课程主页：http://staff.ustc.edu.cn/~ynyang/2022

第十四讲 广义最小二乘

2022.12.2

$$var(x_i) = \sigma_i^2, var\left(\frac{\Sigma x_i/\sigma_i^2}{1/\sigma_i^2}\right) \le var(\bar{x})$$

一个简单例子

例1. 若 $y_1,\ldots,y_n$独立, $y_i \sim (\mu,\sigma_i^2)$, $\underline{\mu = E(y_i)\text{未知}, \sigma_i^2 = \mathrm{var}(\sigma_i^2)\text{已知}}$,
则μ可由OLS或GLS/WLS估计:

普通的最小二乘法(Ordinary LS, OLS): $\min\limits_{\mu} \sum (y_i - \mu)^2 \Rightarrow \hat{\mu}_{OLS} = \bar{y}$

直观上，因为诸y_i方差不等，μ的估计中方差大的数据 应该给予较小权重.

加权最小二乘 (Weighted LS, WLS), 以$1/\sigma_i^2, i = 1,2,\ldots$作为权重:

$$\min_{\mu} \sum_{i=1}^{n} \left(\frac{y_i - \mu}{\sigma_i} \right)^2 = \min_{\mu} \sum_{i=1}^{n} \frac{(y_i - \mu)^2}{\sigma_i^2} \Rightarrow \hat{\mu}_{WLS} = \frac{\sum y_i / \sigma_i^2}{\sum 1/\sigma_i^2}$$

WLS估计方差更小: $\mathrm{var}(\hat{\mu}_{WLS}) = \dfrac{1}{\sum 1/\sigma_i^2} \le \dfrac{\sum \sigma_i^2}{n^2} = \mathrm{var}(\hat{\mu}_{OLS})$.

广义最小二乘（GLS）

异方差模型（方差几乎完全已知）

假设如下异方差模型（误差方差除了一个常数倍之外完全已知）：

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon},\quad \mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0},\quad \mathrm{var}(\boldsymbol{\varepsilon}) = G = \sigma^2 G_0 > 0\ ,\ \boldsymbol{\varepsilon}\text{与}X\text{独立}$$

假设$\boxed{G_0\text{已知},\ \sigma^2\text{未知.}}$

重要情形：

(1) $G = \sigma^2 I_n$,即通常的GM假设。

(2) 抽样调查数据，$G = \sigma^2 diag(1/w_1,\dots,1/w_n), w_1,\dots,w_n$已知权重

因为方差齐性模型中LS估计是最优的，因此我们将其转换为齐性模型：

方程 $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ 两边左乘$G_0^{-1/2}$，记$\tilde{\mathbf{y}} \triangleq G_0^{-1/2}\mathbf{y}, \tilde{X} = G_0^{-1/2}X, \tilde{\boldsymbol{\varepsilon}} = G_0^{-1/2}\boldsymbol{\varepsilon}$，得

$$\tilde{\mathbf{y}} = \tilde{X}\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}},\quad \tilde{\boldsymbol{\varepsilon}} \sim (\mathbf{0}, \sigma^2 I_n) \qquad (*)$$

对(*)应用普通LS (OLS, Ordinary LS)，其误差平方和为

$$(\tilde{\mathbf{y}} - \tilde{X}\boldsymbol{\beta})^\top(\tilde{\mathbf{y}} - \tilde{X}\boldsymbol{\beta}) = (\mathbf{y} - X\boldsymbol{\beta})^\top G_0^{-1}(\mathbf{y} - X\boldsymbol{\beta})$$

所以对模型(*)应用OLS方法（左端）$\Leftrightarrow$ 对原模型极小化右端函数。

方差齐性模型(*)的OLS解为原模型的GLS解，对

$$\tilde{\mathbf{y}} = \tilde{X}\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}, \quad \tilde{\boldsymbol{\varepsilon}} \sim (\mathbf{0}, \sigma^2 I_n) \qquad (*)$$

应用OLS得：

$$\hat{\boldsymbol{\beta}}_{GLS} = (\tilde{X}^\top \tilde{X})^{-1} \tilde{X}^\top \tilde{\mathbf{y}} = (X^\top G_0^{-1} X)^{-1} X^\top G_0^{-1} \mathbf{y}$$

$$\hat{\sigma}^2_{\mathrm{GLS}} = \| \tilde{\mathbf{y}} - \tilde{X}\hat{\boldsymbol{\beta}}_{GLS} \|^2 /(n-p) = (\mathbf{y} - X\hat{\boldsymbol{\beta}}_{GLS})^\top G_0^{-1} (\mathbf{y} - X\hat{\boldsymbol{\beta}}_{GLS}) /(n-p).$$

广义最小二乘

广义最小二乘法（GLS）: GLS极小化加权误差平方和

$$\min_{\boldsymbol{\beta}} (\mathbf{y} - X\boldsymbol{\beta})^\top G^{-1} (\mathbf{y} - X\boldsymbol{\beta}) \Leftrightarrow \min_{\boldsymbol{\beta}} (\mathbf{y} - X\boldsymbol{\beta})^\top G_0^{-1} (\mathbf{y} - X\boldsymbol{\beta})$$

最优解称为GLS估计：$\hat{\boldsymbol{\beta}}_{\mathrm{GLS}} = (X^\top G_0^{-1} X)^{-1} X^\top G_0^{-1} \mathbf{y}$，

σ^2的GLS估计：$\hat{\sigma}^2_{\mathrm{GLS}} = (\mathbf{y} - X\hat{\boldsymbol{\beta}}_{GLS})^\top G_0^{-1} (\mathbf{y} - X\hat{\boldsymbol{\beta}}_{GLS}) /(n-p).$

GLS是BLUE

命题1. $E(\hat{\boldsymbol\beta}_{\mathrm{GLS}} \mid X)=\boldsymbol\beta$, $\mathrm{var}(\hat{\boldsymbol\beta}_{\mathrm{GLS}} \mid X)=\sigma^2(X^\top G_0^{-1}X)^{-1}$,
$\hat{\boldsymbol\beta}_{\mathrm{GLS}}$是$\boldsymbol\beta$的最优线性无偏估计($BLUE$)。

证明：对模型(*),由GM定理，$\hat{\boldsymbol\beta}_{\mathrm{GLS}}$是BLUE，则对原模型它也是BLUE。

GLS 与OLS

异方差模型$\mathbf{y}=X\boldsymbol\beta+\boldsymbol\varepsilon$, $\boldsymbol\varepsilon\sim(\mathbf{0},\ G=\sigma^2G_0)$也可应用普通的最小二乘法（OLS：Ordinary LS）,应用OLS得

$$\hat{\boldsymbol\beta}_{\mathrm{LS}}=(X^\top X)^{-1}X^\top\mathbf{y}=\mathrm{argmin}\,\|\mathbf{y}-X\boldsymbol\beta\|^2$$

容易验证：$E(\hat{\boldsymbol\beta}_{\mathrm{LS}} \mid X)=\boldsymbol\beta$, $\mathrm{var}(\hat{\boldsymbol\beta}_{\mathrm{LS}} \mid X)=(X^\top X)^{-1}X^\top GX(X^\top X)^{-1}$

由命题1, $\mathrm{var}(\hat{\boldsymbol\beta}_{\mathrm{LS}} \mid X)\ge \mathrm{var}(\hat{\boldsymbol\beta}_{\mathrm{GLS}} \mid X)$，即

$$(X^\top X)^{-1}X^\top GX(X^\top X)^{-1}\ge(X^\top G^{-1}X)^{-1}$$

即，对异方差模型应用OLS是一种低效率（方差较大）的做法。

多元正态

定义：若p维随机向量$\mathbf{x}$的密度为

$$f(\mathbf{x})=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right),\quad \mathbf{x}\in \mathrm{R}^{p}$$

其中参数$\boldsymbol{\mu}\in R^{p}$，$\Sigma$为$p\times p$正定矩阵，则$\mathbf{x}\sim N_p(\boldsymbol{\mu},\Sigma)$服从多元正态分布。

正态模型下的MLE和GLS

假设 $\mathbf{y}=X\boldsymbol{\beta}+\boldsymbol{\varepsilon},\quad \boldsymbol{\varepsilon}\sim N(0,G), G=\sigma^2 G_0\quad (G_0$已知，$\sigma^2$未知$)$,似然函数:

$$L(\boldsymbol{\beta})=\frac{1}{(2\pi)^{n/2}|G|^{1/2}}\exp\left(-\frac{1}{2}(\mathbf{y}-X\boldsymbol{\beta})^{\mathsf{T}}G^{-1}(\mathbf{y}-X\boldsymbol{\beta})\right)$$

极大似然函数等价于极小化：

$$-2\log L=\log|\sigma^2 G_0|+(\mathbf{y}-X\boldsymbol{\beta})^{\mathsf{T}}G_0^{-1}(\mathbf{y}-X\boldsymbol{\beta})/\sigma^2.$$

对$\boldsymbol{\beta}$求导（不依赖于σ^2）：$\max\limits_{\boldsymbol{\beta}} L\Leftrightarrow \min\limits_{\boldsymbol{\beta}}(\mathbf{y}-X\boldsymbol{\beta})^{\mathsf{T}}G_0^{-1}(\mathbf{y}-X\boldsymbol{\beta})\Rightarrow \hat{\boldsymbol{\beta}}_{mle}=\hat{\boldsymbol{\beta}}_{GLS},$

对σ^2求导: $n/\sigma^2-RSS/\sigma^4\Rightarrow \hat{\sigma}^2_{\mathrm{mle}}=(\mathbf{y}-X\hat{\boldsymbol{\beta}}_{GLS})^{\mathsf{T}}G_0^{-1}(\mathbf{y}-X\hat{\boldsymbol{\beta}}_{GLS})/n=\frac{n-p}{n}\hat{\sigma}^2_{\mathrm{GLS}}$

MLE与OLS

特别地，若$G=\sigma^2 I_n$, 极大似然估计：

$$\tilde{\boldsymbol{\beta}}_{mle}=(X^{\mathsf{T}}X)^{-1}X^{\mathsf{T}}\mathbf{y}=\hat{\boldsymbol{\beta}}_{LS},\ \tilde{\sigma}^2_{mle}=\|\mathbf{y}-X\tilde{\boldsymbol{\beta}}_{mle}\|^2/n=\hat{\sigma}^2_{LS}\times\frac{n-p}{n}.$$

WLS与抽样调查

加权最小二乘

假设模型$\mathbf{y}=X\boldsymbol{\beta}+\boldsymbol{\varepsilon}, E(\boldsymbol{\varepsilon})=0,\ G=\mathrm{var}(\boldsymbol{\varepsilon})=\sigma^2 W^{-1}$,其中$W=diag(w_1,\ldots,w_n)$,
$w_1,\ldots,w_n$为已知权重，则加权最小二乘目标函数形式如下：

$$(\mathbf{y}-X\boldsymbol{\beta})^{\mathsf{T}}W(\mathbf{y}-X\boldsymbol{\beta})=\sum_{i=1}^{n}w_i(y_i-\tilde{\mathbf{x}}_i^{\mathsf{T}}\boldsymbol{\beta})^2$$

$$\Rightarrow \hat{\boldsymbol{\beta}}_{WLS}=(X^{\mathsf{T}}W\ X)^{-1}X^{\mathsf{T}}W\mathbf{y}=\left(\sum_{i=1}^{n}w_i\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^{\mathsf{T}}\right)^{-1}\left(\sum_{i=1}^{n}w_i\tilde{\mathbf{x}}_i y_i\right),$$

$$\hat{\sigma}^2_{\mathrm{WLS}}=(\mathbf{y}-X\hat{\boldsymbol{\beta}}_{WLS})^{\mathsf{T}}W(\mathbf{y}-X\hat{\boldsymbol{\beta}}_{WLS})/(n-p)=\sum_{i=1}^{n}w_i(y_i-\tilde{\mathbf{x}}_i^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{WLS})^2/(n-p),$$

其中加权残差$e_i=\sqrt{w_i}(y_i-\tilde{\mathbf{x}}_i^{\mathsf{T}}\hat{\boldsymbol{\beta}}_{WLS})$,

例如，在抽样调查cluster survey中，第i个$cluster$的响应通常是类内所有m_i个个体的数据汇总，比如平均或中位数,自变量$\mathbf{x}_i$为第i个$cluster$的特征,则需假设异方差模型

$$y_i=\tilde{\mathbf{x}}_i^{\mathsf{T}}\boldsymbol{\beta}+\varepsilon_i, E(\varepsilon_i)=0,\quad \mathrm{var}(\varepsilon_i)=\sigma^2/m_i,\quad w_i=1/m_i$$

注：大多数cluster汇总统计量的方差与 $1/m_i$成正比，但也有例外，如果汇总是最大值或最小值，则其方差可能与$1/\sqrt{\log(m_i)}$成正比。

例2 (简单线性回归的 WLS). 假设$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \varepsilon_i \sim (0, \sigma^2 / w_i)$, w_i已知. *WLS*极小化

$$\sum_{i=1}^{n} w_i (y_i - \beta_0 - \beta_1 x_i)^2$$

$$\hat{\beta}_1 = \frac{\sum w_i (x_i - \bar{x}_w) y_i}{\sum w_i (x_i - \bar{x}_w)^2}, \quad \bar{x}_w = \frac{\sum w_i x_i}{\sum w_i}; \quad \hat{\beta}_0 = \bar{y}_w - \hat{\beta}_1 \bar{x}_w, \quad \bar{y}_w = \frac{\sum w_i y_i}{\sum w_i},$$

$$\mathrm{var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum w_i (x_i - \bar{x}_w)^2}, \quad \mathrm{var}(\hat{\beta}_0) = \frac{\sigma^2}{\sum w_i} + \frac{\bar{x}_w{}^2 \sigma^2}{\sum w_i (x_i - \bar{x}_w)^2}$$

注意：如果将加权最小 二乘目标函数写成普通 最小二乘的形式(变换y, x):

$$\sum_{i=1}^{n} w_i (y_i - \beta_0 - \beta_1 x_i)^2 = \sum_{i=1}^{n} (\sqrt{w_i} y_i - \sqrt{w_i} \beta_0 - \beta_1 \sqrt{w_i} x_i)^2$$

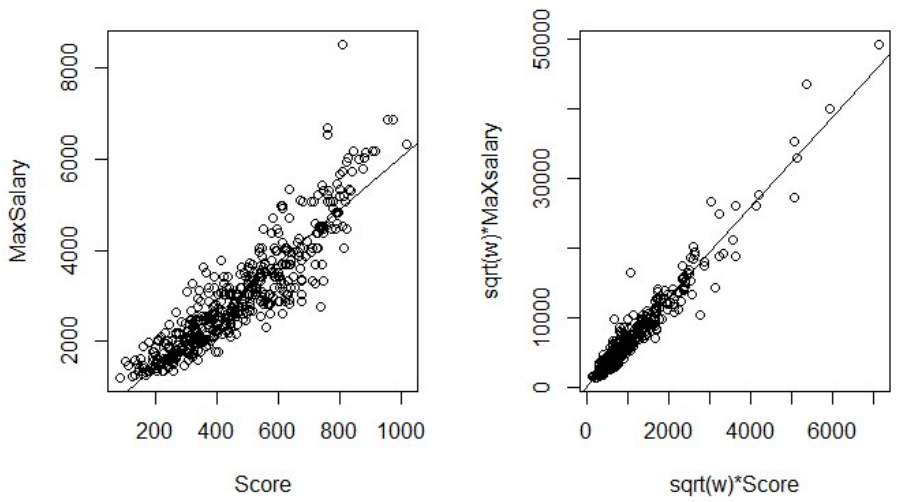
记$\tilde{y}_i = \sqrt{w_i} y_i, \tilde{x}_i = \sqrt{w_i} x_i, \tilde{z}_i = \sqrt{w_i}$,上述右端平方和为$\sum_{i=1}^{n} (\tilde{y}_i - \tilde{z}_i \beta_0 - \tilde{x}_i \beta_1)^2$

对应的方差齐性模型$\tilde{y}_i = \tilde{z}_i \beta_0 + \tilde{x}_i \beta_1 + \tilde{\varepsilon}_i$不含截距项。

例3. 数据集 salarygov (alr4) 汇总了美国政府某部门495种职位的信息，把包括每种职位的最高工资、每种职位的人数、女性人数和难度系数(Score)。目的是研究工资与职位难度的关系。变量具体描述如下：

变量	描述
MaxSalary	职位最高工资
Score	职位难度系数（82-1017）
NE	该职位的雇员总数
NW	该职位的女性人数

JobClass	NW	NE	Score	MaxSalary
Account_clerk	52	68	258	1549
Account_clerk_Interm	26	29	269	1712
Account_clerk_Principal	10	13	321	2182
Account_clerk_Senior	16	24	273	1982
Accountant	1	12	352	2555
...				



WLS：lm函数指定weights：$w_i = \sqrt{\log(NE)}$

$\text{lm}(y \sim x, \text{weights} = (w_1, w_2, \dots, w_n))$

$x_1,\dots,x_n \; iid \sim N(0,1), \quad x_{(n)} = \max(x_i)$

$$\text{var}(x_{(n)}) \approx \frac{c}{\sqrt{\log(n)}}, \quad E(x_{(n)}) \approx \sqrt{2\log(n)}$$

迭代加权最小二乘方法(IRLS）

复杂的误差方差结构$G=G(\boldsymbol{\theta})$下，其中$\boldsymbol{\theta}$是未知参数,通用的方法是极小化

$$Q(\boldsymbol{\theta},\boldsymbol{\beta})=(\mathbf{y}-X\boldsymbol{\beta})^{\mathsf{T}}G^{-1}(\theta)(\mathbf{y}-X\boldsymbol{\beta})+\log|G(\theta)|,$$

可以认为是极小化加权误差平和的时候施加惩罚$\log|G(\theta)|$，实际上该目标函数是正态模型下的对数似然函数。

虽然极小化Q函数可以同时求解最优 的$\boldsymbol{\beta}$和$\boldsymbol{\theta}$，但更简单的方法还是分开求解$\boldsymbol{\beta}$和$\boldsymbol{\theta}$，即

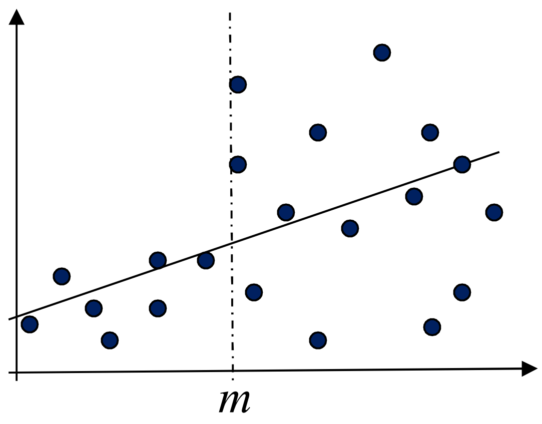
- 给定当前的$\boldsymbol{\theta}$解，利用GLS求解$\boldsymbol{\beta}$；
- 给定$\boldsymbol{\beta}$，利用残差求解$\boldsymbol{\theta}$

上述两步反复迭代直至 收敛。

例4. 右图所示的误差在第 m个样本点处转变,

$$\mathbf{y}=X\boldsymbol{\beta}+\boldsymbol{\varepsilon},\boldsymbol{\varepsilon}\sim(0,G),\ \ G=\mathrm{diag}(\sigma^2,\ldots,\sigma^2,\tau^2,\ldots\ldots,\tau^2),$$

假设m已知， $\boldsymbol{\theta}=(\sigma^2,\tau^2)$。



$$Q(\boldsymbol{\theta},\boldsymbol{\beta}) = \log|G(\boldsymbol{\theta})| + (\mathbf{y}-X\boldsymbol{\beta})^{\mathsf{T}} G^{-1}(\boldsymbol{\theta})(\mathbf{y}-X\boldsymbol{\beta})$$

$$= m\log(\sigma^2) + (n-m)\log(\tau^2) + \sum_{i=1}^{m}\frac{(y_i - \mathbf{x}_i{}^{\mathsf{T}}\boldsymbol{\beta})^2}{\sigma^2} + \sum_{i=m+1}^{n}\frac{(y_i - \mathbf{x}_i{}^{\mathsf{T}}\boldsymbol{\beta})^2}{\tau^2}$$

虽然直接同时极小化 $\boldsymbol{\theta},\boldsymbol{\beta}$很容易，但我们下面分 步对它们求解：

IRLS解法：

Step0. $k=0$, 初始化$\hat{\boldsymbol{\beta}}^{(k)} = \hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$,

Step1. $k=k+1$, 计算残差$\mathbf{e} = \mathbf{y} - X\hat{\boldsymbol{\beta}}^{(k-1)}$,并估计$\sigma^2,\tau^2$：

$$\hat{\sigma}^2_{(k)} = \sum_{i=1}^{m} e_i^2/m, \quad \hat{\tau}^2_{(k)} = \sum_{i=m+1}^{n} e_i^2/(n-m), \quad G_k = diag(\hat{\sigma}^2_{(k)},\ldots,\hat{\sigma}^2_{(k)},\hat{\tau}^2_{(k)}\ldots,\hat{\tau}^2_{(k)})$$

Step2. 计算GLS估计$\hat{\boldsymbol{\beta}}^{(k)} = (X^{\mathsf{T}}G_k{}^{-1}X)^{-1}X^{\mathsf{T}}G_k{}^{-1}\mathbf{y}$,

goto Step1,重复至收敛。

例5．配对数据中每一对中的两个个体接受不同的处理，其响应y_{1j}，y_{2j}不独立，通常假设他们共享一个共同的随机变量β_j，假设混合效应模型（mixed-effects model）:

$$\begin{cases} y_{1j} = \mu + \beta_j + \delta_{1j} \\ y_{2j} = \mu + \beta_j + \alpha + \delta_{2j} \end{cases}, \quad \delta_{ij} \ iid \sim N(0, \sigma^2),\ j = 1, \ldots, J,$$

其中$\beta_1, \ldots, \beta_J \ iid \sim N(0, \tau^2)$称为随机效应(random effects),与$\delta's$独立，$\delta_{1j}$和$\delta_{2j}$分别是第$j$对两个个体的个体效应（随机误差），$\alpha, \mu$是参数（称为固定效应fixed effects），固定效应α代表了两个个体的差异。记每一对的方差协方差矩阵

$$\Psi = \operatorname{var}\begin{pmatrix} y_{1j} \\ y_{2j} \end{pmatrix} = \begin{pmatrix} \tau^2 + \sigma^2 & \tau^2 \\ \tau^2 & \tau^2 + \sigma^2 \end{pmatrix}.$$

记所有响应$\mathbf{y} = (y_{11}, y_{21},\ y_{12}, y_{22},\ \ldots)^{\mathrm{T}}$，$\boldsymbol{\beta} = (\mu, \alpha) \Rightarrow$ 异方差模型

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon},\quad \boldsymbol{\varepsilon} \sim N(0, G),\quad G = \begin{pmatrix} \Psi & 0 & \\ 0 & \Psi & \\ & & \ddots \end{pmatrix} = I_n \otimes \Psi \ (\text{Kronecker乘积}).$$

IRLS: (1)给定τ^2,σ^2时,GLS求解$\boldsymbol{\beta}$;计算残差$e_{1j},e_{1j}, j=1,2,\ldots$

(2)给定$\boldsymbol{\beta}$, 求解τ^2,σ^2有些繁琐，可构造简单（不必最好）的估计:

$$\tau^2 = S_{12} = \sum (e_{1j}-\bar{e}_1)(e_{2j}-\bar{e}_2)/n,$$

$$\sigma^2 = (S_{11}+S_{22})/2 - S_{12}，\ S_{11} = \sum (e_{1j}-\bar{e}_1)^2/n$$

$$\mathrm{var}\begin{pmatrix}\varepsilon_{1j}\\ \varepsilon_{2j}\end{pmatrix} = \begin{pmatrix}\tau^2+\sigma^2 & \tau^2\\ \tau^2 & \tau^2+\sigma^2\end{pmatrix}$$

纵向数据(longitudinal data)：每个个体i有多次测量$(y_{ik},\mathbf{x}_{ik}), k=1,\ldots,n_i$,
（每个个体类似于方差分析中的区组，前述成对数据可看作是最简单的纵向数据），假设混合效应线性模型(既有随机效应，也有固定效应）

$$y_{ik} = \alpha_i + \boldsymbol{\beta}^\top \mathbf{x}_{ik} + \varepsilon_{ik}, \varepsilon_{ik}\ \mathrm{iid} \sim (0,\sigma^2), k=1,\ldots,n_i$$

其中α_i其个体i的效应(~区组效应），代表个个体之间的差异，同一个个体i的所有观测都与α_i有关。若α_i为参数，称为固定效应；若假设α_i为随机变量，比如通常假设$\alpha_i \sim N(\alpha,\tau^2)$,称为随机效应。$\boldsymbol{\beta}$通常是固定的。

很多问题的优化目标函数可写成加权LS的形式：

$$\min_{\boldsymbol{\beta},\boldsymbol{\theta}} \sum w_i(\boldsymbol{\beta},\boldsymbol{\theta})\left(y_i - \mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}\right)^2,\ \text{权与}\boldsymbol{\beta},\boldsymbol{\theta}\text{有关}$$

假设分别优化$\boldsymbol{\beta}$和$\boldsymbol{\theta}$比较容易，可以转化为IRLS：

$$\boldsymbol{\beta}^{(k)} = \arg\min_{\boldsymbol{\beta}} \sum w_i(\boldsymbol{\beta}^{(k-1)},\boldsymbol{\theta}^{(k-1)})\left(y_i - \mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}\right)^2$$

- 给定$\boldsymbol{\theta}=\boldsymbol{\theta}^{(k-1)}$，$\boldsymbol{\beta}=\boldsymbol{\beta}^{(k-1)}$，计算$w_i(\boldsymbol{\beta}^{(k-1)},\boldsymbol{\theta}^{(k-1)})$，$WLS$方法更新$\boldsymbol{\beta}=\boldsymbol{\beta}^{(k)}$；
- 使用更新的$\boldsymbol{\beta}=\boldsymbol{\beta}^{(k)}$，使用残差更新$w_i$中或$G$中的未知参数$\boldsymbol{\theta}$的估计。

线性模型的M估计方法（稳健回归）：模型：$\mathbf{y} = X\boldsymbol{\beta}+\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}\sim(0,\sigma^2 I_n)$,

$$\min \sum \rho(y_i - \mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}),\quad \rho(\cdot)\ge 0\text{关于0对称，}\ \rho(0)=0.$$

当$\rho(t)=|t|$时称为最小一乘法或LAD (least absolute deviation).

对于LAD方法, 改写目标函数：

$$\sum |y_i - \mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}| = \sum w_i(\boldsymbol{\beta})|y_i - \mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}|^2,\ \text{其中}w_i(\boldsymbol{\beta}) = |y_i - \mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}|^{-1}.$$

$$IRLS:\quad w_i(\boldsymbol{\beta}) = |y_i - \mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}|^{-1},\quad \boldsymbol{\beta}^{(k+1)} = \arg\min_{\beta} \sum w_i(\boldsymbol{\beta}^{(k)})|y_i - \mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}|^2.$$
