1. 考虑重复测量(repeat measurements) 的线性模型

$$Y_{i,j} = X_i^T\beta + e_{i,j}\ e_{i,j} \sim iid\ N(0, ||X_i||^2\sigma^2),\ 1 \leqslant j \leqslant n_i,\ 1 \leqslant i \leqslant n$$

其中$Y_{i,j}$ 是响应值，X_i 是$p\times 1$的预测变量，X_i 的第一个元素是1，β是$p\times 1$的未知参数向量，$e_{i,j}$ 是随机误差，σ^2 是未知的方差，$||a||^2 = \sum_{i=1}^p a_i^2,\ a = (a_1, \ldots, a_p)^T$.

1) 试求出β 的最小二乘估计$\hat{\beta}_{LS}$，进而构造σ^2 的无偏估计；

$$\frac{\partial}{\partial\beta}\sum_{i=1}^n\sum_{j=1}^{n_i}(Y_{i,j} - X_i^T\beta)^2 = 2\beta\sum_{i=1}^n n_iX_iX_i^T - 2\sum_{i=1}^n X_i\sum_{j=1}^{n_i}Y_{i,j} \tag{1}$$

$$\frac{\partial^2}{\partial\beta^2}\sum_{i=1}^n\sum_{j=1}^{n_i}(Y_{i,j} - X_i^T\beta)^2 = 2\sum_{i=1}^n n_iX_iX^T \tag{2}$$

Setting (1) to 0, we obtain

$$\sum_{i=1}^n X_i\sum_{j=1}^{n_i}Y_{i,j} = \hat{\beta}\sum_{i=1}^n n_iX_iX_i^T$$

Now we let

$$\begin{aligned}
&X = (\underbrace{X_1^T, \ldots, X_1^T}_{n_1}, \underbrace{X_2^T, \ldots, X_2^T}_{n_2}, \ldots, \underbrace{X_n^T, \ldots, X_n^T}_{n_n})^T\\
&Y = (Y_{1,1}, \ldots, Y_{1,n_1}, Y_{2,1}, \ldots, Y_{2,n_2}, \ldots, Y_{n,1}, \ldots, Y_{n,n_n})^T\\
&e = (e_{1,1}, \ldots, e_{1,n_1}, e_{2,1}, \ldots, e_{2,n_2}, \ldots, e_{n,1}, \ldots, e_{n,n_n})^T\\
&Y = X\beta + e \qquad H = X(X^TX)^{-1}X^T
\end{aligned}$$

Suppose that X is an $\sum_{i=1}^n n_i\ \times p$ constant matrix with rank p, then we know that $X^TX = \sum_{i=1}^n n_iX_iX_i^T$ is invertible, thus

$$\hat{\beta} = (\sum_{i=1}^n n_iX_iX_i^T)^{-1}\sum_{i=1}^n X_i\sum_{j=1}^{n_i}Y_{i,j} = (X^TX)^{-1}X^TY$$

From the fact that (2) > 0, we know that $\hat{\beta}$ is the minimum point, that is,

$$\hat{\beta}_{LS} = \hat{\beta} = (\sum_{i=1}^n n_iX_iX_i^T)^{-1}\sum_{i=1}^n X_i\sum_{j=1}^{n_i}Y_{i,j} = (X^TX)^{-1}X^TY$$

Let

$$\begin{aligned}
&\hat{Y} = X\hat{\beta} = HY\\
&\hat{e} = Y - \hat{Y} = Y - HY
\end{aligned}$$

It's easy to obtain that H is symmetric and idempotent, and $HX = X$. We also know that $I - H$ is symmetric and idempotent. Now we let D denote the following $\sum_{i=1}^{n} n_i \times \sum_{i=1}^{n} n_i$ matrix

$$D = \begin{bmatrix} \frac{1}{||X_1||} I_{n_1} & & \\ & \ddots & \\ & & \frac{1}{||X_n||} I_{n_n} \end{bmatrix}$$

Then we have

$$\begin{aligned}
E||D\hat{e}||^2 &= E(\hat{e}^T D^T D\hat{e}) = E\Big[Y^T (I-H)^T D^T D(I-H)Y\Big] \\
&= E\Big[(e^T + \beta^T X^T)(I-H)^T D^T D(I-H)(X\beta + e)\Big] \\
&= E\Big[e^T (I-H)^T D^T D(I-H)e\Big] = E\Big[tr(e^T (I-H)^T D^T D(I-H)e)\Big] \\
&= E\Big[tr((I-H)^T D^T D(I-H)ee^T)\Big] = tr\Big[(I-H)^T D^T D(I-H)Eee^T\Big] \\
&= \sigma^2 tr D(I-H) \begin{pmatrix} ||X_1||^2 I_{n_1} & & \\ & \ddots & \\ & & ||X_n||^2 I_{n_n} \end{pmatrix} (I-H)^T D^T \\
&= \sigma^2 tr\Big[D(I-H)D^{-1}D^{-T}(I-H)^T D^T\Big] \\
&= \sigma^2 tr\Big[D(I-H)D^{-1}D(I-H)D^{-1}\Big] = tr\Big[D(I-H)D^{-1})\Big] \\
&= tr(I-H) = \sum_{i=1}^{n} n_i - p
\end{aligned}$$

Thus, the following $\hat{\sigma}^2$ is an unbiased estimate of σ^2

$$\hat{\sigma}^2 = \frac{1}{\sum_{i=1}^{n} n_i - p} ||D\hat{e}||^2$$

2) 试求出β 和σ^2 的最大似然估计$\hat{\beta}_{MLE}$ 和$\hat{\sigma}^2_{MLE}$；

We know that $Y_{i,j},\ 1 \leqslant j \leqslant n_i,\ 1 \leqslant i \leqslant n$ are independent, and

$$Y_{i,j} \sim N(X_i^T \beta, ||X_i||^2 \sigma^2),\ 1 \leqslant j \leqslant n_i,\ 1 \leqslant i \leqslant n$$

By the definition of the likelihood function, we write

$$L(\beta, \sigma^2 | data) = \prod_{i=1}^{n} \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi ||X_i||^2 \sigma^2}} \exp\{-\frac{1}{2||X_i||^2 \sigma^2} (Y_{i,j} - X_i^T \beta)^2\}$$

and the log likelihood function

$$l(\beta, \sigma^2) = \sum_{i=1}^{n} (-\frac{n_i}{2}) log(2\pi||X_i||^2\sigma^2) - \sum_{i=1}^{n} \frac{1}{2||X_i||^2\sigma^2} \sum_{j=1}^{n_i} (Y_{i,j} - X_i^T\beta)^2$$

We first fix σ^2 and set

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^{n} \frac{X_i}{||X_i||^2\sigma^2} \sum_{j=1}^{n_i} Y_{i,j} - \beta \sum_{i=1}^{n} \frac{n_i X_i X_i^T}{||X_i||^2\sigma^2} = 0$$

we obtain

$$\hat{\beta} = \left[\sum_{i=1}^{n} \frac{n_i X_i X_i^T}{||X_i||^2} \right]^{-1} \sum_{i=1}^{n} \frac{X_i}{||X_i||^2} \sum_{j=1}^{n_i} Y_{i,j}$$

Since

$$\frac{\partial^2 l}{\partial \beta^2} = -\sum_{i=1}^{n} \frac{n_i X_i X_i^T}{||X_i||^2\sigma^2} < 0$$

Hence, when σ^2 is fixed, $\hat{\beta}$ is the global maximum point of $l(\beta, \sigma^2)$.
Now we fix β and set

$$\frac{\partial l}{\partial \sigma^2} = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} n_i + \frac{1}{\sigma^4} \sum_{i=1}^{n} \frac{1}{2||X_i||^2} \sum_{j=1}^{n_i} (Y_{i,j} - X_i^T\beta)^2 = 0$$

we obtain

$$\hat{\sigma}^2 = \frac{1}{\sum_{i=1}^{n} n_i} \sum_{i=1}^{n} \frac{1}{||X_i||^2} \sum_{j=1}^{n_i} (Y_{i,j} - X_i^T\hat{\beta})^2$$

We also find that

$$\frac{\partial l}{\partial \sigma^2} > 0 \quad if\ \sigma^2 < \hat{\sigma}^2$$
$$\frac{\partial l}{\partial \sigma^2} < 0 \quad if\ \sigma^2 > \hat{\sigma}^2$$

that is, $\hat{\sigma}^2$ is the global maximum point of $l(\hat{\beta}, \sigma^2)$. Therefore

$$\hat{\beta}_{MLE} = \left[\sum_{i=1}^{n} \frac{n_i X_i X_i^T}{||X_i||^2} \right]^{-1} \sum_{i=1}^{n} \frac{X_i}{||X_i||^2} \sum_{j=1}^{n_i} Y_{i,j}$$
$$\hat{\sigma}^2_{MLE} = \frac{1}{\sum_{i=1}^{n} n_i} \sum_{i=1}^{n} \frac{1}{||X_i||^2} \sum_{j=1}^{n_i} (Y_{i,j} - X_i^T\hat{\beta}_{MLE})^2$$

3) β 的最小二乘估计和最大似然估计哪一个更好一些？并给出理由；

Since $E(\hat{\beta}_{LS}) = E(\hat{\beta}_{MLE}) = \beta$, we know they are both unbiased. Next we consider their variance.

Let $D^2 = G$, then we can write $\hat{\beta}_{MLE} = (X^T G X)^{-1} X^T G Y$ We know that

$$\hat{\beta}_{LS} = (X^T X)^{-1} X^T Y = \beta + (X^T X)^{-1} X^T e$$
$$\hat{\beta}_{MLE} = (X^T G X)^{-1} X^T G Y = \beta + (X^T G X)^{-1} X^T G e$$

Thus,

$$\begin{aligned}
Var(\hat{\beta}_{LS}) &= E(\hat{\beta} - \beta)^T (\hat{\beta} - \beta) \\
&= E\Big[(X^T X)^{-1} X^T e\Big]^T \Big[(X^T X)^{-1} X^T e\Big] \\
&= E\Big[e^T X (X^T X)^{-1} (X^T X)^{-1} X^T e\Big] \\
&= tr\Big[X (X^T X)^{-1} (X^T X)^{-1} X^T E e e^T\Big] \\
&= \sigma^2 tr\Big[(X^T X)^{-2} X^T G^{-1} X\Big]
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
Var(\hat{\beta}_{MLE}) &= E(\hat{\beta} - \beta)^T (\hat{\beta} - \beta) \\
&= E\Big[e^T G^T X (X^T G^T X)^{-1} (X^T G X)^{-1} X^T G e\Big] \\
&= tr\Big[G^T X (X^T G X)^{-2} X^T G E e e^T\Big] \\
&= \sigma^2 tr\Big[(X^T G X)^{-2} X^T G X\Big] \\
&= \sigma^2 tr((X^T G X)^{-1})
\end{aligned}$$

Now we are left to compare $tr\Big[(X^T X)^{-2} X^T G^{-1} X\Big]$ and $tr((X^T G X)^{-1})$, where

$$G = D^2 = \begin{bmatrix} \frac{1}{||X_1||^2} I_{n_1} & & \\ & \ddots & \\ & & \frac{1}{||X_n||^2} I_{n_n} \end{bmatrix}$$

By Cauchy-Schwartz inequality, we know that

$$(\sum_{i=1}^{N} X_i X_i^T)^2 \leqslant (\sum_{i=1}^{N} X_i ||X_i||^2 X_i^T)(\sum_{i=1}^{N} \frac{X_i X_i^T}{||X_i||^2})$$

which means

$$(X^T X)^2 \leqslant (X^T G^{-1} X)(X^T G X)$$

Thus,

$$(X^T G X)^{-1} \leqslant (X^T X)^{-2} X^T G^{-1} X$$

That is, $Var(\hat{\beta}_{MLE}) \leqslant Var(\hat{\beta}_{LS})$, which also means $MSE(\hat{\beta}_{MLE}) \leqslant MSE(\hat{\beta}_{LS})$. Therefore, $\hat{\beta}_{MLE}$

is a better estimate of β.

4) $\hat{\sigma}^2_{MLE}$ 是否是σ^2 的无偏估计？如否，将其修正为无偏估计；

From 2) we obtain that

$$\begin{aligned}\hat{\sigma}^2_{MLE} &= \frac{1}{\sum_{i=1}^n n_i}\sum_{i=1}^n \frac{1}{||X_i||^2}\sum_{j=1}^{n_i}(Y_{i,j} - X_i^T\hat{\beta}_{MLE})^2 \\ &= \frac{1}{\sum_{i=1}^n n_i}||D(Y - X\hat{\beta}_{MLE})||^2\end{aligned}$$

We also have $\hat{\beta}_{MLE} = \beta + (X^TGX)^{-1}X^TGe$, $E(ee^T) = \sigma^2G^{-1}$, thus

$$\begin{aligned}E(\hat{\sigma}^2_{MLE}) &= \frac{1}{\sum_{i=1}^n n_i}E\Big[(Y - X\hat{\beta}_{MLE})^TD^TD(Y - X\hat{\beta}_{MLE})\Big] \\ &= \frac{1}{\sum_{i=1}^n n_i}E\Big[e^T[I - X(X^TGX)^{-1}X^TG]^TG[I - X(X^TGX)^{-1}X^TG]e\Big] \\ &= \frac{1}{\sum_{i=1}^n n_i}tr\Big[[I - X(X^TGX)^{-1}X^TG]^TG[I - X(X^TGX)^{-1}X^TG]\sigma^2G\Big] \\ &= \frac{\sigma^2}{\sum_{i=1}^n n_i}tr\Big[(G - G^TX(X^TG^TX)^{-1}X^TG)G^{-1}\Big] \\ &= \frac{\sigma^2}{\sum_{i=1}^n n_i}\Big[tr(I) - tr[G^TX(X^TG^TX)^{-1}X^T]\Big] \\ &= \frac{\sigma^2}{\sum_{i=1}^n n_i}\Big[tr(I_{\sum_{i=1}^n n_i\times\sum_{i=1}^n n_i}) - tr[(X^TG^TX)^{-1}X^TG^TX]\Big] \\ &= \frac{\sigma^2}{\sum_{i=1}^n n_i}[\sum_{i=1}^n n_i - p]\end{aligned}$$

Thus, the following $\tilde{\sigma}^2$ is an unbiased estimate of σ^2

$$\tilde{\sigma}^2 = \frac{\sum_{i=1}^n n_i}{\sum_{i=1}^n n_i - p}\hat{\sigma}^2_{MLE}$$

5) 如何检验如下假设，并说明理由。

$$H_0 : \beta_2 = \cdots = \beta_p = 0,$$

From 2) we know that

$$\begin{aligned}L(\beta, \sigma^2|X, Y) &= \prod_{i=1}^n\Big(\frac{1}{\sqrt{2\pi||X_i||^2\sigma^2}}\Big)^{n_i}\exp\{\sum_{i=1}^n\sum_{j=1}^{n_i} - \frac{(Y_{i,j} - X_i^T\beta)^2}{2||X_i||^2\sigma^2}\} \\ l(\beta, \sigma^2) &= \sum_{i=1}^n\Big(-\frac{n_i}{2}\Big)log(2\pi||X_i||^2\sigma^2) - \sum_{i=1}^n\frac{1}{2||X_i||^2\sigma^2}\sum_{j=1}^{n_i}(Y_{i,j} - X_i^T\beta)^2\end{aligned}$$

By the definition of the likelihood ratio test statistic, we have

$$\lambda(X,Y) = \frac{\sup\limits_{\theta\in\Theta_0} L(\theta|X,Y)}{\sup\limits_{\theta\in\Theta} L(\theta|X,Y)},\ \theta = (\beta, \sigma^2)$$

When $\theta \in \Theta$, we already know that

$$\arg\sup_{\beta,\sigma^2} L(\beta,\sigma^2|X,Y) = (\hat{\beta}_{MLE}, \hat{\sigma}^2_{MLE})$$

When $\theta \in \Theta_0$ *i.e.* $\beta_2 = \cdots = \beta_p = 0$, we can write the log likelihood function as follows

$$l(\theta\in\Theta_0) = l(\beta_1,\sigma^2) = \sum_{i=1}^{n}(-\frac{n_i}{2})log(2\pi||X_i||^2\sigma^2) - \sum_{i=1}^{n}\frac{1}{2||X_i||^2\sigma^2}\sum_{j=1}^{n_i}(Y_{i,j}-\beta_1)^2$$

We first fix σ^2 and set

$$\frac{\partial l}{\partial\beta_1} = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{Y_{i,j}-\beta_1}{||X_i||^2\sigma^2} = 0$$

we obtain

$$\tilde{\beta}_1 = (\sum_{i=1}^{n}\frac{n_i}{||X_i||^2})^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{Y_{i,j}}{||X_i||^2}$$

Since

$$\frac{\partial^2 l}{\partial{\beta_1}^2} = -\sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{1}{||X_i||^2\sigma^2} < 0$$

Hence, when σ^2 is fixed, $\tilde{\beta}_1$ is the global maximum point of $l(\beta_1,\sigma^2)$.
Now we fix β_1 and set

$$\frac{\partial l}{\partial\sigma^2} = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}n_i + \frac{1}{\sigma^4}\sum_{i=1}^{n}\frac{1}{2||X_i||^2}\sum_{j=1}^{n_i}(Y_{i,j}-\beta_1)^2 = 0$$

we obtain

$$\tilde{\sigma}^2 = \frac{1}{\sum_{i=1}^{n}n_i}\sum_{i=1}^{n}\frac{1}{||X_i||^2}\sum_{j=1}^{n_i}(Y_{i,j}-\tilde{\beta}_1)^2$$

We also find that

$$\frac{\partial l}{\partial\sigma^2} > 0 \quad if\ \sigma^2 < \tilde{\sigma}^2$$
$$\frac{\partial l}{\partial\sigma^2} < 0 \quad if\ \sigma^2 > \tilde{\sigma}^2$$

that is, $\tilde{\sigma}^2$ is the global maximum point of $l(\tilde{\beta}_1,\sigma^2)$. Therefore

$$\arg\sup_{\Theta_0} L(\beta_1,\sigma^2) = (\tilde{\beta}_1,\tilde{\sigma}^2)$$

and

$$\lambda(X,Y) = \frac{L(\tilde{\beta}_1,\tilde{\sigma^2})}{L(\hat{\beta}_{MLE},\hat{\sigma}^2_{MLE})} = (\frac{\hat{\sigma}_{MLE}}{\tilde{\sigma}})^{\sum_{i=1}^{n}n_i}$$

For some $0 < c < 1$, We reject H_0 when

$$\lambda(X, Y) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n_i} (Y_{i,j} - X_i^T \hat{\beta}_{MLE})^2 / ||X_i||^2}{\sum_{i=1}^{n} \sum_{j=1}^{n_i} (Y_{i,j} - \tilde{\beta}_1)^2 / ||X_i||^2} \leqslant c$$