回归分析

公式

线性(非线性)模型与Y和X关系无关,指的是Y与参数的关系.

tip: $E[\vec{x}]=\vec{\mu}$, $Cov(\vec{x})=\Sigma$, 则 $E[x^TAx]=\mu^TA\mu+tr(A\Sigma)$

回归参数的OLS估计

定义:涉及p-1个预测变量的多元线性回归模型为 $y=\beta_0+\beta_1x_1+\cdots+\beta_{p-1}x_{p-1}+e$. e为随机误差项, β_0为截距, $\beta_1,\cdots,\beta_{p-1}$为回归系数

tip:对随机误差项e,有Gauss-Markov假设:① $E[e_i]=0$ ② $Var(e_i)=\sigma^2$ (等方差) ③ $Cov(e_i,e_j)=0$;对连续型数据,常假设 $e_1,\cdots,e_n$ i.i.d. $N(0,\sigma^2)$

矩阵表示:样本容量为n, $\vec{y}=(y_1,\cdots,y_n)^T$, $\vec{\beta}=(\beta_0,\cdots,\beta_{p-1})^T$, $\vec{e}=(e_1,\cdots,e_n)^T$.

$$X=\begin{pmatrix}x_{11} & x_{12} & \cdots & x_{1,p-1}\\ x_{21} & x_{22} & \cdots & x_{2,p-1}\\ \vdots & \vdots & & \vdots\\ x_{n1} & x_{n2} & \cdots & x_{n,p-1}\end{pmatrix}$$
.多元线性回归模型表示为 $\vec{y}=X\vec{\beta}+\vec{e}$, $E[\vec{e}]=0$, $Cov(\vec{e})=\sigma^2I_n$

因变量 y_i 的平均的估计为方程: $\hat{Y}=\hat{\beta}_0+\hat{\beta}_1X_1+\cdots+\hat{\beta}_{p-1}X_{p-1}$, $\hat{y}_i$ 称为 y_i 的拟合值. $\hat{e}_i=y_i-\hat{y}_i$ 称残差,最小二乘估计思想为找到使得 $\sum_{i=1}^n\hat{e}_i^2$ 最小的参数值

即 $Q(\vec{\beta})=(y-X\beta)^T(y-X\beta)$ 最小. 对β偏导为0 $\Rightarrow \vec{\beta}=(X^TX)^{-1}X^T\vec{y}$ 　 $\frac{\partial}{\partial\beta}(\beta^TA\beta)=(A+A^T)\vec{\beta}$ 　 $\frac{\partial}{\partial\beta}(a^T\beta)=\frac{\partial}{\partial\beta}(\beta^Ta)=a$.

符号说明:

$m_j=\frac{1}{n}\sum_{i=1}^n x_{ij}$	X_j的样本平均值
$\bar{y}=\frac{1}{n}\sum_{i=1}^n y_i$	Y的样本平均值
$SX_jY=\sum_{i=1}^n(x_{ij}-m_j)(y_i-\bar{y})=\sum_{i=1}^n x_{ij}y_i-\frac{1}{n}\sum_{i=1}^n x_{ij}\sum_{i=1}^n y_i$	X_j与Y的校正叉积和
$SX_jX_j=\sum_{i=1}^n(x_{ij}-m_j)^2=\sum_{i=1}^n x_{ij}^2-n\cdot m_j^2$	X_j的校正平方和
$SYY=\sum_{i=1}^n(y_i-\bar{y})^2=\sum_{i=1}^n y_i^2-n\bar{y}^2$	Y的校正平方和,亦称总平方和

例:简单线性回归 $y_i=\beta_0+\beta_1x_i+e_i$ $i=1,\ldots,n$. 仿β_0,β_1的OLS估计

定义 $Q(\beta_0,\beta_1)=\sum_{i=1}^n(y_i-\beta_0-\beta_1x_i)^2$, 有
$$\begin{cases}\frac{\partial Q}{\partial\beta_0}=-2\sum(y_i-\beta_0-\beta_1x_i)=0\\ \frac{\partial Q}{\partial\beta_1}=-2x_i\sum(y_i-\beta_0-\beta_1x_i)=0\end{cases}\Rightarrow\begin{cases}\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}\\ \hat{\beta}_1=\frac{\sum x_iy_i-n\bar{x}\bar{y}}{\sum x_i^2-n\bar{x}^2}\end{cases}$$

中心化

定义:记 $m_j=\frac{1}{n}\sum_{i=1}^n x_{ij}$ $j=1,2,\cdots,p-1$ 为第j个自变量的样本均值,则 $y_i=\alpha+\beta_1(x_{i1}-m_1)+\beta_2(x_{i2}-m_2)+\cdots+\beta_{p-1}(x_{i,p-1}-m_{p-1})+e_i$ 称为 $y_i=\alpha+\beta_1x_{i1}+\beta_2x_{i2}+\cdots+\beta_{p-1}x_{i,p-1}+e_i$ 的中心化

其中 $\alpha=\beta_0+\beta_1m_1+\cdots+\beta_{p-1}m_{p-1}$

矩阵表示:样本容量为n, $\vec{y}=(y_1,\cdots,y_n)^T$, $\vec{\beta_c}=(\beta_1,\cdots,\beta_{p-1})^T$, $\vec{e}=(e_1,\cdots,e_n)^T$.
$$X_c=\begin{pmatrix}x_{11}-m_1 & x_{12}-m_2 & \cdots & x_{1,p-1}-m_{p-1}\\ x_{21}-m_1 & x_{22}-m_2 & \cdots & x_{2,p-1}-m_{p-1}\\ \vdots & \vdots & & \vdots\\ x_{n1}-m_1 & x_{n2}-m_2 & \cdots & x_{n,p-1}-m_{p-1}\end{pmatrix}$$
.多元线性回归模型表示为 $\vec{y}=\alpha I_n+X_c\vec{\beta}+\vec{e}$

OLS: $Q(\alpha,\beta)=(\vec{y}-\alpha I_n-X_c\vec{\beta_c})^T(\vec{y}-\alpha I_n-X_c\vec{\beta_c})\Rightarrow\begin{cases}\hat{\alpha}=\bar{y}\\ \hat{\vec{\beta}}_c=(X_c^TX_c)^{-1}X_c^T\vec{y} \quad \hat{\beta}_0=\hat{\alpha}-\hat{\vec{\beta}}_c^T\vec{m}\end{cases}$

校正叉积矩阵: $X_c^TX_c=\begin{pmatrix}SX_1X_1 & SX_1X_2 & \cdots & SX_1X_{p-1}\\ \vdots & \vdots & & \\ SX_{p-1}X_1 & SX_{p-1}X_2 & \cdots & SX_{p-1}X_{p-1}\end{pmatrix}$

一元时为 $\frac{SXY}{SXX}$

例:若 $\bar{x}=202.9529$, $\bar{y}=139.60588$, $SXX=530.78235$, $SXY=475.2957$, $SYY=427.7628$. 仿生简单线性回归模型在中心化下的回归参数OLS估计

→ $\hat{\beta}_1=\frac{SXY}{SXX}=0.895$. $\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}=-42.131$. 则 $\hat{y}=-42.131+0.895x$

例:

例1.3 对某种商品销售情况进行调查,考察销售总额Y与居民可支配收入X_1,该类商品的价格指数X_2和其它商品平均价格指数X_3的线性回归关系。共有$n=21$组样本,样本均值如下表

$\bar{y}$	m_1	m_2	m_3
14.27	13.63	12.25	10.42

增广矩阵

$$\mathbf{A}=\begin{pmatrix}\mathbf{X}_c^T\mathbf{X}_c & \mathbf{X}_c^T\vec{y}\\ \vec{y}^T\mathbf{X}_c & SYY\end{pmatrix}=\begin{bmatrix}360.94 & 242.83 & 71.21 & 411.76\\ 242.83 & 168.29 & 48.07 & 279.82\\ 71.21 & 48.07 & 14.65 & 81.05\\ 411.76 & 279.82 & 81.05 & 479.64\end{bmatrix}$$

校正差积阵的逆矩阵

$$\left(\mathbf{X}_c^T\mathbf{X}_c\right)^{-1}=\begin{bmatrix}0.146 & -0.127 & -0.292\\ -0.127 & 0.205 & -0.054\\ -0.292 & -0.054 & 1.667\end{bmatrix}$$

请利用最小二乘估计法给出如下模型的拟合回归方程

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+e,\quad \vec{e}\sim\mathbf{N}(\mathbf{0},\sigma^2I_n),\sigma^2\text{未知.}$$

$\hat{\alpha}=\bar{y}=14.27$

$\hat{\vec{\beta}}_c=(X_c^TX_c)^{-1}X_c^T\vec{y}=\begin{pmatrix}0.839\\0.583\\-0.459\end{pmatrix}$ 　 $\beta_0=\hat{y}-\hat{\beta}_1\bar{x}_1-\hat{\beta}_2\bar{x}_2-\hat{\beta}_3\bar{x}_3$

$\beta_0=\hat{\alpha}-\hat{\vec{\beta}}_c^T\vec{m}=0.479$ ☆

$\Rightarrow \hat{Y}=0.479+0.839X_1+0.583X_2-0.459X_3$

OLS估计的性质

$Cov(\mathbf{A}\vec{Z}+\vec{b})=\mathbf{A}Cov(\vec{Z})\mathbf{A}^T$.

定理:对于线性回归模型 $Y=X\beta+e$, $E[e]=0$, $Cov(e)=\sigma^2I_n$, 其OLS估计 $\hat{\beta}=(X^TX)^{-1}X^T\vec{y}$ 有性质:① $E[\hat{\beta}]=\beta$ ② $Cov(\hat{\beta})=\sigma^2(X^TX)^{-1}$

推论: $\vec{c}=(c_0,\cdots,c_{p-1})^T$, $E[\vec{c}\hat{\beta}]=\vec{c}\beta$. $Cov(\vec{c}\hat{\beta})=\sigma^2c(X^TX)^{-1}c^T$

Gauss-Markov定理:对于线性回归模型 $Y=X\beta+e$, 在$\vec{\beta}$的线性函数$\vec{c}\vec{\beta}$的所有线性无偏估计中, OLS估计$\vec{c}\hat{\beta}$是唯一有最小方差的估计

OLS性质:① $\frac{1}{n}\sum_{i=1}^n\hat{y}_i=\bar{y}$ ② $\hat{Y}=\hat{\beta}_0+\hat{\beta}_1X_1+\cdots+\hat{\beta}_{p-1}X_{p-1}$ 过点 $(m_1,\cdots,m_{p-1},\bar{y})$ ③ $\sum_{i=1}^n\hat{e}_i=0$, $\sum_{i=1}^n\hat{y}_i\hat{e}_i=0$

定义:残差平方和为 $RSS=\sum_{i=1}^n\hat{e}_i^2=\sum_{i=1}^n(y_i-\hat{y}_i)^2$

定理:① $RSS=y^T(I_n-X(X^TX)^{-1}X^T)y$ ② $\hat{\sigma}^2=\frac{RSS}{n-p}$ 是σ^2的无偏估计

定理:对于线性回归模型 $Y=X\beta+e$, 若 $e\sim N_n(0,\sigma^2I_n)$ 则有如下性质

① $\hat{\beta}\sim N_n(\beta,\sigma^2(X^TX)^{-1})$ ② $\frac{RSS}{\sigma^2}\sim\chi^2_{n-p}$ ③ $\hat{\beta}$与RSS相互独立 　 $\hat{\sigma}^2=\frac{RSS}{n-p}$

定理:对中心化回归模型 $\vec{y}=\alpha I_n+X_c\vec{\beta}+\vec{e}$ ① $E[\hat{\alpha}]=\alpha$, $E[\hat{\beta}_c]=\vec{\beta}_c$ 其中 $\hat{\alpha}=\bar{y}$, $\hat{\beta}_c=(X_c^TX_c)^{-1}X_c^T\vec{y}$

② $Cov\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}_c\end{pmatrix}=\sigma^2\begin{pmatrix}\frac{1}{n} & \\ & (X_c^TX_c)^{-1}\end{pmatrix}$ ③ 若 $e\sim N(0,\sigma^2I_n)$, 则 $\hat{\alpha}\sim N(\alpha,\frac{\sigma^2}{n})$, $\hat{\beta}_c\sim N_{p-1}(\vec{\beta}_c,\sigma^2(X_c^TX_c)^{-1})$

正态分布补充:

1. $\mu\in\mathbb{R}^n$, Σ为$n\times n$正定阵. $\vec{X}\sim N_n(\vec{\mu},\Sigma)$ 是指 $f(\vec{x})=\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\}$
   有 $E[\vec{x}]=\vec{\mu}$, $Cov(\vec{x})=\Sigma$ 　 $Cov(\vec{a}^T\vec{x},\vec{b}^T\vec{x})=\vec{a}^TCov(\vec{x})\vec{b}$
2. $\vec{Y}=A\vec{x}+\vec{b}$. 则 $\vec{Y}\sim N_n(A\vec{\mu}+\vec{b},A\Sigma A^T)$
3. $rank(A)=r$. $A^2=A$ 时 $\vec{x}^TA\vec{x}\sim\chi^2_r$
4. $X\in N_n(0,I)$, $BA=0$. 则 $B\vec{x}$ 与 $\vec{x}^TA\vec{x}$ 相互独立

定理 2.2.1 设 $E(\boldsymbol{X})=\boldsymbol{\mu}$, $Cov(\boldsymbol{X})=\boldsymbol{\Sigma}$, 则
$$E(\boldsymbol{X}'A\boldsymbol{X})=\boldsymbol{\mu}'A\boldsymbol{\mu}+tr(A\boldsymbol{\Sigma}).$$

定理 2.4.1 设 $\boldsymbol{X}\sim N_n(\boldsymbol{0},\boldsymbol{\Sigma})$, $\boldsymbol{\Sigma}$ 为正定阵,则 $\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\sim\chi^2_n$.

↑ 正则取 $Y=\Sigma^{-\frac{1}{2}}X$.

定义:拟合优度/测定系数 $R^2=\frac{ESS}{SYY}$, $ESS=\sum_{i=1}^n(\hat{y}_i-\bar{y}_i)^2$, $RSS=\sum_{i=1}^n(\hat{y}_i-y_i)^2$, $RSS+ESS=SYY$. R^2越大,拟合效果越好 ($R^2=1-\frac{RSS}{SYY}$)

例 3.2.1 根据经验知,在人的身高相等的条件下,其血压的收缩压 y 与体重 x_1、年龄 x_2 有关.现收集了 13 名男子的测量数据,见表(3.2.1).试建立 y 关于 x_1,x_2 的线性回归方程.

表 3.2.1 血压数据

序号	x_{i1}	x_{i2}	y_i	序号	x_{i1}	x_{i2}	y_i
1	152	50	120	8	158	50	125
2	183	20	141	9	170	40	132
3	171	20	124	10	153	55	123
4	165	30	126	11	164	40	132
5	158	30	117	12	190	40	155
6	161	50	125	13	185	20	147
7	149	60	123				

利用中心化模型

$$y_i=\alpha+(x_{i1}-\bar{x}_1)\beta_1+(x_{i2}-\bar{x}_2)\beta_2+e_i,\ i=1,\cdots,13.$$

由表 3.2.1 中数据计算可得

$$\bar{x}_1=\frac{1}{13}\sum_{i=1}^{13}x_{i1}=166.8$$
$$\bar{x}_2=\frac{1}{13}\sum_{i=1}^{13}x_{i2}=38.85$$
$$\bar{y}=\frac{1}{13}\sum_{i=1}^{13}y_i=130$$

中心化设计矩阵

$$X_c=\begin{bmatrix}-14.08 & 11.15\\ 16.92 & -18.85\\ 4.92 & -18.85\\ -1.08 & -8.85\\ -8.08 & -8.85\\ -5.08 & 11.15\\ -17.08 & 21.15\\ -8.08 & 11.15\\ 3.92 & 1.15\\ -13.08 & 16.15\\ -2.08 & 1.15\\ 23.92 & 1.15\\ 18.92 & -18.85\end{bmatrix}$$

$ESS=\frac{(SXY)^2}{SXX}$

$RSS=SYY-\frac{(SXY)^2}{SXX}$

中心化正则方程组 $X_c'X_c\boldsymbol{\beta}=X_c'y$ 为

$$\begin{cases}2\,078.92\beta_1-1\,533.85\beta_2=1\,607.00\\ -1\,533.85\beta_1+2\,307.69\beta_2=-715.00\end{cases}$$

解得

$$\hat{\beta}_1=1.068,\quad \hat{\beta}_2=0.400.$$

$\hat{\beta}=(X_c^TX_c)^{-1}X_c^T\vec{y}$

又 $\hat{\alpha}=\bar{y}=130$,

所以得到经验回归方程为

$$\hat{Y}=\hat{\alpha}+(x_1-\bar{x}_1)\hat{\beta}_1+(x_2-\bar{x}_2)\hat{\beta}_2=130+1.068(x_1-166.8)+0.400(x_2-38.85)=-62.963+1.068x_1+0.400x_2.$$

我们还可以计算出判定系数:由于 $SS_{回}=\hat{\beta}'X_c'y=1\,430.276$, $SS_{总}=\sum_{i=1}^{13}(y_i-\bar{y}_i)^2=1\,512$, 所以 $R^2=1\,430.276/1\,512=0.945\,9$. 可见上面得到的二元线性回归方程拟合程度良好.

$(X_c\hat{\beta})^T y$, $((X_c^TX_c)^{-1}(X_c^Ty))^T$

$ESS=\hat{\beta}^TX_c^T\vec{y}$ 回归平方和

$=\hat{\beta}^TX^T\vec{y}-n\bar{y}^2$

$=\frac{(SXY)^2}{SXX}$

约束OLS估计

对线性回归模型 $\vec{y}=X\vec{\beta}+\vec{e}$, $E[\vec{e}]=0$, $Cov(\vec{e})=\sigma^2I_n$, $rank(X)=p$. 其回归参数满足 $A\vec{\beta}=\vec{b}$. 求$\hat{\beta}$.

求解: Step 1: 构造辅助函数: $F(\vec{\beta},\vec{\lambda})=\|\vec{y}-X\vec{\beta}\|^2+2(A\vec{\beta}-\vec{b})^T\vec{\lambda}$ 　 $\vec{\lambda}=(\lambda_1,\cdots,\lambda_k)^T$

求临界点: $D_{\vec{\beta},\vec{\lambda}}F(\vec{\beta},\vec{\lambda})=0$ 即 $\frac{\partial F}{\partial\vec{\beta}}=2(X^TX\vec{\beta}-X^T\vec{y}+A^T\vec{\lambda})=0$; $\frac{\partial F}{\partial\vec{\lambda}}=2(A\vec{\beta}-\vec{b})=0$

Step2: 由 $\frac{\partial F}{\partial \vec\beta}=0 \Rightarrow \hat{\vec\beta}_b=(X^TX)^{-1}(X^T\vec y-A^T\vec\lambda)=\hat{\vec\beta}-(X^TX)^{-1}A^T\vec\lambda \Rightarrow$ 代入 $\frac{\partial F}{\partial \vec\lambda}=0$, 则 $\vec\lambda=(A(X^TX)^{-1}A^T)^{-1}(A\hat{\vec\beta}-\vec b) \Longrightarrow$ 得到 $\hat{\vec\beta}_b=\hat{\vec\beta}-(X^TX)^{-1}A^T(A(X^TX)^{-1}A^T)^{-1}(A\hat{\vec\beta}-\vec b)$

定理: 对上述模型, 满足 $A\vec\beta=\vec b$ 的OLS估计为 $\hat{\vec\beta}_b=\hat{\vec\beta}-(X^TX)^{-1}A^T(A(X^TX)^{-1}A^T)^{-1}(A\hat{\vec\beta}-\vec b)$ —— 计算时不要直接代公式, 用辅助函数重新算

Example (1.3.1)

【[0]例3.3.1】在天文测量中，对天空中三个星位点构成的三角形ABC的三个内角$\theta_1,\theta_2,\theta_3$进行测量，得到的测量值分别为$y_1,y_2,y_3$。由于存在测量误差，所以需要对$\theta_1,\theta_2,\theta_3$进行估计，我们利用线性模型表示有关的量：

$$\begin{cases} y_1=\theta_1+e_1 \\ y_2=\theta_2+e_2 \\ y_3=\theta_3+e_3 \\ \theta_1+\theta_2+\theta_3=\pi \end{cases},$$

其中$e_i, i=1,2,3$表示测量误差。求$\theta_1,\theta_2,\theta_3$的最小二乘估计$\hat\theta_1,\hat\theta_2,\hat\theta_3$，并求出当观测值$y_1=\frac{2}{3}\pi, y_2=\frac{1}{4}\pi, y_3=\frac{1}{6}\pi$时$\hat\theta_1,\hat\theta_2,\hat\theta_3$的具体数值。

- 问题等价于在约束条件$\vec 1_3^T\vec\theta=\pi$下求解如下模型参数的约束OLS估计：
$$\vec y=\vec\theta+\vec e.$$
- 构造辅助函数
$$F(\vec\theta,\lambda)=\|\vec y-\vec\theta\|^2+2\lambda\left(\vec 1_3^T\vec\theta-\pi\right)$$
- 求F的临界点。求解方程
$$\frac{\partial F}{\partial\vec\theta}=2\left(\vec\theta-\vec y+\vec 1_3\lambda\right)=\vec 0.$$
得
$$\vec\theta=\vec y-\lambda\vec 1_3. \quad (5)$$
- 上述等式左右两边同时左乘$\vec 1_3^T$并利用约束条件$\pi=\vec 1_3^T\vec\theta$，可求得$\lambda=(\bar y-\frac{\pi}{3})$.
- 代入(5)，由定理1.3.1即可得
$$\hat{\vec\theta}_{con}=\vec y-(\bar y-\frac{\pi}{3})\vec 1_3. \quad (6)$$
- 将观测值$y_1=\frac{2}{3}\pi, y_2=\frac{1}{4}\pi, y_3=\frac{1}{6}\pi$代入(6)，可得
$$\hat{\vec\theta}_{con}=\begin{pmatrix}\frac{23}{36}\pi\\ \frac{2}{9}\pi\\ \frac{5}{36}\pi\end{pmatrix}.$$

解答完毕。

残差分析

定义: 对经典线性回归模型 $\vec y=X\vec\beta+\vec e$, 有 $\hat{\vec\beta}=(X^TX)^{-1}X^T\vec y$. 有 y 的拟合值 $\hat{\vec y}=X\hat{\vec\beta}=X(X^TX)^{-1}X^T\vec y:=H\vec y$. 称 $H=X(X^TX)^{-1}X^T$ 为帽子矩阵 (因为 $\hat{\vec y}=H\vec y$, 给 $\vec y$ 戴了帽子)

性质: ① $H^T=H$, $H^2=H$, 即 H 为对称幂等阵. ② $\mathrm{rank}\,H=\mathrm{rank}\,X=p$. ③ $\sum_{i=1}^n h_{ii}=p$ ④ $\sum_{i=1}^n h_{ij}=\sum_{j=1}^n h_{ij}=1$ 即 $H\vec 1_n=\vec 1_n$ ⑤ $HX=X$.

定义: 对角元 h_{ii} 称为第 i 个案例的杠杆值(杠杆). h_{ii} 越大, 第 i 个案例的作用越大, 有 $h_{ii}=\frac{1}{n}+(\xi_i-\bar m)^T(X_c^TX_c)^{-1}(\xi_i-\bar m)$, $\xi_i=(1,x_{i1},\dots,x_{i,p-1})$ 为案例点.

定理: 对残差向量 $\hat{\vec e}=\vec y-\hat{\vec y}=(I_n-H)\vec y=(I_n-H)(X\vec\beta+\vec e)=(I_n-H)\vec e$. 有 ① $E[\hat{\vec e}]=0$, $\mathrm{Cov}(\hat{\vec e})=\sigma^2(I_n-H)$

$h_{ii}=\frac{1}{n}+x_{ci}^T(X_c^TX_c)^{-1}x_{ci}$

第 i 个案例, 有 $\mathrm{Var}(\hat e_i)=[\mathrm{Cov}(\hat{\vec e})]_{ii}=\sigma^2(1-h_{ii})$

② $\mathrm{Cov}(\hat{\vec e},\hat{\vec y})=0$, $\mathrm{Cov}(\hat{\vec y})=\sigma^2H$

由此可见 $\xi_i\to$ 中心时 方差比中心时大, 数据不稳定, 考虑标准化

③ 若假设 $\vec e\sim N_n(0,\sigma^2I_n)$, 则 $\hat{\vec e}\sim N_n(0,\sigma^2(I_n-H))$, $\hat{\vec y}\sim N_n(X\vec\beta,\sigma^2H)$

$\mathrm{Cov}(\hat{\vec y})=\sigma^2H$

$\mathrm{Cov}(\hat{\vec e})=\sigma^2(I_n-H)$

$r_i=\frac{\hat e_i}{\hat\sigma\sqrt{1-h_{ii}}}$

定义: 第 i 个案例的学生化残差 $r_i=\frac{\hat e_i}{\hat\sigma\sqrt{1-h_{ii}}}$

定义: 若一个案例不适从某模型但其余数据均适从此模型, 称其为异常值.

$t_i=r_i\sqrt{\frac{n-p-1}{n-p-r_i^2}}\sim t_{n-p-1}$

方法: 考虑均值漂移线性回归模型. 设第 i 个案例可能为异常值. $\begin{cases} y_j=(1,\xi_j^T)\vec\beta+e_j & j\neq i\\ y_i=(1,\xi_i^T)\vec\beta+\eta+e_i\end{cases}$ $\vec e\sim N(0,\sigma^2I_n)$ 在水平 α 下检验 $H_0:\eta=0$ 若不成立, 则第 i 案例为异常值.

有结论: 若 $H_0:\eta=0$ 成立, 则 $t_i=r_i\sqrt{\frac{n-p-1}{n-p-r_i^2}}\sim t_{n-p-1}$, 当 $|t_i|\ge t_{n-p-1}(\frac{\alpha}{2})$ 时判定为异常值

例:

Example (2.1.1)

在Forbes数据中，检验案例12（$x_{12}=204.6, y_{12}=142.44$）是否是异常值，检验水平$\alpha=0.05$。备注：$n=17, \bar x=202.95$, $\hat\sigma=0.379, SXX=530.78$，拟合回归方程$\hat Y=-42.131+0.895X$。

$\hat e_{12}=y_{12}-\hat y_{12}=1.36$

$h_{12,12}=\frac{1}{n}+(x_{12}-\bar x)(SXX)^{-1}(x_{12}-\bar x)=0.0639$ —— 一元回归中 $SXX=\sum(x_i-\bar x)^2=X_c^TX_c$; 15x, 共2个参数.

$r_{12}=\frac{\hat e_{12}}{\hat\sigma\sqrt{1-h_{12,12}}}=3.7078 \Rightarrow t_{12}=r_{12}\sqrt{\frac{n-p-1}{n-p-r_{12}^2}}=12.4>t_{14}(0.025)=2.1448$, 是异常值.

强影响分析

定义: 在某些数据集合中, 若一个案例被删除后观测的拟合估计量有重要改变, 则称之为强影响案例.

定义: Cook距离 $D_i=\frac{(\hat{\vec\beta}-\hat{\vec\beta}_{(i)})^TX^TX(\hat{\vec\beta}-\hat{\vec\beta}_{(i)})}{p\hat\sigma^2}$ p 为参数个数, 包括1; $\hat{\vec\beta}_{(i)}$ 表示剔除第 i 个案例后得到的 $\vec\beta$ 的OLS估计.

定理: $D_i=\frac{1}{p}\left(\frac{h_{ii}}{1-h_{ii}}\right)r_i^2$ $i=1,\dots,n$. h_{ii} 为帽子矩阵 $H=X(X^TX)^{-1}X^T$ 的第 i 个对角元, r_i 为第 i 个案例的学生化残差.

tip: 若 p 固定, D_i 由 r_i 与 h_{ii} 决定. r_i 反映在第 i 个模型处的失真程度, h_{ii} 反映第 i 案例与中心的远近程度.

$D_i=\frac{1}{p}\frac{h_{ii}}{1-h_{ii}}\cdot r_i^2 - F_{p,n-p}(\alpha)$

如何用Cook距离判定某案例是强影响案例?

Lemma: $X\sim N_n(0,\Sigma)$, 则 $X^T\Sigma^{-1}X\sim\chi^2_n$ 类比 $X\sim N(0,\sigma^2)$, $\frac{X^2}{\sigma^2}\sim\chi^2_1$

由上述Lemma得到: 由于 $\hat{\vec\beta}-\vec\beta\sim N(0,\sigma^2(X^TX)^{-1})$, 则 $\frac{(\hat{\vec\beta}-\vec\beta)^TX^TX(\hat{\vec\beta}-\vec\beta)}{p\hat\sigma^2}\sim F_{p,n-p}$

对于给定的 α, $\frac{(\hat{\vec\beta}-\vec\beta)^TX^TX(\hat{\vec\beta}-\vec\beta)}{p\hat\sigma^2}\le F_{p,n-p}(\alpha)$ 的概率为 $1-\alpha$, 以 $\vec\beta$ 为变量, 其表示 p 维空间中以 $\hat{\vec\beta}$ 为中心的椭球.

$\Rightarrow$ 若 $D_i=F_{p,n-p}(\alpha)$, 表示第 i 组数据剔除后 $\hat{\vec\beta}_{(i)}$ 落在了置信椭球上. D_i 越大表明第 i 组数据影响越大

总结: 对线性回归模型 $\vec y=X\vec\beta+\vec e$, $\vec e\sim N(0,\sigma^2I_n)$ 时 $D_i\le F_{p,n-p}(\alpha)$ 则认为 $\vec\beta$ 的两个OLS估计 $\hat{\vec\beta}$, $\hat{\vec\beta}_{(i)}$ 无重大差别, 第 i 个案例不是强影响案例.

Example (2.1.1)

在Forbes数据中，检验案例12（$x_{12}=204.6, y_{12}=142.44$）是否是异常值，检验水平$\alpha=0.05$。备注：$n=17, \bar x=202.95$, $\hat\sigma=0.379, SXX=530.78$，拟合回归方程$\hat Y=-42.131+0.895X$。

Example (2.1)

在Forbes数据中，由本章例1.1得，案例12其$h_{12,12}=0.0639$，学生化残差$r_{12}=3.7078$。判断其是否是强影响数据，检验水平$\alpha=0.05$.

有 $h_{12,12}=0.0639$, $r_{12}=3.7078$. $\Rightarrow$ Cook统计量 $D_{12}=\frac{1}{p}\left(\frac{h_{12,12}}{1-h_{12,12}}\right)\cdot r_{12}^2=0.47<F_{2,15}(0.05)=3.68$ 于是不是强影响数据

Box-Cox 变换

"好"的残差图: 残差方差稳定, 在图中分布均匀; y与x线性. "坏"的残差图: 残差方差不稳定, 在图中分布不均匀

定理(方差稳定化变换): $E[Y]=\mu$, $\mathrm{Var}(Y)=\sigma^2$, 若 $\sigma=\sigma(\mu)$, 则 $Y\to\tilde Y=g(Y)=\int_0^Y\frac{1}{\sigma(\mu)}d\mu$ 的方差基本为常数 (为了满足Gauss-Markov假设)

例: 常用变换: ① $\sqrt Y$ 若 $\mathrm{Var}(Y_i)\propto E[Y_i]$ $Y_i\ge0$ ③ $\frac{1}{Y}$ 若 $\mathrm{Var}(Y_i)\propto E[Y_i]^4$

② $\ln Y$ 若 $\mathrm{Var}(Y_i)\propto E[Y_i]^2$ $Y_i\ge0$ ④ $\sin^{-1}(\sqrt Y)$ 若 $\mathrm{Var}(Y_i)\propto E[Y_i](1-E[Y_i])$ $0\le Y_i\le1$

例: 设 Y 为各个乡镇事故次数, X 为车辆行驶速度. $Y\sim P(\lambda)$, 研究 Y 与 X 关系, 方差稳定

若为线性关系, $Y=\beta_0+\beta_1x+e$. 则 $\beta_0+\beta_1x=E[Y]=\mathrm{Var}(Y)=\lambda$. 又 $\mathrm{Var}(e)=\mathrm{Var}(Y)+\mathrm{Var}(\beta_0+\beta_1x)-2\mathrm{Cov}(Y,\beta_0+\beta_1x)=\mathrm{Var}(Y)$. 不满足Gauss-Markov假设中的等方差性.

于是作变换: $Y\to\tilde Y=g(Y)=\int_0^Y\frac{1}{\sqrt\lambda}d\lambda=2\sqrt Y$. 考虑 $\sqrt Y=\beta_0+\beta_1x+e$. 则 $\mathrm{Var}(\sqrt Y)\approx\left(\frac{1}{2\sqrt\lambda}\right)^2\cdot\lambda=\frac14$ $\left(\mathrm{Var}(g(Y))\approx(g'(\lambda))^2\cdot\mathrm{Var}(Y)\right)$

定义: 假设 $y>0$. 则 Box-Cox变换为 $y\to y^{(\lambda)}=\begin{cases}\frac{y^\lambda-1}{\lambda} & \lambda\neq0\\ \ln y & \lambda=0\end{cases}$ tip: 变换后 $y^{(\lambda)}\sim N_n(X\vec\beta,\sigma^2I_n)$

如何确定 λ 的值进而对 $y^{(\lambda)}=X\beta+e$, $e\sim N(0,\sigma^2I_n)$ 给出 β,σ^2 的OLS估计?

Step1: $y^{(\lambda)}$ 的联合p.d.f.为 $L(\vec\beta,\sigma^2)=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n\exp\{-\frac{1}{2\sigma^2}(y^{(\lambda)}-X\vec\beta)^T(y^{(\lambda)}-X\vec\beta)\}$ 则 $\vec y$ 的p.d.f为 $L(\lambda,\vec\beta,\sigma^2)=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n\exp\{-\frac{1}{2\sigma^2}\|y^{(\lambda)}-X\vec\beta\|^2\}J$ 其中 J 为Box-Cox变换的雅可比. $J=\prod_{i=1}^n\left|\frac{dy_i^{(\lambda)}}{dy_i}\right|=\prod_{i=1}^n y_i^{\lambda-1}$

Step2: 若 λ 已知, 有MLE: $\hat\beta(\lambda)=(X^TX)^{-1}X^Ty^{(\lambda)}$, $\hat\sigma^2(\lambda)=\frac1n y^{(\lambda)T}(I-H)y^{(\lambda)}=\frac{RSS(\lambda)}{n}$

Step3: 将MLE代入, 取对数, 有 $\ln L_{max}(\lambda)=-\frac n2\ln\frac{RSS(\lambda)}{J^{2/n}}-\frac n2\left(\ln\left(\frac{2\pi}{n}\right)+1\right)=-\frac n2\ln\frac{RSS(\lambda)}{J^{2/n}}+\mathrm{Const}$

$RSS=\vec y^T(I_n-H)\vec y=e(I_n-H)e$

Step4: 求导为0无法得到 λ 的解析式, 令 $z_i^{(\lambda)}=\begin{cases}y_i^{(\lambda)}/\left(\prod_{i=1}^n y_i\right)^{\frac{\lambda-1}{n}} & \lambda\neq0\\ (\ln y_i)\cdot\left(\prod_{i=1}^n y_i\right)^{\frac1n} & \lambda=0\end{cases}$, 则 $\frac{RSS(\lambda)}{J^{2/n}}=z^{(\lambda)T}(I-H)z^{(\lambda)}:=RSS(\lambda,z^{(\lambda)})$. 找到一组 λ 中使 $RSS(\lambda,z^{(\lambda)})$ 最小的那一个.

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

广义OLS估计

对线性回归模型 $\vec y=X\vec\beta+\vec e$, $E[\vec e]=0$, $\mathrm{Cov}(\vec e)=\sigma^2\Sigma$, Σ 为正定阵 (不一定为 I_n)

① Σ已知　作线性变换：$T:\mathbb{R}^n \mapsto \mathbb{R}^n$，$\vec{y} \mapsto T\vec{y} = \Sigma^{-\frac{1}{2}}\vec{y}$　代入模型，有 $E[T\vec{e}]=0$，$Cov(T\vec{e})=\sigma^2 I_n$，记 $z=\Sigma^{-\frac{1}{2}}Y$，$U=\Sigma^{-\frac{1}{2}}X$，$\varepsilon=\Sigma^{-\frac{1}{2}}e$，则 $z=U\beta+\varepsilon$，$E[\varepsilon]=0$，$Cov(\varepsilon)=\sigma^2 I$

于是有 $\hat{\beta}_G=(U^TU)^{-1}U^Tz=(X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}\vec{y}$，称其为 β 的广义OLS估计.（GLS估计）　$(X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}\vec{y}$

定理：① $E[\hat{\beta}_G]=\beta$　② $Cov(\hat{\beta}_G)=\sigma^2(X^T\Sigma^{-1}X)^{-1}$　③ 对 $\forall \vec{c}$ 为 $p\times1$ 常向量，$c^T\hat{\beta}_G$ 为 $c^T\beta$ 的唯一最小方差无偏估计（对一般线性回归模型，无偏与最小方差意义下GLS优于OLS估计）

例：若 $\Sigma=Cov(\vec{e})=diag(\sigma_1^2,\cdots,\sigma_n^2)$，求 β 的GLS估计.

记 X 的 n 个行向量为 $\xi_1^T,\cdots,\xi_n^T$，则由公式，$\hat{\beta}_G=\left((\xi_1,\cdots,\xi_n)diag(\frac{1}{\sigma_1^2},\cdots,\frac{1}{\sigma_n^2})\begin{pmatrix}\xi_1^T\\ \vdots\\ \xi_n^T\end{pmatrix}\right)^{-1}(\xi_1,\cdots,\xi_n)diag(\frac{1}{\sigma_1^2},\cdots,\frac{1}{\sigma_n^2})\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix}=\left(\sum_{i=1}^n\frac{\xi_i\xi_i^T}{\sigma_i^2}\right)^{-1}\cdot\sum_{i=1}^n\frac{\xi_i^T y_i}{\sigma_i^2}$　$S_{XX}=X_c^TX_c$

② Σ未知

例：

Example (4.2)

【[0]例3.6.1】在线性回归模型(1)中，若观测向量 $\vec{y}$ 是 n 件样品的某项指标，它们是在k台仪器上测试得到的。不妨设前 n_1 个是在第一台仪器上测得的，试验和测量误差为 σ_1^2，接下来的 n_2 个是在第二台仪器上测得的，测量误差为 σ_2^2，依此类推。并且假设仪器之间的试验和测量误差互不相关。这里 $n_i \geq p$，σ_i^2 未知，$i=1,\ldots,k$。求模型(1)中回归系数 $\vec{\beta}$ 的GLS 估计。

模型可写为 $y_i=X_i\beta+e_i$，$E[e_i]=0$，$Cov(e_i)=\sigma_i^2 I_{n_i}$

对每个 i，有 σ_i^2 无偏估计：$\hat{\sigma}_i^2=\frac{\|y_i-X_i\hat{\beta}_i\|^2}{n_i-p}$

若 σ_i^2 已知，有 β 的GLS估计为 $\hat{\beta}_G=\left(\sum_{i=1}^k\frac{X_i^TX_i}{\sigma_i^2}\right)^{-1}\left(\sum_{i=1}^n\frac{X_i^Ty_i}{\sigma_i^2}\right)$，将得到的 $\hat{\sigma}_i^2$ 代入，有 $\hat{\beta}_G=\left(\sum_{i=1}^k\frac{X_i^TX_i}{\hat{\sigma}_i^2}\right)^{-1}\left(\sum_{i=1}^n\frac{X_i^Ty_i}{\hat{\sigma}_i^2}\right)$

称为 β 的两步估计.

期中部分总结.

$SS_{reg}=\frac{S_{XY}^2}{S_{XX}}$　$RSS=SYY-\frac{S_{XY}^2}{S_{XX}}$

第一章：

tip：无截距时 $SYY \neq RSS+ESS$

1. $\vec{y}=X\vec{\beta}+\vec{e}$，$\vec{e}\sim N_n(0,\sigma^2I_n)$ 即 $E[\vec{e}]=0$，$Cov(\vec{e})=\sigma^2I_n$.

$\hat{\beta}-\beta=(X^TX)^{-1}X^T\vec{e}$

2. $\hat{\beta}=(X^TX)^{-1}X^T\vec{y}$．无截距时 $\hat{\beta}=\frac{S_{XY}+n\bar{x}\bar{y}}{S_{XX}+n\bar{x}^2}$　$RSS=SYY+n\bar{y}^2-\frac{(S_{XY}+n\bar{x}\bar{y})^2}{S_{XX}+n\bar{x}^2}$　$R^2=\frac{(S_{XY}+n\bar{x}\bar{y})^2}{(S_{XX}+n\bar{x}^2)(SYY+n\bar{y}^2)}$

含截距只斜率 → $\hat{\beta}_c=(X_c^TX_c)^{-1}X_c^T\vec{y}$

有截距时 $\hat{\beta}=\frac{S_{XY}}{S_{XX}}=(X_c^TX_c)^{-1}X_c^Ty$　$RSS=\sum_{i=1}^n(y_i-\hat{y}_i)^2=SYY-\frac{S_{XY}^2}{S_{XX}}$　$R^2=\frac{ESS}{SYY}=1-\frac{RSS}{SYY}=\frac{(S_{XY})^2}{S_{XX}\cdot SYY}$

$\hat{e}_i=y_i-\hat{y}_i$　$ESS=\sum_{i=1}^n(\bar{y}_i-\hat{y}_i)^2=\frac{S_{XY}^2}{S_{XX}}$　$SYY=\sum_{i=1}^n(y_i-\bar{y}_i)^2=RSS+ESS$　$\hat{\sigma}^2=\frac{RSS}{n-p}$，$\frac{RSS}{\sigma^2}\sim\chi^2_{n-p}$

→ 是 σ^2 的无偏估计.

3. $\hat{y}=\hat{\alpha}+X_c\hat{\beta}+\vec{e}$　中心化时 $\hat{\alpha}=\bar{y}$，$\hat{\beta}_0=\hat{\alpha}-\hat{\beta}^T\bar{m}$

4. $E[\hat{\beta}]=\beta$，$Cov(\hat{\beta})=\sigma^2(X^TX)^{-1}$．$Cov(\hat{\beta})=\sigma^2(X^TX)^{-1}$

$Cov(\hat{\beta}_c)=\sigma^2(X_c^TX_c)^{-1}$　其中 $\hat{\beta}_c$ 是不含截距 $\hat{\beta}_0$，只有斜率部分的 Cov

$t_i=r_i\sqrt{\frac{n-p-1}{n-p-r_i^2}}$

特别地，一元时，$\hat{\beta}_1=\frac{\sigma^2}{S_{XX}}$　$S_{XX}=X_c^TX_c$ ★

$\frac{1}{n}+(\xi_i-\bar{x})(X_c^TX_c)^{-1}(\xi_i-\bar{x})$．$\hat{e}$

5. 约束OLS估计先构造辅助函数．一般形式用 $+2(A\beta-b)^T\lambda$

$r_i=\frac{\hat{e}}{\hat{\sigma}\sqrt{1-h_{ii}}}$　$D_i=\frac{1}{p}\cdot\frac{h_{ii}}{1-h_{ii}}\cdot r_i^2$

第二章：

6. 帽子矩阵 $H=X(X^TX)^{-1}X^T$　$HX=X$．$h_{11}=x_1^T(X^TX)^{-1}x_1$　$h_{ii}=\frac{1}{n}+x_{ci}^T(X_c^TX_c)^{-1}x_{ci}$　中心化自变量向量，如：$h_{55}=\frac{1}{n}+(x_5-\bar{x})^T(X_c^TX_c)^{-1}(x_5-\bar{x})$

7. 学生化残差 $r_i=\frac{\hat{e}_i}{\hat{\sigma}\sqrt{1-h_{ii}}}$ ⟹ 判断异常值：若 $t_i=|r_i|\sqrt{\frac{n-p-1}{n-p-r_i^2}}\geq t_{n-p-1}(\frac{\alpha}{2})$ 为异常值．

① $E[\hat{e}]=0$，$Cov(\hat{e})=\sigma^2(I_n-H)$

② $Cov(\hat{e},\hat{y})=0$，$Cov(\hat{y})=\sigma^2H$

③ 若假设 $\vec{e}\sim N_n(0,\sigma^2I_n)$，则 $\hat{e}\sim N_n(0,\sigma^2(I_n-H))$，$\hat{y}\sim N_n(X\beta,\sigma^2H)$

$A=\begin{bmatrix}X_c^TX_c & X_c^T\vec{y}\\ \vec{y}^TX_c & SYY\end{bmatrix}$　$ESS=(X_c^Ty)^T(X_c^TX_c)^{-1}(X_c^Ty)$　$\left(\frac{(S_{XY})^2}{S_{XX}}\right)$　$t_i=|r_i|\sqrt{\frac{n-p-1}{n-p-r_i^2}}$

Cook距离：$D_i=\frac{1}{p}\left(\frac{h_{ii}}{1-h_{ii}}\right)\cdot r_i^2$ ⟹ 判断强影响：若 $D_i>F_{p,n-p}(\alpha)$ 为强影响点

8. Box-Cox变换　$\frac{1}{p}\cdot\frac{h_{ii}}{1-h_{ii}}\cdot r_i^2$

GLS估计：$\hat{\beta}=(X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}\vec{y}$

关于OLS估计是BLUE的若干问题

- 问题1：对于线性回归模型(1)，$\hat{\beta}_0$ 是否是 β_0 的BLUE? 也即，$\hat{\beta}_0$ 是否是 β_0 在线性无偏估计类
$$\{\sum_{i=1}^n a_iy_i : a_i\in\mathbb{R}, i=1,\cdots,n\}$$
中的UMVUE ?
- 问题2：对于中心化的线性回归模型
$$\vec{y}=\alpha\cdot\vec{1}_n+X_c\vec{\beta}_c+\vec{e},$$
$$E(\vec{e})=0, Cov(\vec{e})=\sigma^2I_n, rank(X_c)=p-1,$$
其OLS估计 $\hat{\alpha}=\bar{y}$ 是否是 α 的BLUE ?
- 问题3：若rank $(X)<p$，则本章第一节方程(5)的解
$$\hat{\beta}=(X^TX)^-X^T\vec{y}$$
是否成立Gauss-Markov定理？

答案 不一定，参考【1】定理4.1.2.

结论：是

解析：

1. 线性性：截距项OLS估计量 $\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}_1-\ldots-\hat{\beta}_{p-1}\bar{x}_{p-1}$，是 y 的线性函数（$\hat{\beta}_j$ 均为 y 的线性组合），满足线性条件。
2. 无偏性：在Gauss-Markov假设下，$E(\hat{\beta}_0)=\beta_0$，无偏性成立。
3. 最佳性：由Gauss-Markov定理，OLS估计量是所有线性无偏估计中的方差最小者，因此 $\hat{\beta}_0$ 是 β_0 的BLUE。
4. 补充：若进一步假设误差服从正态分布 $\varepsilon\sim N(0,\sigma^2I_n)$，则 $\hat{\beta}_0$ 同时是一致最小方差无偏估计（UMVUE），即所有无偏估计（不限于线性）中方差最小。

结论：是

模型与解析：

中心化线性回归模型：
$\bar{y}=\alpha\cdot1_n+X_c\beta_c+e$，$E(e)=0$，$Cov(e)=\sigma^2I_n$，$rank(X_c)=p-1$
其中 1_n 是全1向量，X_c 是中心化后的解释变量矩阵（列均值为0，因此 $1_n'X_c=0$）。

1. OLS估计推导：
正规方程为：
$$\begin{cases}1_n'1_n\cdot\alpha+1_n'X_c\cdot\beta_c=1_n'y\\ X_c'1_n\cdot\alpha+X_c'X_c\cdot\beta_c=X_c'y\end{cases}$$
由 $1_n'X_c=0$，第一个方程简化为 $n\alpha=n\bar{y}$，因此 $\hat{\alpha}=\bar{y}$。
2. BLUE验证：
 - 线性性：$\hat{\alpha}=\bar{y}=\frac{1}{n}\sum_{i=1}^n y_i$，是 y 的线性函数；
 - 无偏性：$E(\hat{\alpha})=E(\bar{y})=\alpha+\frac{1}{n}1_n'X_c\beta_c+E(\bar{e})=\alpha$（$1_n'X_c=0$，$E(\bar{e})=0$）；
 - 最佳性：由Gauss-Markov定理，中心化模型的OLS估计量 $(\hat{\alpha},\hat{\beta}_c')'$ 是BLUE，因此 $\hat{\alpha}=\bar{y}$ 是 α 的BLUE。
3. 直观理解：
中心化后截距项 α 就是响应变量的均值 $\bar{y}$，样本均值是总体均值的经典BLUE估计量，符合结论。

结论：不一定成立

解析：

1. 核心前提缺失：
Gauss-Markov定理的必要前提是 X 列满秩（$rank(X)=p$），此时 $X'X$ 可逆，OLS估计量 $\hat{\beta}=(X'X)^{-1}X'y$ 唯一、可识别。
当 $rank(X)<p$ 时，$X'X$ 不可逆，只能用广义逆 $(X'X)^-$ 求解，此时：
 - 估计量不唯一：不同广义逆会得到不同的 $\hat{\beta}$；
 - 参数不可识别：无法唯一区分不同解释变量的系数效应（多重共线性完全）；
 - 无偏性仅对可估计函数成立，单个系数 β_j 不再有唯一的无偏估计。
2. 广义逆解的性质：
$\hat{\beta}=(X'X)^-X'y$ 是正规方程的解，但：
 - 仅保证拟合值 $\hat{y}=X\hat{\beta}$ 唯一（与广义逆选择无关）；
 - 单个系数 $\hat{\beta}_j$ 不唯一，因此不满足Gauss-Markov定理对"唯一线性无偏最小方差估计"的要求。
3. 补充说明：
仅当估计的是可估计函数（如 $c'\beta$，其中 $c\in\mathcal{C}(X')$，$\mathcal{C}(\cdot)$ 为列空间）时，$c'\hat{\beta}$ 是唯一的BLUE；但单个系数 β_j 作为不可估计函数，不存在BLUE。

关于Gauss-Markov假设的若干问题

- 问题1：如果随机误差项不满足Gauss-Markov假设，上述性质是否仍成立？　不成立，有偏、非有效
- 问题2：观测值向量 $\vec{y}$ 与其它随机向量的相关关系？
$$Cov(\vec{y},\hat{e})=?\quad Cov(\vec{y},\hat{e})=?\quad Cov(\vec{y},\hat{\beta})=?\quad Cov(\vec{y},\hat{y})=?$$
$=Cov(y,(X^TX)^{-1}X^Ty)=\sigma^2X(X^TX)^{-1}$；$=Cov(y,(I_n-H)y)=Var(y)\cdot(I_n-H)=\sigma^2(I_n-H)$；$=Cov(y,Hy)=\sigma^2H$
- 问题3：拟合值向量 $\hat{y}$ 与其它随机向量的相关关系？
$$Cov(\hat{y},\vec{e})=?\quad Cov(\hat{\vec{e}},\hat{e})=?\quad Cov(\hat{y},\hat{\beta})=?$$
$=\sigma^2H(I_n-H)$　✓　$=X\,Var(\hat{\beta})=\sigma^2X(X^TX)^{-1}$
- 问题4：考虑简单线性回归模型 $y_i=\beta x_i+e_i$，$i=1,2,\cdots,n$，e_i 满足Gauss-Markov假设。若 x_i，y_i 均为随机，
 - 此时是否仍成立 $E(y_i|x_i)=\beta x_i$？ 成立
 - 假设 $Var(x_i)=\tau^2$，$i=1,\cdots,n$，且 x_i 与 e_i 独立，计算 $Var(y_i)$ 和 $Var(y_i|x_i)$ $=Var(e|x_i)=\sigma^2$

 $y_i=\beta x_i+e_i$　$Var(y_i)=\beta^2\tau^2+\sigma^2$

- 问题1：性质(3)中第二个等式表明残差向量与拟合值向量相互正交，那么残差向量与解释向量 $\{\vec{x}_1,\cdots,\vec{x}_{p-1}\}$ 呢？

答案 相互正交。

- 问题2：对于无截距项的回归模型，
$$y_i=\beta_1x_{i1}+\beta_2x_{i2}+\cdots+\beta_{p-1}x_{i,p-1}+e_i, i=1,2,\cdots,n,$$
e_i, $i=1,\cdots,n$满足Gauss-Markov假设。性质(3)中两个等式是否仍然成立？性质(1)和(2)是否仍然满足？

答案 除了性质(3)中第二个等式仍然成立，，其余性质(1)-(3)均不一定成立。

$S_{XX}=X_c^TX_c$.

- 约束最小二乘估计为 $\hat{\beta}_{con}$ 是否是 $\vec{\beta}$ 的无偏估计？

答案 在约束条件下，是。无约束条件下，是否无偏？

- 对于回归参数有线性约束条件的回归模型(1)，在任一关于 $\vec{\beta}$ 的线性函数 $\vec{c}^T\vec{\beta}$ 的所有线性无偏估计中，OLS估计 $\vec{c}^T\hat{\beta}$ 是否仍是唯一具有最小方差的估计？

答案 不是。约束OLS估计 $\vec{c}^T\hat{\beta}_{con}$ 是无偏的且方差可以更小。参考无约束条件下OLS估计的Gauss-Markov定理证明。

- 定理1.3.1结果对于无截距模型是否成立？

答案 仍成立。

- 中心化模型下，假设回归参数有约束条件 $A\vec{\beta}_c=\vec{b}$，求此时参数的约束最小二乘估计？

答案

$\hat{\alpha}_{con}=\bar{y}$,

$\hat{\beta}_{c,con}=\hat{\beta}_c-(X_c^TX_c)^{-1}A^T\left(A(X_c^TX_c)^{-1}A^T\right)^{-1}(A\hat{\beta}_c-\vec{b})$.

Problem

与最小二乘（OLS）估计 $\hat{\eta}=\vec{c}^T\hat{\beta}$ 相比，参数 $\eta=\vec{c}^T\vec{\beta}$ 的另一估计 $\hat{\eta}_G=\vec{c}^T\hat{\beta}_G$ 方差更小还是更大？

答案 当随机误差项 $\vec{e}$ 满足 $E(\vec{e})=0$，$Cov(\vec{e})=\sigma^2I$ 时，$\vec{c}^T\hat{\beta}$ 方差最小；而若 $Cov(\vec{e})=\sigma^2\Sigma\neq\sigma^2I$ 时，由定理4.1（c）可知，对于任意 $\vec{c}\in\mathbb{R}^p$，
$$Var\left(\vec{c}^T\hat{\beta}_G\right)\leq Var\left(\vec{c}^T\hat{\beta}\right).$$

- 即对于一般线性回归模型(1)，以线性无偏和最小方差意义而言，广义最小二乘（GLS）估计优于最小二乘(OLS)估计。

方差分析表 1. 最简

问题：$\vec{e}\sim N(0,\sigma^2 I)$，两模型为 $Y=\beta_0+\beta_1X_1+\cdots+\beta_{p-1}X_{p-1}+e$，$Y=\beta_0+e$，是否有显著区别？检验水平为 α.

答：检验 $H_0:\beta_1=\cdots=\beta_{p-1}=0 \leftrightarrow H_1:\beta_1,\cdots,\beta_{p-1}$ 中至少有1个非0.

若 σ^2 已知，$\frac{ESS}{\sigma^2}\sim\chi^2_{p-1}$. 当 $\frac{ESS}{\sigma^2}>\chi^2_{p-1}(\alpha)$ 时认为两模型有显著区别

若 σ^2 未知，有 $\hat{\sigma}^2=\frac{RSS}{n-p}$，$\frac{RSS}{\sigma^2}\sim\chi^2_{n-p}$ 故 $\frac{ESS/p-1}{\hat{\sigma}^2}\sim F_{p-1,n-p}$，当 $\frac{ESS/p-1}{\hat{\sigma}^2}>F_{p-1,n-p}(\alpha)$ 时认为两模型有显著区别

定义 $MSS\triangleq\frac{ESS}{p-1}$

- 记回归均方为 $MS_{reg}=SS_{reg}/(p-1)$，则上述讨论可归结为如下表
- 方差分析表（ANOVA）

来源	d.f.	SS	MS	F
X的回归	$p-1$	SS_{reg}	$SS_{reg}/(p-1)$	$MS_{reg}/\hat{\sigma}^2$
残差	$n-p$	RSS	$RSS/(n-p)$	
总的	$n-1$	SYY		

2. 单个自变量的显著性检验

问题：$\vec{e}\sim N(0,\sigma^2 I)$，两模型为 $Y=\beta_0+\beta_1X_1+\cdots+\beta_{p-1}X_{p-1}+e$，$Y=\beta_0+\beta_2X_2+\cdots+\beta_{p-1}X_{p-1}+e$ 是否有显著区别？检验水平为 α.

答：检验 $H_0:\beta_1=0 \leftrightarrow H_1:\beta_1\neq0$

定义 $(C_{ij})_{i,j=1,\cdots,p-1}=(X_c^TX_c)^{-1}$，$C_{00}=\frac{1}{n}+\vec{m}^T(X_c^TX_c)^{-1}\vec{m}$ 以 $t_j=\frac{\hat{\beta}_j}{\hat{\sigma}\sqrt{C_{jj}}}\sim t_{n-p}$. 则当 $|t_1|>t_{n-p}(\frac{\alpha}{2})$ 时拒绝 H_0. 认为两模型有显著区别

一般地，对 $H_0:\beta_j=0 \Leftrightarrow H_1:\beta_j\neq0$. $|t_j|>t_{n-p}(\frac{\alpha}{2})$ 时拒绝 H_0.

或：完整模型下残差平方和为RSS，去除 X_1 模型下残差平方和为 $RSS_{(1)}$

定义 $\widetilde{SS}_{reg}^{(1)}=RSS-RSS_{(1)}$，$\widetilde{MS}_{reg}^{(1)}=\frac{\widetilde{SS}_{reg}^{(1)}}{1}$. 有 $F_1=\frac{\widetilde{MS}_{reg}^{(1)}}{\hat{\sigma}^2}\sim F_{1,n-p}$. $F_1>F_{1,n-p}(\alpha)$ 时认为两模型有显著区别

3. 两个自变量的显著性检验

问题：$\vec{e}\sim N(0,\sigma^2 I)$，两模型为 $Y=\beta_0+\beta_1X_1+\cdots+\beta_{p-1}X_{p-1}+e$，$Y=\beta_0+\beta_3X_3+\cdots+\beta_{p-1}X_{p-1}+e$ 是否有显著区别？检验水平为 α.

或：完整模型下残差平方和为RSS，去除 $X_{1,2}$ 模型下残差平方和为 $RSS_{(1,2)}$

定义 $\widetilde{SS}_{reg}^{(1,2)}=RSS-RSS_{(1,2)}$，$\widetilde{MS}_{reg}^{(1,2)}=\frac{\widetilde{SS}_{reg}^{(1,2)}}{2}$. 有 $F_2=\frac{\widetilde{MS}_{reg}^{(1,2)}}{\hat{\sigma}^2}\sim F_{2,n-p}$. $F_2>F_{2,n-p}(\alpha)$ 时认为两模型有显著区别

↳ 一般地，$\widetilde{MS}_{reg}^{(1,\cdots,k-1)}=\frac{\widetilde{SS}_{reg}^{(1,\cdots,k-1)}}{k-1}$ 有 $F_{k-1}=\frac{\widetilde{MS}_{reg}^{(1,\cdots,k-1)}}{\hat{\sigma}^2}\sim F_{k-1,n-p}$. $F_{k-1}>F_{k-1,n-p}(\alpha)$ 时认为两模型有显著区别

一般线性假设与应用

思考：考虑 $y=X\beta+e$，$e\sim N(0,\sigma^2 I)$. 检验 β 是否满足假设 $H_0:A\beta=b$.

若不考虑限制有 $RSS=y^T(I-X(X^TX)^{-1}X^T)y$

附加约束性假设，有 $RSS_H=(y-X\hat{\beta}_H)^T(y-X\hat{\beta}_H)$，$\hat{\beta}_H=\hat{\beta}-(X^TX)^{-1}A^T(A(X^TX)^{-1}A^T)^{-1}(A\hat{\beta}-b)$

} 明显地，加约束条件后 β 变化范围缩小，则 RSS_H 变大. 故 $RSS_H\geq RSS$

若参数的确满足 $A\beta=b$，对数据拟合程度一样，RSS_H-RSS 较小，反之较大. 于是我们在 RSS_H-RSS 较大时拒绝 H_0.

定理：① $RSS/\sigma^2\sim\chi^2_{n-p}$ ② 若 $A\beta=b$ 成立，$\frac{RSS_H-RSS}{\sigma^2}\sim\chi^2_m$ ③ RSS 与 RSS_H-RSS 相互独立 ④ $A\beta=b$ 时 $\frac{(RSS_H-RSS)/m}{RSS/(n-p)}\sim F_{m,n-p}$. $(m=\mathrm{rank}A)$

tip：由④知 $\frac{(RSS_H-RSS)/m}{RSS/(n-p)}>F_{m,n-p}(\alpha)$ 时拒绝 H_0，$RSS_H-RSS=(A\hat{\beta}-b)^T(A(X^TX)^{-1}A^T)^{-1}(A\hat{\beta}-b)$

回归方程的显著性检验

思考：$H:\beta_1=\cdots=\beta_{p-1}=0$. 若拒绝 H_0，Y 线性依赖于至少一个 X；若接受 H，可认为相对于误差而言，所有自变量与 Y 的影响不重要.

明显地，这是上一节中 $A=(0,I_{p-1})$，$b=0$ 的特殊情况，于是 $m=p-1$

结论：由上述分析，$\frac{(RSS_H-RSS)/(p-1)}{RSS/(n-p)}>F_{p-1,n-p}(\alpha)$ 时拒绝 H.

因变量的预测

考虑线性回归模型 $y_i=(1,\xi_i^T)\beta+e_i$　$e_i\in N(0,\sigma^2)$，$1\leq i\leq n$. 给定 $\xi_0=(x_{01},\cdots,x_{0,p-1})^T$，$y_0$ 未知. 若也满足 $y_0=(1,\xi_0^T)\beta+e_0$. $e_0\in N(0,\sigma^2)$

结论：y_0 的点预测：$\hat{y}_0=(1,\xi_0^T)\hat{\beta}$.

性质：① $\hat{y}_0$ 为 y_0 的无偏预测 ② 在 y_0 的一切线性无偏预测中，$\hat{y}_0$ 具有最小方差. tip：此处"无偏"指二者均值相同.

定义：使预测值落在其中的概率达到预定值的区间称为预测区间.

记 $d_0=\frac{1}{n}+(\xi_0-\vec{m})^T(X_c^TX_c)^{-1}(\xi_0-\vec{m})$

y_0 的预测区间：σ^2 已知时 $\frac{y_0-\hat{y}_0}{\sigma\sqrt{1+d_0}}\sim N(0,1)$，置信度为 $1-\alpha$ 的置信区间为 $[\hat{y}_0-\sigma\sqrt{1+d_0}\,u_{\frac{\alpha}{2}},\ \hat{y}_0+\sigma\sqrt{1+d_0}\,u_{\frac{\alpha}{2}}]$

σ^2 未知时 $\frac{y_0-\hat{y}_0}{\hat{\sigma}\sqrt{1+d_0}}\sim t_{n-p}$，置信度为 $1-\alpha$ 的置信区间为 $[\hat{y}_0-\hat{\sigma}\sqrt{1+d_0}\,t_{n-p}(\frac{\alpha}{2}),\ \hat{y}_0+\hat{\sigma}\sqrt{1+d_0}\,t_{n-p}(\frac{\alpha}{2})]$　$(\frac{(n-p)\hat{\sigma}^2}{\sigma^2}\sim\chi^2_{n-p})$

$E[y_0]$ 的预测区间：σ^2 已知时 $\frac{E[y_0]-\hat{y}_0}{\sigma\sqrt{d_0}}\sim N(0,1)$，置信度为 $1-\alpha$ 的置信区间为 $[\hat{y}_0-\sigma\sqrt{d_0}\,u_{\frac{\alpha}{2}},\ \hat{y}_0+\sigma\sqrt{d_0}\,u_{\frac{\alpha}{2}}]$

σ^2 未知时 $\frac{E[y_0]-\hat{y}_0}{\hat{\sigma}\sqrt{d_0}}\sim t_{n-p}$，置信度为 $1-\alpha$ 的置信区间为 $[\hat{y}_0-\hat{\sigma}\sqrt{d_0}\,t_{n-p}(\frac{\alpha}{2}),\ \hat{y}_0+\hat{\sigma}\sqrt{d_0}\,t_{n-p}(\frac{\alpha}{2})]$　$(\frac{(n-p)\hat{\sigma}^2}{\sigma^2}\sim\chi^2_{n-p})$

共线性相关

定义：设 $\vec{\theta}$ 为未知参数向量，$\hat{\theta}$ 为 $\vec{\theta}$ 的一个估计，$\hat{\theta}$ 的均方误差定义为 $MSE(\hat{\theta})=E[\|\hat{\theta}-\vec{\theta}\|^2]$，$\|\cdot\|$ 为 $\mathbb{R}^p$ 上的欧氏距离

定理：$MSE(\tilde{\theta})=\mathrm{tr}\,Cov(\tilde{\theta})+\|E[\tilde{\theta}]-\vec{\theta}\|^2$

考虑 $\vec{y}=\alpha 1_n+X_s\beta_s+e$　X_s 为中心化标准化后的设计矩阵 $X_{s,ij}=\frac{x_{ij}-m_j}{s_j}$　$s_j=\sqrt{S_{x_jx_j}}$. $\hat{\alpha}=\bar{y}$. $\hat{\beta}_s=(X_s^TX_s)^{-1}X_s^T\vec{y}$ 为 β_s 的无偏估计

$MSE(\hat{\beta}_s)=\mathrm{tr}Cov(\hat{\beta}_s)=\sigma^2\mathrm{tr}(X_s^TX_s)^{-1}=\sigma^2\sum_{i=1}^{p-1}\frac{1}{\lambda_i}$　λ_i 为 $X_s^TX_s$ 的特征值

定义：$X_s^TX_s$ 的特征值为 $\lambda_1\geq\cdots\geq\lambda_{p-1}>0$. 条件数 $\kappa=\frac{\lambda_1}{\lambda_{p-1}}$ 为最大与最小特征值之比
- $\kappa<100$ 认为复共线性程度很小
- $100\leq\kappa\leq1000$ 认为复共线性程度较强
- $\kappa>1000$ 认为复共线性程度严重

问题：认为预测变量 $x_1,\cdots,x_{p-1}$ 之间存在较强/严重的共线性时，怎么找共线性关系？

回答：对正定阵 $X_s^TX_s$，$\exists$ 正交阵 Φ s.t. $X_s^TX_s=\Phi\,\mathrm{diag}(\lambda_1,\cdots,\lambda_{p-1})\Phi^T$. 其中 $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_{p-1}>0$.

若 $\lambda_j\approx0$. $\vec{\varphi}_j=(\varphi_{j,1},\cdots,\varphi_{j,p-1})^T$. $X_s=(\vec{x}_{s,1},\vec{x}_{s,2},\cdots,\vec{x}_{s,p-1})$. 有 $\varphi_{j,1}\vec{x}_{s,1}+\cdots+\varphi_{j,p-1}\vec{x}_{s,p-1}\approx\vec{0} \Rightarrow \varphi_{j,1}\frac{X_1-m_1}{s_1}+\cdots+\varphi_{j,p-1}\frac{X_{p-1}-m_{p-1}}{s_{p-1}}\approx0$

Example (4.2.1)

【[0]例3.7.1】考虑一个有六个回归自变量的线性回归模型，

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\beta_5X_5+\beta_6X_6+e,$$

原始数据见[0]表3.7.1，共12组数据。
对于如上模型，记中心化和标准化设计阵为$\mathbf{X}_s$，则自变量$X_1,\cdots,X_6$的样本相关系数矩阵$\mathbf{X}_s^T\mathbf{X}_s$的上三角阵为

$$\begin{array}{cccccc}1.000 & 0.052 & -0.343 & -0.498 & 0.417 & -0.192\\ & 1.000 & -0.432 & -0.371 & 0.485 & -0.317\\ & & 1.000 & -0.355 & -0.505 & 0.494\\ & & & 1.000 & -0.215 & -0.087\\ & & & & 1.000 & -0.123\\ & & & & & 1.000\end{array}$$

注 从非对角元的绝对值看，任两个回归自变量之间似乎不存在较严重的线性依赖关系，但事实是，多个自变量之间存在严重的共线性。

- 计算矩阵$\mathbf{X}_s^T\mathbf{X}_s$的六个特征根，分别为

$$\lambda_1=2.24879,\lambda_2=1.54615,\lambda_3=0.92208,$$
$$\lambda_4=0.79399,\lambda_5=0.30789,\lambda_6=0.00111.$$

- 样本均值和样本标准方差

$$m_1=2.5,\quad m_2=2,\quad m_3=2.5,\quad m_4=3.083,$$
$$s_1=11.358,\quad s_2=10.1,\quad s_3=13.077,\quad s_4=13.89.$$

- $\lambda_6\approx0$，因此存在共线性关系；同时共线性的度量：条件数

$$\kappa=\frac{\lambda_1}{\lambda_6}=\frac{2.24879}{0.00111}=2025.94>1000.$$

- 根据我们本节约定的标准，模型的设计阵（或自变量之间）存在严重的（复）共线性。
- 利用$\lambda_6=0.00111\approx0$，我们进一步算出$\mathbf{X}_s^T\mathbf{X}_s$对应于$\lambda_6$的特征向量为

$$\vec{\varphi}=(-0.45,-0.42.-0.54,-0.57,-0.006,-0.002)^T,$$

- 因而回归自变量之间有如下（复）共线关系

$$0.45\frac{X_1-m_1}{s_1}+0.42\frac{X_2-m_2}{s_2}+0.54\frac{X_3-m_3}{s_3}+0.57\frac{X_4-m_4}{s_4}\approx0. \tag{4}$$

其中由于$\varphi_5,\varphi_6\approx0$，已删除相关变量。

- 代入样本均值和样本标准差，可得

$$0.0396X_1+0.0416X_2+0.0413X_3+0.0410X_4\approx0.4120$$

岭估计

考虑回归模型 $\hat{y}=\alpha 1_n+X_s\beta_s+\vec{e}$. $\vec{e}\sim N(0,\sigma^2I_n)$

定义：对上述模型，$\vec{\beta}_s$的岭估计为 $\hat{\beta}_s(k)=(X_s^TX_s+kI_{p-1})^{-1}X_s^Ty$

这里$k>0$称岭参数或随参数

$\hat{\beta}_s(k)$是一个估计类，$\hat{\beta}_s(0)$即为平常的OLS估计

定理：在MSE意义下，岭估计优于OLS估计. 即 $\exists k>0$. $MSE(\hat{\beta}_s(k))<MSE(\hat{\beta}_s)$

岭参数的选择

Hoerl-Kennard公式：$\hat{k}=\frac{\hat{\sigma}^2}{\max\limits_i\hat{\alpha}_i^2}$ $\vec{\alpha}=T^T\hat{\beta}_s$ 称为典则回归系数 → 一些公式：$\hat{\alpha}_j=\hat{\beta}_j\cdot\frac{s_j}{s_y}$ $\hat{\beta}_j(k)=\hat{\alpha}_j(k)\cdot\frac{s_y}{s_j}$. $\hat{\beta}_0(k)=\bar{y}-\sum_{j=1}^{p-1}\hat{\beta}_j(k)\bar{x}_j$

岭迹法：思想：找参数k，使得回归系数估计值趋于稳定，且k的值不至于过大

定义：记$\hat{\beta}_{sj}(k)$为$\hat{\beta}_s(k)$的第j分量，为k一元函数，$k\in[0,\infty)$变化时，$\hat{\beta}_{sj}(k)$的图形称为岭迹.

- 岭迹法选择k的具体操作：
1. 将$\hat{\beta}_{s1}(k),\ldots,\hat{\beta}_{s,p-1}(k)$的岭迹画在同一图上；
2. 根据岭迹的变化趋势选择k值，使得各个回归系数的岭估计大体上稳定，并且各个回归系数岭估计值的符号比较合理；
3. 同时选取的k尽量接近零，使得残差平方和不要上升太多。

主成分估计

考虑线性回归模型. $y=\alpha 1_n+X_s\beta_s+e$ $e\sim N(0,\sigma^2I_n)$. $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_p>0$ $X_s^TX_s$的特征值. $\varphi_1,\cdots,\varphi_p$为相应正交化特征向量. $T=(\varphi_1,\cdots,\varphi_p)$为$p\times p$正交阵.

记$Z=X_sT$. $\alpha=T^T\beta_s$. 则上述模型为 $y=\alpha_0 1_n+Z\alpha+e$. α为典则回归系数.

定义：$Z=(\vec{z}_1,\vec{z}_2,\cdots,\vec{z}_{p-1})$. 称$\vec{z}_k$为第k主成分

算法：将对应特征值近似为0的$(p-1)-r$个主成分从模型中剔除，再对剩下的r个主成分做OLS估计. 得到主成分回归方程. 此过程称为主成分回归

定理：当设计阵存在复共线关系时，适当选择保留的主成分个数可使主成分估计有比OLS估计更小的MSE. 即 $MSE(\hat{\beta}_{s,p})<MSE(\hat{\beta}_s)$

$\frac{\hat{\beta}_j}{\hat{\sigma}\sqrt{c_{jj}}}$ $C_{00}=\frac{1}{n}+\bar{x}^TA_{xx}^{-1}\bar{x}$ $\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}$

1. 单个检验：$H_0:\beta_j=0$，则取 $t_j=\frac{\hat{\beta}_j}{\hat{\sigma}\sqrt{C_{jj}}}\sim t_{n-p}(\frac{\alpha}{2})$ $\hat{\sigma}=\sqrt{MSE}=\sqrt{\frac{RSS}{n-p}}$ $C_{00}=\frac{1}{n}+\bar{m}^T(X_c^TX_c)^{-1}\bar{m}$ （一元为 $\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}$）

C_{ii} 为 $(X_c^TX_c)^{-1}$ 的第 i 对角元（一元为 $\frac{1}{S_{xx}}$）（$S_{xx}=X_c^TX_c$）

$R_{SS}=\frac{S_{xy}^2}{S_{xx}}$

2. $SYY=SS_{reg}+RSS \Rightarrow$ 残差均方 $MSE=\frac{RSS}{df_E}$

$\Downarrow$

回归均方 $MS_{reg}=\frac{SS_{reg}}{df_R}$ ↗ 一定是所有自变量系数均为0.全回归方程的检验

对检验 $H_0:\beta_1=\cdots=\beta_k=0$. 则 $df_R=k$. $df_E=n-k-1$. F统计量：$F=\frac{MS_{reg}}{MSE}\sim F_{k,n-k-1}(\alpha)$ （$n-p$）

表1

	来源	d.f.	SS	MS	F
$p-1$	关于所有X的回归	δ_1 k	δ_2 SYY-RSS（回归平方和 SS_{reg}）	δ_3 MS_{reg}	δ_4 = $\frac{MS_{reg}}{MSE}=\frac{SS_{reg}/(p-1)}{RSS/(n-p)}$
$n-p$	残差	δ_5 n-p	δ_6 (n-p) MSE	4816	
$n-1$	总的	δ_7 n-1	216571（残差平方和 RSS）		

- 记回归均方为 $MS_{reg}=SS_{reg}/(p-1)$，则上述讨论可归结为如下表
- 方差分析表（ANOVA）

来源	d.f.	SS	MS	F
X的回归	$p-1$	SS_{reg}	$SS_{reg}/(p-1)$	$MS_{reg}/\hat{\sigma}^2$
残差	$n-p$	RSS	$RSS/(n-p)$	
总的	$n-1$	SYY		

$F=\frac{(RSS_H-RSS)/m}{RSS/(n-p)}$ → 受检验影响

$\frac{(RSS_H-RSS)/m}{RSS/(n-p)}$ ↘ 除去受检验部分. P为总的自变量数. 如 $Y=\beta_0+\beta_1X+e$，$Y=\beta_0+e$ 中 $p=2$

如：$Y=\beta_0+\beta_1X_1+\beta_2X_2+e$ 总为 $Y=\beta_0+\cdots+\beta_5X_5+e$.

$Y=\beta_0+\beta_1X_1+e$

$\frac{(A\hat{\beta}-b)^T(A(X^TX)^{-1}A^T)^{-1}(A\hat{\beta}-b)/m}{RSS/(n-p)}\sim F_{m,n-p}$

则分子为 X_1（新增变量）的 RSS，$m=1$ 为新增自变量数

分母为 X_3,X_4,X_5 的 RSS.（$RSS_{full}=SYY-SS_{reg}(3,4,5)$） $n-p=n-3$ $p=3$ 为检验中总变量数（β_0,β_1,β_2）

3. 线性约束检验统计量：

对 $Y=X\beta+e$. $e\sim N(0,\sigma^2I_n)$，$H_0:A\beta=b$ 的F统计量为 $F=\frac{(A\hat{\beta}-b)^T(A(X^TX)^{-1}A^T)^{-1}(A\hat{\beta}-b)/m}{\hat{\sigma}^2}\sim F_{m,n-p}$. m为约束个数 = rank A $\hat{\sigma}^2=\frac{RSS}{n-p}$

tip：若A只对截距项. $A=(a_0,0,0,\cdots)^T$，则 $A(X^TX)^{-1}A^T=\frac{1}{n}+m^T(X_c^TX_c)^{-1}m$ $m=(m_1,\cdots,m_{p-1})$

$d_0=\frac{1}{n}+(\xi-\bar{\xi})^T(X_c^TX_c)^{-1}(\xi-\bar{\xi})$ 不对 ‥ $A=(0,a_1,\cdots)^T$，则 ‥ $=(a_1,\cdots)(X_c^TX_c)^{-1}(a_1,\cdots)^T$

不过若A不对截距项.一般用 $X_c^TX_c$ 计算.会直接消去.不存在此情况. $A\hat{\beta}-b\to A\hat{\beta}_c-b$. $\hat{\beta}_c$ 为去除 $\hat{\beta}_0$ 的 $(\hat{\beta}_1,\cdots)$

$(X^TX)^{-1}\to(X_c^TX_c)^{-1}$

4. $d_0=\frac{1}{n}+(\xi_0-\bar{\xi})^T(X_c^TX_c)^{-1}(\xi_0-\bar{\xi})$ $d_0=\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}$

$\frac{1}{n}+(\xi-\bar{\xi})^T(X_c^TX_c)^{-1}(\xi-\bar{\xi})$ σ^2 未知. y_0 预测：$\hat{y}_0\pm\hat{\sigma}\sqrt{1+d_0}\cdot t_{n-p}(\frac{\alpha}{2})$ $E(y_0)$：$\hat{y}_0\pm\hat{\sigma}\sqrt{d_0}\cdot t_{n-p}(\frac{\alpha}{2})$ $\hat{y}_0\pm\hat{\sigma}\sqrt{1+d_0}\,t_{n-p}(\frac{\alpha}{2})$

$\hat{y}_0\pm\hat{\sigma}\sqrt{1+d_0}\,t_{n-p}(\frac{\alpha}{2})$ β_j 预测：$\hat{\beta}_j\pm\hat{\sigma}\sqrt{C_{jj}}\cdot t_{n-p}(\frac{\alpha}{2})$ $\hat{\beta}_j\pm\hat{\sigma}\sqrt{C_{jj}}\,t_{n-p}(\frac{\alpha}{2})$

5. 共线性：若 $\lambda_i\approx0$，可判断有共线性. $k_i=\frac{\lambda_1}{\lambda_i}>1000$ 则有强共线性. 记标准化变量为 $Z_i=\frac{X_i-m_i}{S_i}$

寻找具体共线性关系：若 $\vec{\varphi}=(a_1,\cdots,a_m)$，则为 $a_1Z_1+\cdots+a_mZ_m=0$（可去除 <0.01 的 a_i 部分）

6. 系数表中 t检验 $=\frac{系数}{标准误}$

7. ① 岭参数与岭回归方程的计算：

$\hat{\beta}_c\to\hat{\beta}_s\to\hat{\alpha}=\Phi^T\hat{\beta}_s\to\hat{k}=\frac{\hat{\sigma}^2}{\max\hat{\alpha}_i^2}$

若有中心化后OLS系数（≈原系数）$\hat{\beta}_c$，则有 $\hat{\beta}_s=\hat{\beta}_c\odot(s_1,s_2,\cdots)^T$，有主成分坐标下 $\hat{\alpha}=\Phi^T\hat{\beta}_s$. Φ 为特征向量矩阵（↗表示逐元素相乘）

$\downarrow$ $\hat{\alpha}(\hat{k})=(kI+\Lambda)^{-1}\Lambda\hat{\alpha}$

由 Hoerl-Kennard 公式：$\hat{k}=\frac{\hat{\sigma}^2}{\max_i\hat{\alpha}_i^2}$ 求得 $\hat{k}$. 则岭估计在主成分坐标下为 $\hat{\alpha}(\hat{k})=(\Lambda+\hat{k}I)^{-1}\Lambda\hat{\alpha}$（$\Lambda$ 为 $diag(\lambda_1,\cdots)$ 特征值对角阵）

$\downarrow$ $\hat{\beta}_s(\hat{k})=\Phi\hat{\alpha}(\hat{k})$

引进线性回归模型(1)的典则形式：

$$\vec{y}=\alpha_0\vec{1}_n+Z\vec{\alpha}+\vec{e},\ E(\vec{e})=0,\ Cov(\vec{e})=\sigma^2I. \quad (3)$$

其中设计阵 $Z=X_s\Phi$，而 $\vec{\alpha}=\Phi^T\vec{\beta}_s$ 称为**典则回归系数**。这里 $\Phi=(\vec{\varphi}_1,\cdots,\vec{\varphi}_{p-1})$，$\Lambda=diag(\lambda_1,\cdots,\lambda_{p-1})$，其中 $\lambda_1\geq\cdots\geq\lambda_{p-1}$ 为 $X_s^TX_s$ 的特征根，$\vec{\varphi}_1,\cdots,\vec{\varphi}_{p-1}$ 为对应的标准正交化特征向量。

标准化的岭参数：$\hat{\beta}_s(\hat{k})=\Phi\hat{\alpha}(\hat{k})$ 原尺度：$\hat{\beta}_c(k)=\hat{\beta}_s(k)\oslash(s_1,s_2,\cdots s_j)^T$ $\downarrow$ $\hat{\beta}_c(\hat{k})=\hat{\beta}_s(\hat{k})\oslash(s_1,\cdots)$

↘表示逐元素相除

岭 $\sum\frac{k\lambda_i}{(\lambda_i+k)^2}\hat{\alpha}_i^2$ 对 $\bar{y}=\hat{\beta}_0+\bar{m}^T\hat{\beta}_c$. 有岭回归方程：$\hat{Y}=\bar{y}+\hat{\beta}_{c_1}(\hat{k})(X_1-m_1)+\cdots$ $\downarrow$ $\bar{y}=\hat{\beta}_0+\bar{m}\cdot\hat{\beta}_c(\hat{k})$

② 岭回归造成的RSS增加：OLS: RSS(0) 岭: RSS(k) $RSS(k)-RSS(0)=k^2\sum_{i=1}^{r}\frac{\lambda_i}{(\lambda_i+k)^2}\hat{\alpha}_i^2$ 代入 $\hat{k}$ 即可 $\downarrow$ $\hat{y}=\bar{y}+\hat{\beta}_{c_1}(\hat{k})(X_1-m_1)+\cdots$

8. ① 主成分回归方程：先找 累计解释比例达到的.若 r 个贡献率 ≈1 $\tilde{\beta}_{s,p}=\Phi_r\hat{\alpha}_r$（前r项） $\tilde{\beta}_{c,p}=\tilde{\beta}_{s,p}\oslash(s_1,\cdots,s_r)^T$

则 $\tilde{Y}=\bar{y}+\hat{\beta}_{c,p_1}(X_1-m_1)+\cdots$

若舍去第 m 个主成分

② 主成分回归造成的RSS增加：$RSS_p-RSS=\lambda_m\hat{\alpha}_m^2$

R^2 损失：$SYY=SS_{reg}+RSS$，$R^2-R_p^2=\frac{RSS_p-RSS}{SYY}$ $RSS_p-RSS=\sum_{舍去}\lambda_m\hat{\alpha}_m^2$ $\sum\lambda_m\frac{k^2}{(\lambda_m+k)^2}\hat{\alpha}_m^2$ ✓岭

系数估计向量长度缩减：OLS：$\|\hat{\beta}_c\|$. 主：$\|\hat{\beta}_{c,p}\|$ ⇒ 缩了 $\|\hat{\beta}_c\|-\|\hat{\beta}_{c,p}\|$

9. $SS_{reg}=(X_c^Ty)^T(X_c^TX_c)^{-1}(X_c^Ty)$ ⇒ 对子集S.有 G_s 为 $X_c^TX_c$ 对应的主子矩阵. V_s 为 X_c^Ty 对应分量. $SS_{reg}(S)=V_s^TG_s^{-1}V_s$ $RSS(S)=SYY-SS_{reg}(S)$ $X_c^Ty=X_c^T\hat{y}=X_c^TX_c\hat{\beta}_c$

模型公式：q为自变量数+1.（含截距） （$\hat{\sigma}_{full}^2=\frac{RSS}{n-p}$ 对应的全模型）（此p不是子集S的p.是全模型的p）

$RMS_q=\frac{RSS_q}{n-q}$ $C_p=\frac{RSS_q}{\hat{\sigma}_{full}^2}-n+2q$. $AIC=n\ln RSS_q+2q$ $BIC=n\ln RSS_q+q\ln n$

10. 变量选择法.

方差分析表			
方差来源	自由度	平方和	均方
回归	3	73.506	24.502
残差	66	282.873	4.286
系数表			
变量	系数	标准误差	t检验
常数项	-0.070	0.251	-0.28
FAM	1.101	1.411	0.78
PEER	2.322	1.481	1.57
SCHOOL	-2.281	2.220	-1.03

Table: 4.1.1

问题 考察学校资源设施(*SCHOOL*)这个变量在模型(1)中的作用，检验其是否有存在的必要，检验水平$\alpha = 0.05$。

- 问题等价于检验$H_0 : \beta_3 = 0$，检验统计量

$$|t_3| = |\frac{\hat{\beta}_3}{\sqrt{\widehat{var}(\hat{\beta}_3)}}| = |-\frac{2.281}{2.220}| = 1.03 < t_{66}(\frac{0.05}{2}) = 1.997.$$

因此接受H_0，即*SCHOOL*这一变量在模型(1)中没有存在的必要。

注 这里$\widehat{var}$指代样本方差，从而$\widehat{var}(\hat{\beta}_3) = c_{3,3}\hat{\sigma}^2$，其中$c_{3,3}$是校正叉积矩阵的逆$(X_c^T X_c)^{-1}$的第三个对角元。

- 从而

$$MSE\left(\hat{\vec{\beta}}_s\right) = \sigma^2 tr\{\left(\mathbf{X}_s^T\mathbf{X}_s\right)^{-1}\} = \sigma^2\sum_{j=1}^{p-1}\frac{1}{\lambda_j}.$$

- 由此可得
  1. 如果$\mathbf{X}_s^T\mathbf{X}_s$至少有一个特征根非常小，即非常接近于零，那么$MSE\left(\hat{\vec{\beta}}_s\right)$就会很大。从均方误差的标准来看，这时的最小二乘估计$\hat{\vec{\beta}}$就不是一个好的估计。

注意 这一点和Gauss-Markov定理并无抵触，因为Gauss-Markov定理仅仅保证了最小二乘(OLS)估计在线性无偏估计类中的方差最小性。但在$\mathbf{X}_s^T\mathbf{X}_s$至少有一个特征值很小时，这个最小的方差值本身却很大，因而导致了很大的均方误差。

  2. 对最小二乘估计模长的影响: 由

$$\mathbb{E}\left\|\hat{\vec{\beta}}_s\right\|^2 = \left\|\vec{\beta}_s\right\|^2 + MSE\left(\hat{\vec{\beta}}_s\right) = \left\|\vec{\beta}_s\right\|^2 + \sigma^2\sum_{j=1}^{p-1}\frac{1}{\lambda_j}.$$

可知，当$\mathbf{X}_s^T\mathbf{X}_s$至少有一个特征根很小时，OLS估计$\hat{\vec{\beta}}_s$的长度平均说来要比真正的未知向量$\vec{\beta}_s$的长度大得多。这就导致了$\hat{\vec{\beta}}$的某些分量的绝对值太大。

总结 当$\mathbf{X}_s^T\mathbf{X}_s$至少有一个特征值很小时，最小二乘(OLS)估计不再是一个好的估计。

Definition (4.3.2)

记$\hat{\beta}_{s,j}(k)$为$\hat{\vec{\beta}}_s(k)$的第j个分量，它是k的一元函数。当k在$[0,\infty)$上变化时，$\hat{\vec{\beta}}_{s,j}(k)$的图形称为**岭迹**（ridge trace）。

- 岭迹法选择k的具体操作：
  1. 将$\hat{\beta}_{s,1}(k),\ldots,\hat{\beta}_{s,p-1}(k)$的岭迹画在同一图上；
  2. 根据岭迹的变化趋势选择k值，使得各个回归系数的岭估计大体上稳定，并且各个回归系数岭估计值的符号比较合理；
  3. 同时选取的k尽量接近零，使得残差平方和不要上升太多。

总结 岭回归方法可对线性回归方程通常的最小二乘（OLS）估计的稳定性提供一个判断的工具：

- 在严重复共线性的情况下，数据的一个微小的变动（扰动）会造成回归系数估计的很大的变动，而岭回归能揭示这种现象。
- 相对于OLS估计来说，岭回归估计具有抵抗数据扰动的稳健性。
- 岭回归的**稳健性定义**：它不受数据中微小的变动的影响。

- 相对于OLS估计，岭估计具有更小的均方误差，因此
  - 回归系数的岭估计更倾向于接近回归系数的真实值；
  - 对于不在数据集中的预测变量的值，预测其相应的响应的变量的值时，用岭回归方法去预测将获得更精确地预报结果。

指标	OLS ($k=0$)	最优k的岭回归	k持续增大
训练 RSS	全局最小值	变大	持续单调增大
参数偏差	0（无偏）	轻微>0	持续单调飙升
参数方差	最大、易爆炸	大幅缩小	持续单调减小
参数总 MSE	基准很高	显著变小	先降后升，过大k后反弹

一句话终极总结

岭回归本质是**主动牺牲一点点训练拟合精度（RSS变大）、换取参数稳定性，解决OLS多重共线性方差爆炸的问题，最终实现整体估计误差MSE下降、泛化预测能力变强。**

C_p准则选择的标准之一：C_p与q的靠近程度。

图形法进行变量子集的选择（C_p图）：

- 对于每一个变量子集，在二维图上画一个点(q, C_p)，
- 在图上画一条直线$C_p(q) = q$，即第一象限平分线；
- 选点(q, C_p)，最接近第一象限平分线且C_p值最小的的那个点所对应的自变量子集就是建立回归模型的最优子集。

- 问题：当自变量个数很多时，计算所有可能子集回归的计算量很大。
- 解决：不需要计算所有可能子集回归的变量选择算法————逐步回归算法。
- **基本思想**：
  - 将变量一个一个引入，引入变量的条件是其偏回归平方和经检验是显著的。
  - 同时，每引入一个新的变量，对已入选方程的老变量逐个进行检验，将经检验认为不显著的变量剔除，以保证所得自变量子集中的每一个变量都是显著的。
  - 此过程经若干步直到不能再引入新变量，也没有已入选方程的老变量需要剔除为止。

注 经逐步回归后，最终，回归方程中所有自变量对因变量Y都是显著的，而不在回归方程中的变量对Y经检验均不显著。

- 逐步回归法选择变量的两个基本步骤：
  1. 引入新变量到回归方程中；
  2. 从回归方程中剔除经检验不显著的变量。

一 利用主成分回归处理预测变量之间的非正交性，这种非正交性导致模型的共线性。

- 利用主成分的方法可以减缓数据中的共线性现象。
- 其方法就是用部分主成分去解释预测变量的变动。
- 若主成分回归中使用了全部主成分作为解释变量，那么主成分回归就等价于通常的最小二乘法。

二 主成分回归与回归系数$\hat{\vec{\beta}}_s$之间的约束条件有关。

岭回归：

$$RSS_R(k) = RSS_{OLS} + \sum_{j=1}^{p-1}\frac{k^2\lambda_j}{(\lambda_j+k)^2}\hat{\alpha}_j^2.$$

主成分回归：

$$RSS_P(r) = RSS_{OLS} + \sum_{j=r+1}^{p-1}\lambda_j\hat{\alpha}_j^2.$$

岭回归是对所有主成分方向做连续收缩；主成分回归是直接舍去后面的主成分方向。因此二者的 RSS 增量分别来自：

$$\text{岭回归：}\quad \sum_{j=1}^{p-1}\frac{k^2\lambda_j}{(\lambda_j+k)^2}\hat{\alpha}_j^2,$$

$$\text{主成分回归：}\quad \sum_{j=r+1}^{p-1}\lambda_j\hat{\alpha}_j^2.$$

二者差别在于：

1. 岭回归不删除任何主成分，而是把第 j 个方向按
$$\frac{\lambda_j}{\lambda_j+k}$$
进行连续收缩；
2. 主成分回归直接保留前 r 个主成分，并把后面的主成分系数估计为 0；
3. 岭回归的偏差来自所有方向的收缩；
4. 主成分回归的偏差只来自被舍去的主成分方向。