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

第五讲 线性回归模型

2022.9.30

线性回归模型形式上与去相关化相同，但假设误差与自变量独立（而不是不相关）。

Recap

$\mathbf{y}_{q\times1}$与 $\mathbf{x}_{p\times1}$整体相关性 （典则相关系数或决定系数）

去相关化：$\mathbf{y}^{\perp} = \mathbf{y} - \Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\mathbf{x} = \mathbf{y} - B\mathbf{x}$

1) 正交分解： $\mathbf{y} = \Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\mathbf{x} + \mathbf{y}^{\perp} = B\mathbf{x} + \mathbf{y}^{\perp}$， $\mathbf{y}^{\perp}$与$\mathbf{x}$不相关
（线性模型有更强的假设： $\mathbf{y} = \mathbf{a} + B\mathbf{x} + \boldsymbol{\varepsilon}$， $\boldsymbol{\varepsilon}$与$\mathbf{x}$独立）

2) 方差分解：$\Sigma_{\mathbf{yy}} = \Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}} + \Sigma_{\mathbf{yy\bullet x}}$

$\mathbf{x}$解释$\mathbf{y}$的方差的比例（ $q \times q$ 矩阵）：

$$\Phi = \Phi_{\mathbf{yx}} = \Sigma_{\mathbf{yy}}^{-1/2}\left(\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}\right)\Sigma_{\mathbf{yy}}^{-1/2}$$

3) $\mathbf{x}$ 和$\mathbf{y}$的整体相关性：

$q = 1$: 决定系数 $R^2 = \Phi$.

$q > 1$: 典则相关系数 $\sqrt{\lambda_{\max}(\Phi)}$

$\mathbf{y}_{q\times 1}$与 $\mathbf{x}_{p\times 1} = \begin{pmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{pmatrix}$的局部相关性：控制$\mathbf{x}_2$时，$\mathbf{y}$与$\mathbf{x}_1$的相关性

- $\mathbf{y}$与$\mathbf{x}_1$关于$\mathbf{x}_2$去相关化：

$$\mathbf{y}^{\perp} = \mathbf{y} - \Sigma_{\mathbf{y}2}\Sigma_{22}^{-1}\mathbf{x}_2,\quad \mathbf{x}_1^{\perp} = \mathbf{x}_1 - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{x}_2$$

- 考虑 $\mathbf{y}^{\perp}$，$\mathbf{x}_1^{\perp}$的相关性（应用上页 1)-3)）。

$\Sigma_{\mathbf{y}2} = \Sigma_{\mathbf{y}\mathbf{x}_2}$
$\Sigma_{22} = \Sigma_{\mathbf{x}_2\mathbf{x}_2}$

具体如下：

1) 去相关化（$\mathbf{y}^{\perp}$中消去$\mathbf{x}_1^{\perp}$）：$(\mathbf{y}^{\perp})^{\perp} = \mathbf{y}^{\perp} - \Sigma_{\mathbf{y}^{\perp}\mathbf{x}_1^{\perp}}\Sigma_{\mathbf{x}_1^{\perp}\mathbf{x}_1^{\perp}}^{-1}\mathbf{x}_1^{\perp}$

   方差分解：$\Sigma_{\mathbf{yy}\bullet 2} = \Sigma_{\mathbf{y}1\bullet 2}\Sigma_{11\bullet 2}^{-1}\Sigma_{1\mathbf{y}\bullet 2} + \Sigma_{\mathbf{yy}\bullet\mathbf{x}}$

2) $\mathbf{x}_1^{\perp}$解释$\mathbf{y}^{\perp}$的方差的“比例”

$$\Phi^{\perp} = \Phi_{\mathbf{y}1\bullet 2} = \Sigma_{\mathbf{yy}\bullet 2}^{-1/2}\left(\Sigma_{\mathbf{y}1\bullet 2}\Sigma_{11\bullet 2}^{-1}\Sigma_{1\mathbf{y}\bullet 2}\right)\Sigma_{\mathbf{yy}\bullet 2}^{-1/2}$$

3) $\mathbf{y}$与$\mathbf{x}_1$的局部相关性：

   $q = 1$: $R_{y1\bullet 2}^2 \triangleq \Phi^{\perp}$ 部分决定系数

   $q > 1$: $\sqrt{\lambda_{max}(\Phi^{\perp})}$ 部分典则相关系数

具体计算步骤如下（仅考虑$q = 1$情形）

(1) 计算随机变量y和随机向量$\mathbf{x}$之间的决定系数

- $\begin{pmatrix} y \\ \mathbf{x} \end{pmatrix}$的协方差矩阵：$\Sigma = \begin{pmatrix} \Sigma_{yy} & \Sigma_{y\mathbf{x}} \\ \Sigma_{\mathbf{x}y} & \Sigma_{\mathbf{xx}} \end{pmatrix}$
- $R^2 = \Sigma_{y\mathbf{x}} \Sigma_{\mathbf{xx}}^{-1} \Sigma_{\mathbf{x}y} / \Sigma_{yy}$

(2) 控制$\mathbf{z}$时，计算随机变量y和随机向量$\mathbf{x}$之间的部分决定系数$R^2_{y\mathbf{x}\bullet\mathbf{z}}$

- $\begin{pmatrix} \boxed{\begin{matrix} y \\ \mathbf{x} \end{matrix}} \\ \mathbf{z} \end{pmatrix} \triangleq \begin{pmatrix} \boxed{\mathbf{w}} \\ \mathbf{z} \end{pmatrix}$的协方差矩阵: $\Sigma = \begin{pmatrix} \boxed{\begin{matrix} \Sigma_{yy} & \Sigma_{y\mathbf{x}} \\ \Sigma_{\mathbf{x}y} & \Sigma_{\mathbf{xx}} \end{matrix}} & \begin{matrix} \Sigma_{y\mathbf{z}} \\ \Sigma_{\mathbf{xz}} \end{matrix} \\ \begin{matrix} \Sigma_{\mathbf{z}y} & \Sigma_{\mathbf{zx}} \end{matrix} & \Sigma_{\mathbf{zz}} \end{pmatrix} = \begin{pmatrix} \boxed{\Sigma_{\mathbf{ww}}} & \Sigma_{\mathbf{wz}} \\ \Sigma_{\mathbf{zw}} & \Sigma_{\mathbf{zz}} \end{pmatrix}$
- 计算$\mathbf{w} = \begin{pmatrix} y \\ \mathbf{x} \end{pmatrix}$的关于$\mathbf{z}$的偏协方差矩阵$\Sigma_{\mathbf{ww}\bullet\mathbf{z}} = \Sigma_{\mathbf{ww}} - \Sigma_{\mathbf{wz}} \Sigma_{\mathbf{zz}}^{-1} \Sigma_{\mathbf{zw}}$:

$$\begin{pmatrix} \Sigma_{yy\bullet\mathbf{z}} & \Sigma_{y\mathbf{x}\bullet\mathbf{z}} \\ \Sigma_{\mathbf{x}y\bullet\mathbf{z}} & \Sigma_{\mathbf{xx}\bullet\mathbf{z}} \end{pmatrix} = \begin{pmatrix} \Sigma_{yy} & \Sigma_{y\mathbf{x}} \\ \Sigma_{\mathbf{x}y} & \Sigma_{\mathbf{xx}} \end{pmatrix} - \begin{pmatrix} \Sigma_{y\mathbf{z}} \\ \Sigma_{\mathbf{xz}} \end{pmatrix} \Sigma_{\mathbf{zz}}^{-1} (\Sigma_{\mathbf{z}y}, \Sigma_{\mathbf{zx}})$$

- $R^2_{y\mathbf{x}\bullet\mathbf{z}} = \Sigma_{y\mathbf{x}\bullet\mathbf{z}} \Sigma_{\mathbf{xx}\bullet\mathbf{z}}^{-1} \Sigma_{\mathbf{x}y\bullet\mathbf{z}} / \Sigma_{yy\bullet\mathbf{z}}$

例1. 三个变量 (y, x, z) 的相关系数矩阵：

$$\Sigma = \begin{pmatrix} 1 & \rho_{yx} & \rho_{yz} \\ \rho_{xy} & 1 & \rho_{xz} \\ \rho_{zy} & \rho_{zx} & 1 \end{pmatrix}$$

（1）y与(x, z)的决定系数 :

$$R^2 = \left(\rho_{yx}, \rho_{yz}\right) \begin{pmatrix} 1 & \rho_{xz} \\ \rho_{zx} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho_{xy} \\ \rho_{zy} \end{pmatrix} / \Sigma_{yy} = \frac{\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz}}{1 - \rho_{xz}^2}$$

（2）控制z条件下，y与x的偏决定系数:

首先$\begin{pmatrix} \Sigma_{yy\bullet z} & \Sigma_{yx\bullet z} \\ \Sigma_{xy\bullet z} & \Sigma_{xx\bullet z} \end{pmatrix} = \begin{pmatrix} 1 & \rho_{yx} \\ \rho_{xy} & 1 \end{pmatrix} - \begin{pmatrix} \rho_{yz} \\ \rho_{xz} \end{pmatrix} \left(\rho_{zy}, \rho_{zx}\right) = \begin{pmatrix} 1 - \rho_{yz}^2 & \rho_{yx} - \rho_{yz}\rho_{xz} \\ \rho_{yx} - \rho_{yz}\rho_{xz} & 1 - \rho_{xz}^2 \end{pmatrix}$

$$\Rightarrow R_{yx\bullet z}^2 = \Sigma_{yx\bullet z}\Sigma_{xx\bullet z}^{-1}\Sigma_{xy\bullet z} / \Sigma_{yy\bullet z} = (\rho_{yx} - \rho_{yz}\rho_{xz})^2 / (1 - \rho_{yz}^2)(1 - \rho_{xz}^2)$$

$= (r_{yx\bullet z})^2 =$ 偏相关系数的平方。

数值例子：

第二讲例3(续）中，阅读能力y，身高x,年龄z的相关系数矩阵为

$$\Sigma = \begin{array}{c} \\ y \\ x \\ z \end{array}\begin{array}{c} \begin{array}{ccc} y & x & z \end{array} \\ \begin{pmatrix} 1 & 0.56 & 0.7 \\ 0.56 & 1 & 0.8 \\ 0.7 & 0.8 & 1 \end{pmatrix} \end{array}$$

$$R^2 = (0.56, 0.7)\begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}^{-1}\begin{pmatrix} 0.56 \\ 0.57 \end{pmatrix} = 0.49$$

$$\begin{pmatrix} \Sigma_{yy\bullet z} & \Sigma_{yx\bullet z} \\ \Sigma_{xy\bullet z} & \Sigma_{xx\bullet z} \end{pmatrix} = \begin{pmatrix} 1 & 0.56 \\ 0.56 & 1 \end{pmatrix} - \begin{pmatrix} 0.7 \\ 0.8 \end{pmatrix}(0.7, 0.8) = \begin{array}{c} \\ y \\ x \end{array}\begin{array}{c} \begin{array}{cc} y & x \end{array} \\ \begin{pmatrix} 0.51 & 0 \\ 0 & 0.36 \end{pmatrix} \end{array}$$

$$\Rightarrow R^2_{yx\bullet z} = \Sigma_{yx\bullet z}\Sigma^{-1}_{xx\bullet z}\Sigma_{xy\bullet z} / \Sigma_{yy\bullet z} = 0 = (r_{yx\bullet z})^2$$

线性回归模型

即使控制变量，相关分析也不足以推断因果关系，这是因为不相关（去相关化）不等同于独立。

线性回归模型有更强的假设，即误差与自变量独立。该条件成立时，参数才能正确地估计（无偏），有好的理论性质。

多元线性回归模型

定义：假设$\mathbf{y}_{q\times1},\mathbf{x}_{p\times1}$满足：

$$\mathbf{y}_{q\times1}=\mathbf{a}_{q\times1}+B_{q\times p}\mathbf{x}_{p\times1}+\boldsymbol{\varepsilon}_{q\times1},\quad E(\boldsymbol{\varepsilon})=0,\ \operatorname{var}(\boldsymbol{\varepsilon})=\Omega,\quad \underline{\underline{\boldsymbol{\varepsilon}\text{不可观测,与}\mathbf{x}\text{独立}}}$$

称为多元线性回归模型。

注：对比去相关化的正交分解：

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

线性模型对误差项有更高的要求：$\mathbf{y}^{\perp}$与$\mathbf{x}$独立 。

但独立性条件在实际操作中无法执行，数据分析方法（比如最小二乘）实质上只利用了误差与自变量的不相关性。因此，从实际操作的角度看，线性模型与去相关化本质相同。

线性模型与去相关化

命题1. 多元线性回归模型的参数$\mathbf{a}$，B，Ω由$\mathbf{y},\mathbf{x}$的均值和方差决定：

$$B=\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1},\quad \mathbf{a}=\boldsymbol{\mu}_{\mathbf{y}}-B\boldsymbol{\mu}_{\mathbf{x}},\quad \Omega=\Sigma_{\mathbf{yy}\bullet\mathbf{x}}$$

注：由命题，线性回归模型可写成去相关化的形式：

$$\mathbf{y}-\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\mathbf{x}=\boldsymbol{\varepsilon}+\boldsymbol{\mu}_{\mathbf{y}}-\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\boldsymbol{\mu}_{\mathbf{x}}$$

差别在于（1）模型要求$\mathbf{y}-\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\mathbf{x}$与$\mathbf{x}$独立，（2）$E(\boldsymbol{\varepsilon})=0$导致多出一个截距项。

证明：前面求解相关化 时除了均值，已经基本 验证，再次证明如下：

- $\boldsymbol{\varepsilon}$与$\mathbf{x}$独立

$$\Rightarrow 0=\mathrm{cov}(\boldsymbol{\varepsilon},\mathbf{x})=\mathrm{cov}(\mathbf{y}-\mathbf{a}-B\mathbf{x},\mathbf{x})=\Sigma_{\mathbf{yx}}-B\Sigma_{\mathbf{xx}}\quad\Rightarrow B=\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}$$

- 方程两边求均值

$$\Rightarrow \boldsymbol{\mu}_{\mathbf{y}}=\mathbf{a}+B\boldsymbol{\mu}_{\mathbf{x}}\Rightarrow \mathbf{a}=\boldsymbol{\mu}_{\mathbf{y}}-B\boldsymbol{\mu}_{\mathbf{x}}=\boldsymbol{\mu}_{\mathbf{y}}-\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\boldsymbol{\mu}_{\mathbf{x}}$$

- 方程两边求方差

$$\Rightarrow \Sigma_{\mathbf{yy}}=\mathrm{var}(B\mathbf{x})+\mathrm{var}(\boldsymbol{\varepsilon})=B\Sigma_{\mathbf{xx}}B^{\top}+\Omega=\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}+\Omega$$

$$\Rightarrow \Omega=\Sigma_{\mathbf{yy}}-\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}}$$

(一元)线性回归模型

定义：(一元)线性回归模型假设一元响应变量y和自变量向量$\mathbf{x}$满足方程：

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

其中误差项ε是不可观测的随机变量，满足Gauss-Markov假设

(1) $E(\varepsilon) = 0$

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

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

注1：Gauss-Markov假设(1),(2)简单记为$\varepsilon \sim (0, \sigma^2)$。

注2：一般教科书上认为$\mathbf{x}$是给定的，通常不提条件(3)。

命题2(命题1的$q=1$情况). 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：

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

其中$R^2 = \dfrac{\mathrm{var}(\mathbf{b}^{\mathsf{T}}\mathbf{x})}{\mathrm{var}(y)} = \Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y} / \Sigma_{yy}$。

参数的矩估计

所有参数可由矩估计方法得到。假设样本$(y_i, \mathbf{x}_i)$来自于线性模型

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

假设样本方差矩阵和样本均值为

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

则参数a，$\mathbf{b}$，σ^2的矩估计

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

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

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

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

考虑$y=a+\mathbf{b}^{\mathsf{T}}\mathbf{x}+\varepsilon$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：

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

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

（2）$\mathbf{b}_1=\Sigma_{11\bullet2}^{-1}\Sigma_{1y\bullet2}$

(3) 当$\mathbf{x}_1$是随机变量时，$b_1\propto\rho_{1y\bullet2}$与偏相关系数成正比。

（4）部分 / 偏决定系数$R_{1y\bullet2}^2=\dfrac{\mathrm{var}(\mathbf{b}_1^{\mathsf{T}}\mathbf{x}_1^{\perp})}{\mathrm{var}(y^{\perp})}=\dfrac{\Sigma_{y1\bullet2}\Sigma_{11\bullet2}^{-1}\Sigma_{1y\bullet2}}{\Sigma_{yy\bullet2}}$.

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,

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

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

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

（2）对于模型(1)

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

应用命题2，我们有

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

（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)部分决定系数$R_{y1\bullet 2}^2$即模型(1)的决定系数

$$R_{y1\bullet 2}^2=\frac{\mathrm{var}(\mathbf{b}_1^{\mathsf{T}}\mathbf{x}_1^{\perp})}{\mathrm{var}(y^{\perp})}=\frac{\mathbf{b}_1^{\mathsf{T}}\,\mathrm{var}(\mathbf{x}_1^{\perp})\mathbf{b}_1}{\mathrm{var}(y^{\perp})}=\frac{\mathbf{b}_1^{\mathsf{T}}\Sigma_{11\bullet 2}\mathbf{b}_1}{\Sigma_{yy\bullet 2}}=\frac{\Sigma_{y1\bullet 2}\Sigma_{11\bullet 2}^{-1}\Sigma_{1y\bullet 2}}{\Sigma_{yy\bullet 2}}$$

(2) $\mathbf{b}_1=\Sigma_{11\bullet 2}^{-1}\Sigma_{1y\bullet 2}$的直接证明(利用分块矩阵的逆矩阵公式化):

由命题2，我们知$y=a+\mathbf{b}^{\mathsf{T}}\mathbf{x}+\varepsilon$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$\mathbf{b}_1=\Sigma_{11\bullet 2}^{-1}\Sigma_{1y}-\Sigma_{11\bullet 2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{2y}=\Sigma_{11\bullet 2}^{-1}(\Sigma_{1y}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{2y})=\Sigma_{11\bullet 2}^{-1}\Sigma_{1y\bullet 2}$

命题4. 若$(y,\mathbf{x})$满足线性模型的Gauss - Markov假设(1),(2),(3)，则

（i）$E(y\mid\mathbf{x})=a+\mathbf{b}^{\mathsf{T}}\mathbf{x}$ （回归函数/条件期望$E(y\mid\mathbf{x})$是线性函数）

（ii）$\mathrm{var}(y\mid\mathbf{x})=\sigma^2$ （条件方差$\mathrm{var}(y\mid\mathbf{x})$是常数，不依赖于$\mathbf{x}$）

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

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

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

注：条件（i）、（ii）可导出（1),(2),但不能得到(3)。

有时线性模型以稍弱于Gauss – Markov条件的（i）、（ii）定义（参见下页）。

另一种模型定义方式

y: 响应变量，$\mathbf{x}$: 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：$E(y\,|\,\mathbf{x}) = a + \mathbf{b}^{\mathsf{T}}\mathbf{x}$，

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

条件(i),(ii)与条件(1),(2),(3)不等价：

$(1),(2),(3) \Rightarrow (i),(ii)$

$(i),(ii) \Rightarrow (1),(2)$, 但不蕴含“$\varepsilon$与$\mathbf{x}$独立”。

验证：令$\varepsilon = y - E(y\,|\,\mathbf{x}) = y - (a + \mathbf{b}^{\mathsf{T}}\mathbf{x})$，则

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

$(2)\, \mathrm{var}(\varepsilon) = \mathrm{var}(y - a - \mathbf{b}^{\mathsf{T}}\mathbf{x}) = \mathrm{var}(E(y - a - \mathbf{b}^{\mathsf{T}}\mathbf{x}\,|\,\mathbf{x})) + E(\mathrm{var}(y - a - \mathbf{b}^{\mathsf{T}}\mathbf{x}\,|\,\mathbf{x}))$

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

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

附：关于回归函数 $E(y|\mathbf{x})$

性质1.假设二阶矩存在，以$\mathbf{x}$的函数$f(\mathbf{x})$逼近y的平方误差定义为

$E(y-f(\mathbf{x}))^2$，则 $f(\mathbf{x})=E(y\mid\mathbf{x})$时误差最小.

性质2：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）令$\varepsilon=y-E(y\mid\mathbf{x})$，则 ε与$\mathbf{x}$不相关且$E(\varepsilon)=0$，

我们有正交分解：$y=E(y\mid\mathbf{x})+\varepsilon$

(3) Law of total variance（正交分解两边求方差）:

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

注：由（3）我们可定义决定系数(y的方差中$\mathbf{x}$能解释的比例):

$$R^2=\operatorname{var}(E(y\mid\mathbf{x}))/\operatorname{var}(y)$$