第五讲 线性回归模型

2023.10.13

$$y = a + \mathbf{b}^\top \mathbf{x} + \varepsilon,\ \ \varepsilon \perp\!\!\!\perp \mathbf{x},\ \ \varepsilon \sim (0, \sigma^2)$$

线性回归模型

为了研究x, y之间是否存在因果关系，所有影响y的其它因素需要与x独立（即不存在干扰因素）。所以，即使控制变量，偏相关分析和去相关化也不足以推断因果关系，这是因为不相关不等同于独立。

线性回归模型

定义1: (一元)线性回归模型假设响应变量y和自变量向量$\mathbf{x}$满足方程:

$$y = a + \mathbf{b}^{\mathrm{T}}\mathbf{x} + \varepsilon,$$

其中误差项ε是不可观测的随机变量，满足Gauss-Markov(GM)假设

(1) $E(\varepsilon) = 0$； (截距a可识别)

(2) $\mathrm{var}(\varepsilon) = \sigma^2$ (方差齐性, Homoscedasticity)

(3) ε与$\mathbf{x}$独立 (外生性, Exogeneity)

- 误差项ε代表了除了自变量$\mathbf{x}$之外所有其它影响y的因素。
- 模型假设$\varepsilon \perp\!\!\!\perp \mathbf{x}$ (ε与$\mathbf{x}$独立)，即ε不是干扰因素，是推断因果必需的。
- 条件(1)、(2)简记作$\varepsilon \sim (0, \sigma^2)$
- $(y, \mathbf{x})$服从多元正态时，GM假设都成立。

命题1. 若$(y, \mathbf{x})$满足线性模型的Gauss-Markov假设(1),(2),(3)，则

（i）$E(y \mid \mathbf{x}) = a + \mathbf{b}^\top \mathbf{x}$ （回归函数线性）

（ii）$\mathrm{var}(y \mid \mathbf{x}) = \sigma^2$ （条件方差常数）

证明：因为ε与$\mathbf{x}$独立，所以$E(\varepsilon \mid \mathbf{x}) = E(\varepsilon) = 0$，所以

$$E(y \mid \mathbf{x}) = E(a + \mathbf{b}^\top \mathbf{x} + \varepsilon \mid \mathbf{x}) = a + \mathbf{b}^\top \mathbf{x} + E(\varepsilon \mid \mathbf{x}) = a + \mathbf{b}^\top \mathbf{x},$$

$$\mathrm{var}(y \mid \mathbf{x}) = \mathrm{var}(a + \mathbf{b}^\top \mathbf{x} + \varepsilon \mid \mathbf{x}) = \mathrm{var}(\varepsilon \mid \mathbf{x}) = \mathrm{var}(\varepsilon) = \sigma^2.$$

注：线性回归模型蕴含了条件期望

$$E(y|\mathbf{x}) = a + \mathbf{b}^\top \mathbf{x}$$

关于$\mathbf{x}$线性。条件期望$E(y|\mathbf{x})$也称为回归函数，是平方损失下所有$\mathbf{x}$的函数中对y的最佳逼近（参见P6）。所以线性回归模型实际上假设了“$\mathbf{x}$的线性函数是所有y的函数$g(\mathbf{x})$逼近中最佳的逼近”。

命题1中的(i),(ii) 与定义1中的(1)、(2)、(3)几乎等价，后者更强：

若(i),(ii)成立，令$\varepsilon = y - E(y\,|\,\mathbf{x}) = y-(a+\mathbf{b}^\top\mathbf{x})$，则我们有

$$y = a+\mathbf{b}^\top\mathbf{x}+\varepsilon,$$

下面验证ε满足GM假设(1),(2),以及(3)的弱版本：

(1)$E(\varepsilon)=E(y-E(y\,|\,\mathbf{x}))=0,$

(2) $\mathrm{var}(\varepsilon)=\mathrm{var}(y-a-\mathbf{b}^\top\mathbf{x})$

$=\mathrm{var}(E(y-a-\mathbf{b}^\top\mathbf{x}\,|\,\mathbf{x}))+E(\mathrm{var}(y-a-\mathbf{b}^\top\mathbf{x}\,|\,\mathbf{x}))$

$=0+E(\mathrm{var}(y\,|\,\mathbf{x}))=\sigma^2.$

(3*)由条件期望的性质，$\varepsilon = y - E(y\,|\,\mathbf{x})$与$\mathbf{x}$不相关（但不一定独立）。

我们只需在$(i),(ii)$基础上增加假设"$y - E(y\,|\,\mathbf{x})$与$\mathbf{x}$独立"，即得到等价定义

模型的等价定义

定义2：假设y是响应变量，$\mathbf{x}$是自变量(向量)，$E(y|\mathbf{x})$称为回归函数，线性回归模型假设：

(i) 线性回归函数：$E(y|\mathbf{x}) = a + \mathbf{b}^\top \mathbf{x}$.

(ii) 方差常数/齐性：$\mathrm{var}(y|\mathbf{x}) = \sigma^2$.

(iii) $y - E(y|\mathbf{x})$与$\mathbf{x}$独立。

注1：一般教材通常假设$\mathbf{x}$是给定的常数，并以条件(i)、(ii)或GM假设(1)、(2)定义线性模型，而不假设ε与$\mathbf{x}$的独立性。

注2：其它（非线性）回归模型通常对回归函数做假设，比如

(1) logistic回归：响应变量y是伯努利变量，它与自变量$\mathbf{x}$有关，logistic回归模型假设$y|\mathbf{x} \sim B(1, \pi(\mathbf{x}))$,

$$\pi(\mathbf{x}) = P(y=1|\mathbf{x}) = 1 - P(y=0|\mathbf{x}) = E(y|\mathbf{x}) = \frac{\exp(a + \mathbf{b}^\top \mathbf{x})}{1 + \exp(a + \mathbf{b}^\top \mathbf{x})}$$

(2)Poisson回归：假设y是计数变量，服从Poisson分布$Pois(\lambda(\mathbf{x}))$,

$$\lambda(\mathbf{x}) = E(y|\mathbf{x}) = \exp(a + \mathbf{b}^\top \mathbf{x}).$$

向量空间中的投影

向量空间或线性空间V是一个定义了加法和数乘的封闭集合，集合的元素称作向量。假设赋予V内积：$\langle v,u\rangle$，$u,v\in V$。假设$U\subset V$是向量子空间，则任何$v\in V$在U上的正交投影（记作$\hat{v}\in U$）满足

$$\langle v-\hat{v},u\rangle=0,\forall\ u\in U$$

最常见的向量空间是函数空间，包括R^n和随机变量空间，以及量子力学的量子空间。

关于回归函数

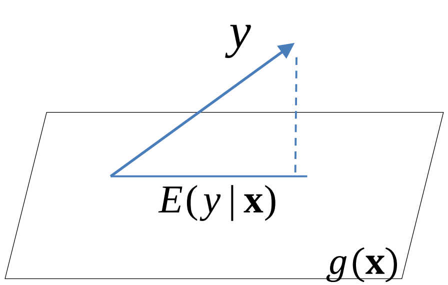
定义(条件期望/回归函数)：$\varphi(\mathbf{x})$是y给定$\mathbf{x}$的条件期望, 记作$E(y|\mathbf{x})$, 若对任何可测$g(\mathbf{x})$, 内积$\langle y-\varphi(\mathbf{x}),g(\mathbf{x})\rangle=E(y-\varphi(\mathbf{x}))g(\mathbf{x})=0$.

定义表明，$y-\varphi(\mathbf{x})\perp L^2(\sigma(\mathbf{x}))$，其中

$$L^2(\sigma(\mathbf{x}))=\{g(\mathbf{x}):g\text{可测},Eg(\mathbf{x})^2<\infty\}$$

是所有$\sigma(\mathbf{x})$可测函数组成的随机变量空间。因此，$E(y|\mathbf{x})$是y在$\mathbf{x}$的函数组成的空间上的正交投影，从而是$\mathbf{x}$的所有可测函数中对y的最佳逼近:

$$E(y|\mathbf{x})=\operatorname{argmin}_g E(y-g(\mathbf{x}))^2\qquad(*)$$



条件期望$E(y|\mathbf{x})$是y在所有$\mathbf{x}$的函数组成空间上的正交投影。

证明(*):这是正交投影的基本性质，但这里不妨直接证明如下：

首先，注意到$\min_{a\in R} E(y-a)^2$的最小值在$a=E(y)$处达到。

其次，为了极小化$\min_g E(y-g(\mathbf{x}))^2$，只需极小化$E[(y-g(\mathbf{x}))^2\mid\mathbf{x}]$,

当$\mathbf{x}$给定时，常数$g(\mathbf{x})$等于y的期望$E(y\mid\mathbf{x})$时达到极小。

性质：假设y是任一随机变量,$\mathbf{x}$是随机向量,则

(1) Law of total expectation/Tower property: $E(E(y\mid\mathbf{x}))=E(y)$。

(2) Law of total variance（方差分解/勾股定理）:

$$\operatorname{var}(y)=\operatorname{var}[E(y\mid\mathbf{x})]+E[\operatorname{var}(y\mid\mathbf{x})]$$

因此定义决定系数

$$\Phi=\frac{\operatorname{var}(E(y\mid\mathbf{x}))}{\operatorname{var}(y)}$$

证明：（1）条件期望定义中取$g(\mathbf{x})=1$,则$E[y-E(y\mid x)]\times 1=0$.

(2):令$\varepsilon=y-E(y\mid\mathbf{x})$,定义中取$g(\mathbf{x})=x_i$, $E[y-E(y\mid x)]x_i=0$

即ε与x_i不相关，ε与$\mathbf{x}$不相关，正交分解$y=E(y\mid\mathbf{x})+\varepsilon$两边同时求方差：

$$\operatorname{var}(y)=\operatorname{var}[E(y\mid\mathbf{x})]+\operatorname{var}[\varepsilon],$$

而误差的方差$\operatorname{var}[\varepsilon]=\operatorname{var}[y-E(y\mid\mathbf{x})]=\ldots=E[\operatorname{var}(y\mid\mathbf{x})]$。

参数的含义

命题2. 假设线性回归模型(y随机变量，$\mathbf{x}_{p\times 1}$自变量):

$$y=a+\mathbf{b}^{\mathsf{T}}\mathbf{x}+\varepsilon,\ E(\varepsilon)=0,\ \mathrm{var}(\varepsilon)=\sigma^2,\ \varepsilon\text{与}\mathbf{x}\text{独立}$$

则参数a，$\mathbf{b}$，σ^2由$y,\mathbf{x}$的均值和方差决定：

$$\mathbf{b}=\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y},\quad a=\mu_y-\mathbf{b}^{\mathsf{T}}\boldsymbol{\mu}_{\mathbf{x}},\quad \sigma^2=\Sigma_{yy\bullet\mathbf{x}}=\Sigma_{yy}(1-\Phi)$$

其中$\Phi=\dfrac{\mathrm{var}(\mathbf{b}^{\mathsf{T}}\mathbf{x})}{\mathrm{var}(y)}=\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y}/\Sigma_{yy}$为决定系数。

证：ε与$\mathbf{x}$独立$\Rightarrow\varepsilon=y-a-\mathbf{b}^{\mathsf{T}}\mathbf{x}$与$\mathbf{x}$不相关，是$\mathbf{y}$的去相关化，所以必定$\mathbf{b}^{\mathsf{T}}=\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}$，再次验证如下：

$$0=\mathrm{cov}(\varepsilon,\mathbf{x})=\mathrm{cov}(y-a-\mathbf{b}^{\mathsf{T}}\mathbf{x},\mathbf{x})=\Sigma_{y\mathbf{x}}-\mathbf{b}^{\mathsf{T}}\Sigma_{\mathbf{xx}}\quad\Rightarrow\mathbf{b}=\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y}$$

方程两边求均值：$\mu_{\mathbf{y}}=a+\mathbf{b}^{\mathsf{T}}\boldsymbol{\mu}_{\mathbf{x}}\Rightarrow a=\mu_y-\mathbf{b}^{\mathsf{T}}\boldsymbol{\mu}_{\mathbf{x}}=\mu\ -\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\boldsymbol{\mu}_{\mathbf{x}}$

方程两边求方差：$\Sigma_{yy}=\mathrm{var}(\mathbf{b}^{\mathsf{T}}\mathbf{x})+\mathrm{var}(\varepsilon)=\mathbf{b}^{\mathsf{T}}\Sigma_{\mathbf{xx}}\mathbf{b}+\sigma^2\Rightarrow\sigma^2=\Sigma_{yy\bullet\mathbf{x}}$

$\Phi=\mathrm{var}(\mathbf{b}^{\mathsf{T}}\mathbf{x})/\mathrm{var}(y)=\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y}/\Sigma_{yy}\Rightarrow\sigma^2=\Sigma_{yy\bullet\mathbf{x}}=\Sigma_{yy}(1-\Phi)$。

注1：由命题2，线性回归模型以均值和方差可表示为：

$$y = a + \Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\mathbf{x} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon}\text{与}\ \mathbf{x}\text{独立}$$

与去相关化类比几乎相同：

$$y = \Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\mathbf{x} + y^{\perp}, \qquad y^{\perp}\text{与}\ \mathbf{x}\text{不相关}$$

其中$y^{\perp}$可看作是误差项，它与$\mathbf{x}$不相关，但不一定独立

- 因为要求$E(\boldsymbol{\varepsilon}) = 0$，回归模型多出了常数项$\mathbf{a}$（不影响相关性和独立性）；
- 回归模型假设$\boldsymbol{\varepsilon} = \mathbf{y} - \Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\mathbf{x} - \mathbf{a}$与$\mathbf{x}$独立，比"$\mathbf{y}^{\perp}$与$\mathbf{x}$不相关"要求更高。
- 两个变量不相关可以通过去相关化做到，而独立性只是个假设，难以做到(无法做到)。

注2：响应变量$\mathbf{y}$是向量的情形，线性模型类似：

定义：（多元线性回归模型）

假设随机向量$\mathbf{y}_{q\times1},\mathbf{x}_{p\times1}$满足：

$$\mathbf{y}_{q\times1}=\mathbf{a}_{q\times1}+B_{q\times p}\mathbf{x}_{p\times1}+\boldsymbol{\varepsilon}_{q\times1},\quad \boldsymbol{\varepsilon}\sim(0,\ \Sigma),\quad \underline{\underline{\boldsymbol{\varepsilon}\text{与}\mathbf{x}\text{独立}}}$$

称为多元线性回归模型。其中$\mathbf{y}$称为响应，$\mathbf{x}$称为自变量，可观测；$\boldsymbol{\varepsilon}$不可观测,一般称为误差。$\mathbf{a}$，B，Σ是参数，它们完全由$\mathbf{y},\mathbf{x}$的均值和方差决定：$B=\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}$，$\mathbf{a}=\boldsymbol{\mu}_{\mathbf{y}}-B\boldsymbol{\mu}_{\mathbf{x}}$，$\Sigma=\Sigma_{\mathbf{yy\bullet x}}$。

今后一般不再考虑多元回归模型。

考虑$y=a+\mathbf{b}^\top\mathbf{x}+\varepsilon$中部分变量的回归系数。假设$\mathbf{x}=\begin{pmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{pmatrix}$, $\mathbf{b}=\begin{pmatrix}\mathbf{b}_1\\ \mathbf{b}_2\end{pmatrix}$,

其中$\mathbf{x}_1$是感兴趣的变量，$\mathbf{x}_2$是控制变量/干扰因素。

命题3. 假设模型：

$$y=a+\mathbf{b}^\top\mathbf{x}+\varepsilon=a+\mathbf{b}_1^\top\mathbf{x}_1+\mathbf{b}_2^\top\mathbf{x}_2+\varepsilon，\ \varepsilon\sim(0,\sigma^2)\text{与}\mathbf{x}\text{独立}$$

其中$\mathbf{x}=\begin{pmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{pmatrix}$独立，$\mathbf{b}=\begin{pmatrix}\mathbf{b}_1\\ \mathbf{b}_2\end{pmatrix}$。令$\mathbf{x}_1^\perp=\mathbf{x}_1-\Sigma_{12}\Sigma_{22}^{-1}\mathbf{x}_2$，$y^\perp=y-\Sigma_{y2}\Sigma_{22}^{-1}\mathbf{x}_2$，则

(1) $y^\perp=a+\mathbf{b}_1^\top\mathbf{x}_1^\perp+\varepsilon$，$E(\varepsilon)=0$, $\mathrm{var}(\varepsilon)=\sigma^2$, ε与$\mathbf{x}_1^\perp$独立.

(2) $\mathbf{b}_1=\Sigma_{11\bullet 2}^{-1}\Sigma_{1y\bullet 2}$

(3) 当$\mathbf{x}_1$是随机变量时，$b_1\propto\rho_{1y\bullet 2}$与偏相关系数成正比。

(4) 部分/偏决定系数$R_{1y\bullet 2}^2=\dfrac{\mathrm{var}(\mathbf{b}_1^\top\mathbf{x}_1^\perp)}{\mathrm{var}(y^\perp)}=\dfrac{\Sigma_{y1\bullet 2}\Sigma_{11\bullet 2}^{-1}\Sigma_{1y\bullet 2}}{\Sigma_{yy\bullet 2}}$.

证明：（1）原方程为

$$y^{\perp} + \Sigma_{y2}\Sigma_{22}^{-1}\mathbf{x}_2 = y = a + \mathbf{b}_1^{\mathsf{T}}(\mathbf{x}_1^{\perp} + \Sigma_{12}\Sigma_{22}^{-1}\mathbf{x}_2) + \mathbf{b}_2^{\mathsf{T}}\mathbf{x}_2 + \varepsilon$$

所以

$$y^{\perp} = a + \mathbf{b}_1^{\mathsf{T}}\mathbf{x}_1^{\perp} + \underline{\left(\mathbf{b}_1^{\mathsf{T}}\Sigma_{12}\Sigma_{22}^{-1} + \mathbf{b}_2^{\mathsf{T}} - \Sigma_{y2}\Sigma_{22}^{-1}\right)\mathbf{x}_2} + \varepsilon \overset{\text{记为}}{=} a + \mathbf{b}_1^{\mathsf{T}}\mathbf{x}_1^{\perp} + \underline{\mathbf{c}^{\mathsf{T}}\mathbf{x}_2} + \varepsilon$$

其中下划线$\mathbf{c}^{\mathsf{T}}\mathbf{x}_2$部分一定为0。

这是因为方程中的其它各项都与$\mathbf{x}_2$不相关,

$$0 = \mathrm{cov}(y^{\perp} - a - \mathbf{b}_1^{\mathsf{T}}\mathbf{x}_1^{\perp} - \varepsilon, \mathbf{x}_2) = \mathrm{cov}(\mathbf{c}^{\mathsf{T}}\mathbf{x}_2, \mathbf{x}_2) = \mathbf{c}^{\mathsf{T}}\Sigma_{22}$$

$\Rightarrow \mathbf{c}^{\mathsf{T}} = 0$，因此我们有

$$y^{\perp} = a + \mathbf{b}_1^{\mathsf{T}}\mathbf{x}_1^{\perp} + \varepsilon$$

（2）对于模型(1)

$$y^{\perp}=a+\mathbf{b}_1^{\top}\mathbf{x}_1^{\perp}+\varepsilon，\ E(\varepsilon)=0,\ \mathrm{var}(\varepsilon)=\sigma^2,\ \varepsilon\text{与}\mathbf{x}_1^{\perp}\text{独立}$$

应用命题2，我们有

$$\mathbf{b}_1=\Sigma_{\mathbf{x}_1^{\perp}\mathbf{x}_1^{\perp}}^{-1}\Sigma_{\mathbf{x}_1^{\perp}y^{\perp}}=\Sigma_{11\bullet 2}^{-1}\Sigma_{1y\bullet 2}$$

(2) 也可利用分块矩阵的逆矩阵公式证明如下：

由命题2，我们知$y=a+\mathbf{b}^{\top}\mathbf{x}+\varepsilon$中$\mathbf{b}=\begin{pmatrix}\mathbf{b}_1\\ \mathbf{b}_2\end{pmatrix}=\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y}$,

其中$\Sigma_{\mathbf{xx}}=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix},\Sigma_{\mathbf{x}y}=\begin{pmatrix}\Sigma_{1y}\\ \Sigma_{2y}\end{pmatrix}$，所以

$$\begin{pmatrix}\mathbf{b}_1\\ \mathbf{b}_2\end{pmatrix}=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}^{-1}\begin{pmatrix}\Sigma_{1y}\\ \Sigma_{2y}\end{pmatrix}$$

$$=\begin{pmatrix}\Sigma_{11\bullet 2}^{-1} & -\Sigma_{11\bullet 2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\\ * & *\end{pmatrix}\begin{pmatrix}\Sigma_{1y}\\ \Sigma_{2y}\end{pmatrix}=\begin{pmatrix}\Sigma_{11\bullet 2}^{-1}\Sigma_{1y}-\Sigma_{11\bullet 2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{2y}\\ *\end{pmatrix}$$

故$\mathbf{b}_1=\Sigma_{11\bullet 2}^{-1}\Sigma_{1y}-\Sigma_{11\bullet 2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{2y}=\Sigma_{11\bullet 2}^{-1}(\Sigma_{1y}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{2y})=\Sigma_{11\bullet 2}^{-1}\Sigma_{1y\bullet 2}$

（3）特别地，当$\mathbf{x}_1$是随机变量时， $b_1 = \Sigma_{y1\bullet 2} / \Sigma_{11\bullet 2}$

$$= \frac{\Sigma_{y1\bullet 2}}{\sqrt{\Sigma_{11\bullet 2}}\sqrt{\Sigma_{yy\bullet 2}}} \times \frac{\sqrt{\Sigma_{yy\bullet 2}}}{\sqrt{\Sigma_{11\bullet 2}}} \propto \rho_{1y\bullet 2}$$

(4)部分 / 偏决定系数$R^2_{y1\bullet 2}$即模型(1)的决定系数

$$R^2_{y1\bullet 2} = \frac{\mathrm{var}(\mathbf{b}_1^\top \mathbf{x}_1^\perp)}{\mathrm{var}(y^\perp)} = \frac{\mathbf{b}_1^\top \mathrm{var}(\mathbf{x}_1^\perp)\mathbf{b}_1}{\mathrm{var}(y^\perp)} = \frac{\mathbf{b}_1^\top \Sigma_{11\bullet 2}\mathbf{b}_1}{\Sigma_{yy\bullet 2}} = \frac{\Sigma_{y1\bullet 2}\Sigma_{11\bullet 2}^{-1}\Sigma_{1y\bullet 2}}{\Sigma_{yy\bullet 2}}$$

参数的矩估计

假设样本$(y_i, \mathbf{x}_i)$，$i=1,\dots,n$来自于线性模型

$$y = a + \mathbf{b}^\top \mathbf{x} + \varepsilon,\quad E(\varepsilon)=0,\ \mathrm{var}(\varepsilon)=\sigma^2,\ \varepsilon\text{与}\mathbf{x}\text{独立}$$

假设样本方差矩阵和样本均值为

$$S = \begin{pmatrix} S_{yy} & S_{y\mathbf{x}} \\ S_{\mathbf{x}y} & S_{\mathbf{xx}} \end{pmatrix},\quad \begin{pmatrix} \bar{y} \\ \bar{\mathbf{x}} \end{pmatrix}$$

由命题2知,参数a，$\mathbf{b}$，σ^2的矩估计可取为:

$$\begin{cases} \hat{\mathbf{b}} = S_{\mathbf{xx}}^{-1} S_{\mathbf{x}y}, \\ \hat{a} = \bar{y} - \hat{\mathbf{b}}^\top \bar{\mathbf{x}}, \\ \tilde{\sigma}^2 = S_{yy\bullet\mathbf{x}} = S_{yy} - S_{y\mathbf{x}} S_{\mathbf{xx}}^{-1} S_{\mathbf{x}y}. \end{cases}$$

另外，$R^2 = \hat{\Phi} = S_{y\mathbf{x}} S_{\mathbf{xx}}^{-1} S_{\mathbf{x}y} / S_{yy}$。

注：后面将会看到，除了误差方差之外， 所有参数的最小二乘估计等于上述矩估计。误差方差的最小二乘估计

$$\hat{\sigma}^2 = \frac{n-1}{n-p}\tilde{\sigma}^2 \approx \tilde{\sigma}^2$$