函数

$$P(R=2k)=\frac{2\,C_{n_1-1}^{k-1}C_{n_2-1}^{k-1}}{C_{n_1+n_2}^{n_1}}$$

$$P(R=2k+1)=\frac{C_{n_1-1}^{k}C_{n_2-1}^{k-1}+C_{n_1-1}^{k-1}C_{n_2-1}^{k}}{C_{n_1+n_2}^{n_1}}$$

回归分析

$$\langle R\rangle = \frac{2n_1 n_2}{n_1 + n_2} + 1$$

$$\sigma^2(R) = \frac{2n_1 n_2(2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}$$

$R_{1\alpha}$ $\quad P(R \le R_1) = \alpha$ $\quad$ 最大整数R_1

$R_{2\alpha}$ $\quad P(R \ge R_2) = \alpha$ $\quad$ 最小整数R_2

$R < R_{1\alpha}$ $\quad R > R_{2\alpha}$ $\quad$ 拒绝残差序列为随机的假设

样本的游程检验: 确定中位数m， $x>m$记为+

$x<m$记为−

$$n_1, n_2 > 10 \quad Z = \frac{R - \langle R\rangle + \frac{1}{2}}{\sigma^2(R)} \sim N(0,1)$$

§ 逐步回归分析

一 基本思想

最优回归方程：1. 包含的自变量是显著的

2. 剩余标准差小

基本思想：根据对因变量影响的大小，逐个引入自变量到回归方程，进行显著性检验加以选择 多元回归

二 线性方程组的求解求逆紧凑变换

解正规方程组 → 回归系数

系数矩阵逆矩阵 → 偏回归平方和 显著检验

逐步回归实质：无回代消去法 + 挑选重要变量

求解求逆并行

回归分析

方程组
$$
\begin{cases}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = a_{1n+1} \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = a_{2n+1} \\
\cdots\cdots \\
a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = a_{nn+1}
\end{cases}
$$

其增广矩阵
$$
\mathbf{A}^{(0)} = \begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} & a_{1n+1} \\
a_{21} & a_{22} & \cdots & a_{2n} & a_{2n+1} \\
\vdots & & \ddots & & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{nn} & a_{nn+1}
\end{bmatrix}
$$

进行初等变换
$$
a_{ij}^{(1)} = a_{ij} - a_{ik}a_{kj} / a_{kk} \qquad i \neq k, j = 1 \sim n+1
$$
$$
a_{kj}^{(1)} = a_{kj} / a_{kk} \qquad j = 1 \sim n+1
$$

回归分析

得到矩阵 $$\mathbf{A}^{(1)}=\begin{bmatrix} a_{11}^{(1)} & \cdots & 0 & \cdots & a_{1n+1}^{(1)} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{k1}/a_{kk} & \cdots & 1 & \cdots & a_{kn+1}/a_{kk} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n11}^{(1)} & \cdots & 0 & \cdots & a_{nn+1}^{(1)} \end{bmatrix}$$

对$\mathbf{A}^{(0)}$进行第l次变换

$$\begin{cases} a_{ij}^{(l+1)} = a_{ij}^{(l)} - a_{ik}^{(l)} a_{kj}^{(l)} / a_{kk}^{(l)} & i \neq k, j = 1 \sim n+1 \\ a_{kj}^{(l+1)} = a_{kj}^{(l)} / a_{kk}^{(l)} & j = 1 \sim n+1 \end{cases}$$

第n步后，得到增广矩阵

$$\mathbf{A}^{(n)}=\begin{bmatrix} 1 & 0 & \cdots & 0 & a_{1n+1}^{(n)} \\ 0 & 1 & \cdots & 0 & a_{2n+1}^{(n)} \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & 1 & a_{nn+1}^{(n)} \end{bmatrix}$$

回归分析

$$
\mathbf{A}^{(0)}=\begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} & a_{1n+1} & 1 & 0 & \cdots & 0\\
a_{21} & a_{22} & \cdots & a_{2n} & a_{2n+1} & 0 & 1 & \cdots & 0\\
\vdots & & \ddots & & \vdots & \vdots & & \ddots & \vdots\\
a_{n1} & a_{n2} & \cdots & a_{nn} & a_{nn+1} & 0 & 0 & \cdots & 1
\end{bmatrix}
$$

$$\Big\Downarrow\ n次变换$$

$$
\mathbf{A}^{(n)}=\begin{bmatrix}
1 & 0 & \cdots & 0 & a_{1n+1}^{(n)} & a_{1n+2}^{(n)} & & a_{12n}^{(n)} & a_{12n+1}^{(n)}\\
0 & 1 & \cdots & 0 & a_{2n+1}^{(n)} & a_{2n+2}^{(n)} & & a_{22n}^{(n)} & a_{22n+1}^{(n)}\\
\vdots & & \ddots & & \vdots & & & & \\
0 & 0 & \cdots & 1 & a_{nn+1}^{(n)} & a_{nn+2}^{(n)} & & a_{n2n}^{(n)} & a_{n2n+1}^{(n)}
\end{bmatrix}
$$

$$
x_j=a_{jn+1}^{(n)} \qquad \mathbf{A}^{-1}=\begin{bmatrix}
a_{1n+2}^{(n)} & a_{1n+3}^{(n)} & \cdots & a_{12n}^{(n)} & a_{12n+1}^{(n)}\\
a_{2n+2}^{(n)} & a_{2n+3}^{(n)} & \cdots & a_{22n}^{(n)} & a_{22n+1}^{(n)}\\
\vdots & & \ddots & & \vdots\\
a_{nn+2}^{(n)} & a_{nn+3}^{(n)} & \cdots & a_{n2n}^{(n)} & a_{n2n+1}^{(n)}
\end{bmatrix}
$$

回归分析

对 $\mathbf{A}^{(l-1)}$ 作消去x_k的变换前，第$n+1+k$列的元素为

$$\begin{cases} a_{kn+1+k}^{(l-1)} = 1 \\ a_{in+1+k}^{(l-1)} = 0 \quad i = 1 \sim n, i \neq k \end{cases}$$

对 $\mathbf{A}^{(l-1)}$ 作消去x_k的变换后，第$n+1+k$列的元素为

$$\begin{cases} a_{kn+1+k}^{(l)} = 1 \big/ a_{kk}^{(l-1)} \\ a_{in+1+k}^{(l)} = -a_{ik}^{(l-1)} \big/ a_{kk}^{(l-1)} \quad i = 1 \sim n, i \neq k \end{cases}$$

可将 $\mathbf{A}^{(l)}$ 第$n+1+k$列的元素放到第k列上

$$\begin{cases} a_{kk}^{(l)} = 1 \big/ a_{kk}^{(l-1)} \\ a_{ik}^{(l)} = -a_{ik}^{(l-1)} \big/ a_{kk}^{(l-1)} \quad i = 1 \sim n, i \neq k \end{cases}$$

$$\begin{cases} a_{ij}^{(l)} = a_{ij}^{(l-1)} - a_{ik}^{(l-1)} a_{kj}^{(l-1)} \big/ a_{kk}^{(l-1)} & i \neq k, j \neq k \\ a_{kj}^{(l)} = a_{kj}^{(l-1)} \big/ a_{kk}^{(l-1)} & j \neq k \\ a_{ik}^{(l)} = -a_{ik}^{(l-1)} \big/ a_{kk}^{(l-1)} & i \neq k \\ a_{kk}^{(l)} = 1 \big/ a_{kk}^{(l-1)} & \end{cases}$$

节省内存单元

求解求逆紧凑变换

回归分析

$$\mathbf{A}^{(l)}=L_k\mathbf{A}^{(l-1)}$$

性质： 1. 对$\mathbf{A}^{(0)}$作消去$x_k(k=k_1,k_2,\ldots,k_l)$的变换得到$\mathbf{A}^{(l)}$，则$\mathbf{A}^{(l)}$与消去变量$x_k$的次序无关；

2. $L_k(L_k\mathbf{A}^{(l)})=\mathbf{A}^{(l)}$。对$\mathbf{A}^{(l)}$连续施行两次同样的变换结果不变；

3. 对原方程组每经过一次变换L_k，就可以得到子方程组的解及该子方程系数矩阵的逆矩阵。

三 计算步骤

1. 计算相关阵

1) 计算均值$\bar{x}$、$\bar{y}$

2) 标准化 $x'_{ki}=\dfrac{x_{ki}-\bar{x}_i}{\sigma_i}$ $y'_k=\dfrac{y_k-\bar{y}}{\sigma_y}$

3) 计算相关阵 $l_{ij}=\sum_{k=1}^{n}(x_{ki}-\bar{x}_i)(x_{kj}-\bar{x}_j)$ $r_{ij}=\dfrac{l_{ij}}{\sqrt{l_{ii}l_{jj}}}$

2. 逐步计算

1) 挑选变量（已计算l步）

I. 计算所有未选变量的偏回归平方和

$$p_i^{(l)} = [r_{iy}^{(l)}]^2 / r_{ii}^{(l)}$$

II. 引进变量检验

$$Q^{(l+1)} = Q^{(l)} - p_k^{(l)} = r_{yy}^{(l)} - p_k^{(l)}$$

$$F_1 = \frac{p_k^{(l)}[n-(l+1)-1]}{r_{yy}^{(l)} - p_k^{(l)}}$$

$F_1 < F_\alpha(1, n-l-2)$，结束挑选；

$F_1 > F_\alpha(1, n-l-2)$，引进变量。

2) 剔除变量

I. 计算已选变量的偏回归平方和

$$p_i^{(l+1)} = [r_{iy}^{(l+1)}]^2 / r_{ii}^{(l+1)}$$

II. 剔除变量检验

$$Q^{(l+1)} = r_{yy}^{(l)} - p_k^{(l)} = r_{yy}^{(l)} - [r_{ky}^{(l+1)}]^2 / r_{kk}^{(l+1)} = r_{yy}^{(l+1)}$$

$$F_2 = \frac{p_k^{(l+1)}(n-l-2)}{r_{yy}^{(l+1)}}$$

$F_2 > F_\alpha(1,n–l–2)$，结束剔除；

$F_2 < F_\alpha(1,n–l–2)$，继续剔除检验。

3. 计算回归结果

1) 标准回归系数 $b_i' = r_{iy}^{(l)}$

2) 剩余平方和 $Q'^{(l)} = r_{yy}^{(l)}$

3) 实际回归系数 $b_i = b_i' \sigma_y / \sigma_i \quad b_0 = \bar{y} - \sum_{i=1}^{m} b_i \bar{x}_i$

4) m元回归方程 $y = b_0 + \sum_{i=1}^{m} b_i x_i$

5) 实际剩余平方和 $Q = \sigma_y^2 Q' = \sigma_y^2 r_{yy}^{(l)}$

6) 回归方程标准差 $S_y = \sigma_y \sqrt{r_{yy}^{(l)} / (n-m-1)}$

7) 回归系数标准差 $S_i = S_y \sqrt{r_{ii}^{(l)}} / \sigma_i$

4. 几点说明

1) 临界值F_α的选取

   多选变量，F_1、F_2小；少选变量，F_1、F_2大

2) 控制计算量

   R_{ii}不能等于或接近0

3) 没有唯一的最优子集
4) 不合理的“最优”子集