课程主页：http://staff.ustc.edu.cn/~ynyang/2023

第十一讲 最小二乘法II

2023.12.1

$$\boxed{\begin{gathered} {A^{\perp}}^{\top} B^{\perp} = {A^{\perp}}^{\top} B = A^{\top} B^{\perp} \\ \Sigma(a_i - \bar{a})(b_i - \bar{b}) = \Sigma(a_i - \bar{a})b_i = \Sigma a_i(b_i - \bar{b}) \end{gathered}}$$

例子：体重指数

例1. 第7讲例6（P24）使用简单回归模型研究了$\log(W) \sim \log(H)$的关系，得到了体重指数$W/H^{2.27}$，$W=$体重，H$=$身高。因为W,H都与性别有关，因此我们考虑控制性别因素$Sex(1=male; 0=female)$:

$$\log(W) = a + b\times\log(H) + c\times Sex + \varepsilon,\ \varepsilon \sim N(0,\sigma^2),$$

标准化：$z = \dfrac{\log(W) - (a + b\times\log(H) + c\times Sex)}{\sigma} \sim N(0,1)$

$z \ge z_0 \Leftrightarrow W/H^b > e^{a+z_0\sigma+c\times sex}$ (Lab2，练习3(e))

```
> lm(formula = log(weight) ~ log(height) + sex, data = hw)
Coefficients:
(Intercept)    log(height)      sex
    3.0            2.0               0.124
Residual standard error: 0.1145
```

$$\hat{b}=2, \hat{a}=3, \hat{c}=0.124, \hat{\sigma}=0.1145$$

$z \ge 1.645$（标准正态上5%）

$$\Leftrightarrow BMI = W/H^2 \ge e^{3+0.124\times sex+1.645\times 0.1145} = \begin{cases} 28 & male \\ 24 & female \end{cases}$$

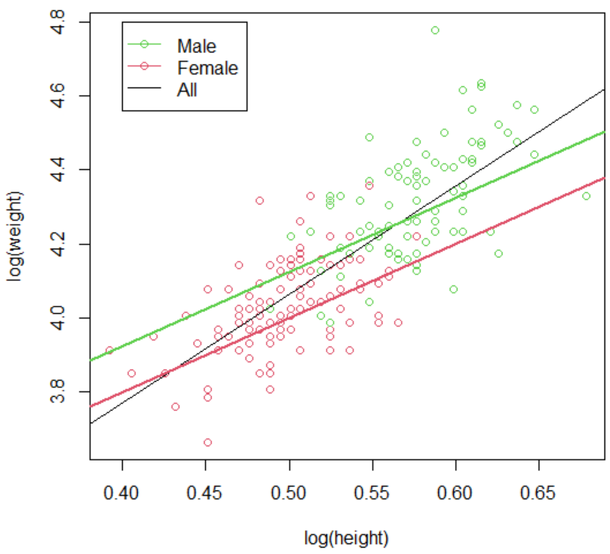
BMI的阈值应该区分性别。

(可加)模型 $\log(W)=a+b\times\log(H)+c\times Sex+\varepsilon$意味着对于男性和女性模型分别为

$$sex=1（绿线）:\ \log(W)=(a+c)+b\times\log(H)$$

$$sex=0（红线）:\ \log(W)=a+b\times\log(H)$$

即$\log(H)$的系数b与性别无关：男性和女性的$\log(H)$的系数都是b,但截距不同。这意味着，男性和女性数据具有大致平行的形状，才能应用线性可加模型。



男性组内（绿）、
女性组内（红）
的斜率b相同，即W-H关系平行，差别体现在截距上（即模型中 $c\times sex$ 一项）。

如果两组数据不平行即男性组和女性组斜率不同，可引入交互作用项：

$$\log(W)=a+b\log(H)+c\mathrm{Sex}+d\mathrm{Sex}\times\log(H)+error$$

此时两组的斜率分别是$b,b+d$

中心化

$\boldsymbol{\beta}$的整体LS估计$\hat{\boldsymbol{\beta}} = (X^\top X)^- X^\top \mathbf{y}$。下面考虑$\boldsymbol{\beta}$分量的LS估计。

$$X = (\mathbf{1}, Z),\quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \mathbf{b} \end{pmatrix}$$

因为截距β_0的特殊性，我们常常需要考虑将$\mathbf{1}$与 Z 独立开来（正交化），令Z关于$\mathbf{1}$的正交化（称为Z的中心化矩阵）：

$$Z_c = Z - P_{\mathbf{1}} Z = Z - \frac{\mathbf{1}\mathbf{1}^\top}{n} Z = Z - \mathbf{1}\bar{\mathbf{x}}^\top$$

其中$\bar{\mathbf{x}} = Z^\top \mathbf{1}/n$是所有自变量的样本平均值组成的向量（$Z$的列平均), 则样本方差矩阵

$$S = Z_c^\top Z_c/(n-1)$$

我们将反复应用投影分解

$$P_X = P_{\mathbf{1}} + P_{Z_c} = \mathbf{1}\mathbf{1}^\top/n + Z_c(Z_c^\top Z_c)^- Z_c^\top$$

一元情形

- 样本：$y_1,\ldots,y_n \in R$；$\mathbf{y}=(y_1,\ldots,y_n)^{\mathsf{T}}$
- 样本均值 $\bar{y}=(y_1+\ldots+y_n)/n=\mathbf{1}^{\mathsf{T}}\mathbf{y}/n$, 其中$\mathbf{1}=(1,\ldots,1)^{\mathsf{T}}$
- 中心化：$\mathbf{y}_c=(y_1-\bar{y},\ldots,y_n-\bar{y})^{\mathsf{T}}=\mathbf{y}-\mathbf{1}\bar{y}=\mathbf{y}-P_{\mathbf{1}}\mathbf{y}$
- 样本方差：$s^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(y_1-\bar{y})^2=\dfrac{1}{n-1}\mathbf{y}_c^{\mathsf{T}}\mathbf{y}_c=\dfrac{1}{n-1}\mathbf{y}^{\mathsf{T}}(I_n-P_{\mathbf{1}})\mathbf{y}.$
- 标准化：$\mathbf{y}_c=(y_1-\bar{y},\ldots,y_n-\bar{y})^{\mathsf{T}}/s=(\mathbf{y}-\mathbf{1}\bar{y})/s$

多元情形

- 样本：$\mathbf{x}_1,\ldots,\mathbf{x}_n \in R^m$；数据矩阵：$Z=(\mathbf{x}_1,\ldots,\mathbf{x}_n)^{\mathsf{T}}$
- 样本均值：$\bar{\mathbf{x}}=(\mathbf{x}_1+\ldots+\mathbf{x}_n)/n=Z^{\mathsf{T}}\mathbf{1}/n.$
- 中心化：$Z_c=(\mathbf{x}_1-\bar{\mathbf{x}},\ldots,\mathbf{x}_n-\bar{\mathbf{x}})^{\mathsf{T}}=Z-\mathbf{1}\bar{\mathbf{x}}^{\mathsf{T}}=Z-P_{\mathbf{1}}Z$
- 样本协方差：$S=\dfrac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^{\mathsf{T}}=\dfrac{1}{n-1}Z_c^{\mathsf{T}}Z_c=\dfrac{1}{n-1}Z^{\mathsf{T}}(I_n-P_{\mathbf{1}})Z$
- 标准化：$Z_s=D^{-1/2}Z_c$, 其中$D=diag(S)$,
- 相关系数矩阵$R=Z_s^{\mathsf{T}}Z_s/(n-1)$

线性模型的中心化

模型：$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{1}\beta_0 + Z\mathbf{b} + \boldsymbol{\varepsilon},\ X_{n\times p} = (\mathbf{1}_{n\times 1}, Z_{n\times(p-1)}),\ \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \mathbf{b} \end{pmatrix}$

- Z中心化：$Z_c = Z - \mathbf{1}\bar{\mathbf{x}}^{\mathsf{T}} = Z - P_{\mathbf{1}}Z$，模型改写为

$$\mathbf{y} = \mathbf{1}\beta_0 + (Z_c + \mathbf{1}\bar{\mathbf{x}}^{\mathsf{T}})\mathbf{b} + \boldsymbol{\varepsilon} = \mathbf{1}(\beta_0 + \bar{\mathbf{x}}^{\mathsf{T}}\mathbf{b}) + Z_c\mathbf{b} + \boldsymbol{\varepsilon} \overset{\Delta}{=} \mathbf{1}\beta_0^* + Z_c\mathbf{b} + \boldsymbol{\varepsilon},$$

其中$\beta_0^* = \beta_0 + \bar{\mathbf{x}}^{\mathsf{T}}\mathbf{b}$。 $P_X = P_{\mathbf{1},Z_c} = P_{\mathbf{1}} + P_{Z_c}$。

- $\mathbf{y}, Z$都中心化：$\mathbf{y}_c = \mathbf{y} - \mathbf{1}\bar{y}, Z_c = Z - \mathbf{1}\bar{\mathbf{x}}^{\mathsf{T}}$，模型改写为

$$\mathbf{y}_c = \mathbf{1}(\beta_0 + \bar{\mathbf{x}}^{\mathsf{T}}\mathbf{b} - \bar{y}) + Z_c\mathbf{b} + \boldsymbol{\varepsilon} \overset{\Delta}{=} \mathbf{1}\beta_0^{**} + Z_c\mathbf{b} + \boldsymbol{\varepsilon}$$

其中$\beta_0^{**} = \beta_0 + \bar{\mathbf{x}}^{\mathsf{T}}\mathbf{b} - \bar{y}$

后面将看到，$\hat{\beta}_0 = \bar{y} - \bar{\mathbf{x}}^{\mathsf{T}}\hat{\mathbf{b}}$，$\hat{\beta}_0^* = \bar{y}$，$\hat{\beta}_0^{**} \equiv 0$，故第二个模型中$\beta_0^{**}$通常不用写出来。

注1：中心化不影响回归系数**b**，但改变了截距项。

注2：我们一般采用第一种中心化（Z中心化）

拟合优度：决定系数 R²

模型： $\mathbf{y}=X\boldsymbol{\beta}+\boldsymbol{\varepsilon}=\mathbf{1}\beta_0+Z\mathbf{b}+\boldsymbol{\varepsilon},\quad \boldsymbol{\varepsilon}\sim(0,\sigma^2 I_n)$

拟合值：$\hat{\mathbf{y}}=P_X\mathbf{y}\in C(X)$； 残差：$\mathbf{e}=\mathbf{y}^{\perp}=\mathbf{y}-\hat{\mathbf{y}}=(I_n-P_X)\mathbf{y}\perp C(X)$

$\mathbf{e}\perp C(X)$,特别地$\mathbf{e}\perp\mathbf{1}\Rightarrow\sum e_i=\sum(y_i-\hat{y}_i)=0\Rightarrow\bar{y}=\sum y_i/n=\sum\hat{y}_i/n=\bar{\hat{y}}$。

或：因为$P_X\mathbf{1}=\mathbf{1}$, 所以$\hat{\mathbf{y}}=(\hat{y}_1,\dots,\hat{y}_n)^{\top}$的样本均值:

$$\bar{\hat{y}}=\mathbf{1}^{\top}\hat{\mathbf{y}}/n=\mathbf{1}^{\top}P_X\mathbf{y}/n=\mathbf{1}^{\top}\mathbf{y}/n=\bar{y}$$

定义：总平方和$\mathrm{SS}_{总}=s_{yy}=\sum_{i=1}^{n}(y_i-\bar{y})^2=\|\mathbf{y}-\mathbf{1}\bar{y}\|^2$

回归平方和$\mathrm{SS}_{回}=s_{\hat{y}\hat{y}}=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2=\|\hat{\mathbf{y}}-\mathbf{1}\bar{y}\|^2$

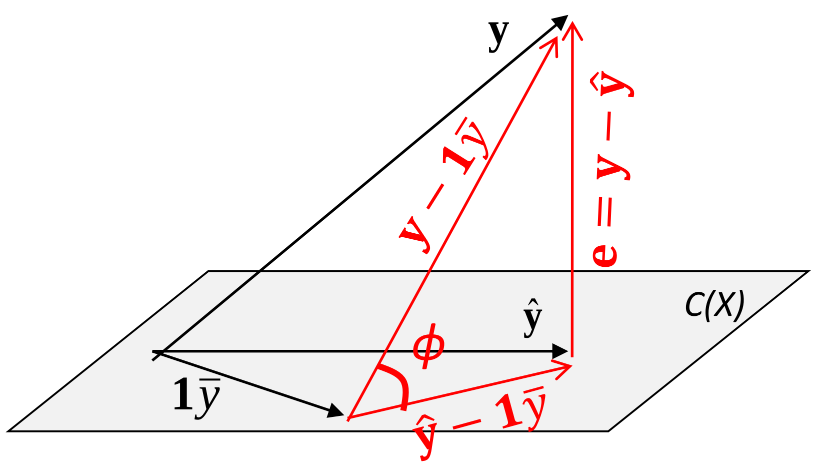
决定系数$R^2=\dfrac{\mathrm{SS}_{回}}{\mathrm{SS}_{总}}=\dfrac{\|\hat{\mathbf{y}}-\mathbf{1}\bar{y}\|^2}{\|\mathbf{y}-\mathbf{1}\bar{y}\|^2}=\dfrac{VAR(\hat{\mathbf{y}})}{VAR(\mathbf{y})}$， VAR表示样本方差。

决定系数度量了模型拟合的程度(拟合优度)/自变量对响应的解释能力。

因为$\mathbf{e}\perp\hat{\mathbf{y}}$，$\mathbf{e}\perp\mathbf{1}$，所以$\mathbf{e}\perp(\hat{\mathbf{y}}-\mathbf{1}\bar{y})\Rightarrow \mathbf{y}-\mathbf{1}\bar{y}=(\hat{\mathbf{y}}-\mathbf{1}\bar{y})\oplus\mathbf{e}$，（下图红色三角形）有平方分解：

$$\begin{array}{ccccc} \|\mathbf{y}-\mathbf{1}\bar{y}\|^2 & = & \|\hat{\mathbf{y}}-\mathbf{1}\bar{y}\|^2 & + & \|\mathbf{e}\|^2 \\ \mathrm{SS}_{总} & = & \mathrm{SS}_{回} & + & \mathrm{RSS} \end{array}$$

所以，$\mathrm{SS}_{回}\le\mathrm{SS}_{总}$, $0\le R^2\le 1$。



$$上图表明R^2=\frac{\|\hat{\mathbf{y}}-\mathbf{1}\bar{y}\|^2}{\|\mathbf{y}-\mathbf{1}\bar{y}\|^2}=\cos^2(\phi)=\left(r_{y\hat{y}}\right)^2$$

命题1. $R^2=\left(r_{y\hat{y}}\right)^2=S_{y\mathbf{x}}S_{\mathbf{xx}}^{-}S_{\mathbf{x}y}/S_{yy}$，其中$r_{y\hat{y}}$为$(y_i,\hat{y}_i),i=1,\ldots,n$的样本相关系数，$S_{\mathbf{xx}}=\frac{1}{n-1}\sum(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^{\top}$，$S_{\mathbf{x}y}=\frac{1}{n-1}\sum(\mathbf{x}_i-\bar{\mathbf{x}})(y_i-\bar{y})$。

证明：（1）因为$\mathbf{e}=\mathbf{y}-\hat{\mathbf{y}}\perp\hat{\mathbf{y}}$，$\mathbf{e}\perp\mathbf{1}\Rightarrow\langle\mathbf{e},\hat{\mathbf{y}}-\mathbf{1}\bar{y}\rangle=0$

$$\langle\mathbf{y}-\mathbf{1}\bar{y},\hat{\mathbf{y}}-\mathbf{1}\bar{y}\rangle=\langle\hat{\mathbf{y}}-\mathbf{1}\bar{y},\hat{\mathbf{y}}-\mathbf{1}\bar{y}\rangle=\|\hat{\mathbf{y}}-\mathbf{1}\bar{y}\|^2$$

所以$r_{y\hat{y}}=\dfrac{\langle\mathbf{y}-\mathbf{1}\bar{y},\hat{\mathbf{y}}-\mathbf{1}\bar{y}\rangle}{\|\mathbf{y}-\mathbf{1}\bar{y}\|\|\hat{\mathbf{y}}-\mathbf{1}\bar{y}\|}=\dfrac{\|\hat{\mathbf{y}}-\mathbf{1}\bar{y}\|}{\|\mathbf{y}-\mathbf{1}\bar{y}\|}\Rightarrow r_{y\hat{y}}{}^2=R^2$

$A^{\perp\top}B^{\perp}=A^{\perp\top}B$
其中A,B 为矩阵或向量
$A^{\perp}=A-P_CA$
$B^{\perp}=B-P_CB$

$Z_c^{\top}\mathbf{y}=Z_c^{\top}\mathbf{y}_c$

(2)因为$\hat{\mathbf{y}}=P_{\mathbf{1},Z}\mathbf{y}=P_{\mathbf{1},Z_c}\mathbf{y}=P_{\mathbf{1}}\mathbf{y}+P_{Z_c}\mathbf{y}=\mathbf{1}\bar{y}+Z_c(Z_c{}^{\top}Z_c)^{-}Z_c{}^{\top}\mathbf{y}$，

所以由$Z_c{}^{\top}Z_c/(n-1)=S_{\mathbf{xx}},Z_c{}^{\top}\mathbf{y}/(n-1)=S_{\mathbf{x}y}$，

$$\mathrm{SS}_{回}=\|\hat{\mathbf{y}}-\mathbf{1}\bar{y}\|^2=\left\|Z_c(Z_c{}^{\top}Z_c)^{-}Z_c{}^{\top}\mathbf{y}\right\|^2=\mathbf{y}^{\top}Z_c(Z_c{}^{\top}Z_c)^{-}Z_c{}^{\top}\mathbf{y}=(n-1)S_{y\mathbf{x}}S_{\mathbf{xx}}^{-}S_{\mathbf{x}y},$$

而$\mathrm{SS}_{总}=(n-1)S_{yy}$，所以$R^2=S_{y\mathbf{x}}S_{\mathbf{xx}}^{-}S_{\mathbf{x}y}/S_{yy}$

命题2. $R^2=\max\limits_{\mathbf{u}\in L(X)}(r_{\mathbf{u},\mathbf{y}})^2$, 其中$r_{\mathbf{u},\mathbf{y}}$为样本相关系数，

且最大值在$\mathbf{u}=\hat{\mathbf{y}}=X\hat{\boldsymbol\beta}$处达到，其中$\hat{\boldsymbol\beta}=(X^\top X)^- X^\top\mathbf{y}$。

第3讲命题3的特殊情形

证明：记$\mathbf{y}_c\triangleq\mathbf{y}-\mathbf{1}\bar{y}$。对任何$\mathbf{u}=X\boldsymbol\beta=\mathbf{1}\beta_0+Z\mathbf{b}\in C(X)$,

记$\mathbf{u}$的中心化$\mathbf{u}_c=\mathbf{u}-P_\mathbf{1}\mathbf{u}=\mathbf{1}\beta_0+Z\mathbf{b}-(P_\mathbf{1}\mathbf{1}\beta_0+P_\mathbf{1}Z\mathbf{b})=Z\mathbf{b}-P_\mathbf{1}Z\mathbf{b}=Z_c\mathbf{b}$

则 $(r_{\mathbf{u},\mathbf{y}})^2=\dfrac{(\mathbf{u}_c^\top\mathbf{y}_c)^2}{\|\mathbf{u}_c\|^2\|\mathbf{y}_c\|^2}=\dfrac{(\mathbf{b}^\top Z_c^\top\mathbf{y}_c)^2}{\mathbf{b}^\top Z_c^\top Z_c\mathbf{b}\times\|\mathbf{y}_c\|^2}=\dfrac{\left(\mathbf{w}^\top\left[(Z_c^\top Z_c)^{-1/2}Z_c^\top\mathbf{y}_c\right]\right)^2}{\mathbf{w}^\top\mathbf{w}\cdot\|\mathbf{y}_c\|^2}$

$\mathbf{w}=(Z_c^\top Z_c)^{1/2}\mathbf{b}$
$\mathbf{b}=(Z_c^\top Z_c)^{-1/2}\mathbf{w}$

$\le\dfrac{\mathbf{w}^\top\mathbf{w}\ \times\ \mathbf{y}_c^\top Z_c(Z_c^\top Z_c)^- Z_c^\top\mathbf{y}_c}{\mathbf{w}^\top\mathbf{w}\times\|\mathbf{y}_c\|^2}$ （Cauchy-Schwartz不等式）

$=\dfrac{\mathbf{w}^\top\mathbf{w}\times\|P_{Z_c}\mathbf{y}_c\|^2}{\mathbf{w}^\top\mathbf{w}\times\|\mathbf{y}_c\|^2}=\dfrac{\|P_{Z_c}\mathbf{y}\|^2}{\|\mathbf{y}_c\|^2}=\dfrac{\|\hat{\mathbf{y}}-\mathbf{1}\bar{y}\|^2}{\|\mathbf{y}-\mathbf{1}\bar{y}\|^2}=R^2$

$P_{Z_c}\mathbf{y}_c=P_{Z_c}\mathbf{y}=\hat{\mathbf{y}}-\mathbf{1}\bar{y}$

当$\mathbf{w}\propto(Z_c^\top Z_c)^{-1/2}Z_c^\top\mathbf{y}_c\Leftrightarrow\mathbf{b}=(Z_c^\top Z_c)^{-1/2}\mathbf{w}\propto(Z_c^\top Z_c)^- Z_c^\top\mathbf{y}_c$时，等号成立.

此时$\mathbf{u}_c=Z_c\mathbf{b}\propto Z_c(Z_c^\top Z_c)^- Z_c^\top\mathbf{y}_c=Z_c\hat{\mathbf{b}}$,特别地在$\mathbf{u}=Z_c\hat{\mathbf{b}}+\mathbf{1}\bar{y}=X\hat{\boldsymbol\beta}$达到极大。

回归系数b的LS估计

我们将假设X列满秩。自变量的回归系数**b**的LS估计仅依赖于样本协方差矩阵，截距项LS估计与样本均值有关。

命题3. 假设模型$\mathbf{y}=X\boldsymbol{\beta}+\boldsymbol{\varepsilon}=\mathbf{1}\beta_0+Z\mathbf{b}+\boldsymbol{\varepsilon},\boldsymbol{\varepsilon}\sim(0,\ \sigma^2 I_n)$与$X$独立。

假设$X=(\mathbf{1},Z)$列满秩, $\boldsymbol{\beta}=\begin{pmatrix}\beta_0\\ \mathbf{b}\end{pmatrix}, Z=(\mathbf{x}_1,\ldots,\mathbf{x}_n)^{\mathrm{T}}, Z_c=Z-\mathbf{1}\bar{\mathbf{x}}^{\mathrm{T}}$。则

(1) LS估计 $\hat{\mathbf{b}}=(Z_c Z_c^{\mathrm{T}})^{-1}Z_c^{\mathrm{T}}\mathbf{y}=S_{\mathbf{xx}}^{-1}S_{\mathbf{x}y}$, $\hat{\beta}_0=\bar{y}-\bar{\mathbf{x}}^{\mathrm{T}}\hat{\mathbf{b}}$.

(2) $\mathrm{var}(\hat{\mathbf{b}}\mid X)=\sigma^2(Z_c Z_c^{\mathrm{T}})^{-1}=\sigma^2 S_{\mathbf{xx}}^{-1}/(n-1)$.

(3) $\hat{\sigma}^2=\dfrac{n-1}{n-p}S_{yy\bullet\mathbf{x}}\approx S_{yy\bullet\mathbf{x}}=S_{yy}-S_{y\mathbf{x}}S_{\mathbf{xx}}^{-1}S_{\mathbf{x}y}$

$=S_{yy}(1-R^2),\quad R^2=S_{y\mathbf{x}}S_{\mathbf{xx}}^{-1}S_{\mathbf{x}y}/S_{yy}$.

对比第5讲命题2(总体版本)

$\mathbf{b}=\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y}$

$\beta_0=\mu_y-\mathbf{b}^{\mathrm{T}}\boldsymbol{\mu}_{\mathbf{x}}$

$\sigma^2=\Sigma_{yy\bullet\mathbf{x}}=(1-\Phi)\Sigma_{yy}$

$\Phi=\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y}/\Sigma_{yy}$

注：将**b**、β_0、σ^2中的总体方差$\Sigma_{\mathbf{xx}}$、$\Sigma_{\mathbf{x}y}$换成样本方差$S_{\mathbf{xx}}$、$S_{\mathbf{x}y}$, 总体均值换成样本均值即得矩估计。**b**、β_0的矩估计与LS估计相同，σ^2的估计略有差异(相差倍数 $(n-1)/(n-p)\approx 1$）。

证1:（1）先将Z中心化（即Z与**1**正交化），

令$Z_c = Z - P_{\mathbf{1}}Z = Z - \mathbf{1}\bar{\mathbf{x}}^{\mathsf{T}}$，改写模型

$$\mathbf{y} = \mathbf{1}\beta_0 + Z\mathbf{b} + \boldsymbol{\varepsilon} = \mathbf{1}\beta_0^* + Z_c\mathbf{b} + \boldsymbol{\varepsilon}, \quad \beta_0^* = \beta_0 + \bar{\mathbf{x}}^{\mathsf{T}}\mathbf{b}$$

则$\mathbf{y}$在$C(X) = C(\mathbf{1}, Z) = C(\mathbf{1}, Z_c)$上的投影：

$$\hat{\mathbf{y}} = P_X\mathbf{y} = P_{\mathbf{1}}\mathbf{y} + P_{Z_c}\mathbf{y} = \mathbf{1}\underline{\underline{\mathbf{1}^{\mathsf{T}}\mathbf{y}/n}} + Z_c\,\underline{\underline{(Z_c{}^{\mathsf{T}}Z_c)^{-1}Z_c{}^{\mathsf{T}}\mathbf{y}}} = \mathbf{1}\hat{\beta}_0^* + Z_c\hat{\mathbf{b}}$$

1的系数为$\hat{\beta}_0^*$，
Z的系数为$\hat{\mathbf{b}}$

所以$\hat{\mathbf{b}} = (Z_c{}^{\mathsf{T}}Z_c)^{-1}Z_c{}^{\mathsf{T}}\mathbf{y}$，$\hat{\beta}_0^* = \mathbf{1}^{\mathsf{T}}\mathbf{y}/n = \bar{y} \Rightarrow \hat{\beta}_0 = \hat{\beta}_0^* - \bar{\mathbf{x}}^{\mathsf{T}}\hat{\mathbf{b}} = \bar{y} - \bar{\mathbf{x}}^{\mathsf{T}}\hat{\mathbf{b}}.$

或对中心化模型

$$\mathbf{y} = \mathbf{1}\beta_0^* + Z_c\mathbf{b} + \boldsymbol{\varepsilon} \overset{\Delta}{=} X^*\boldsymbol{\beta}^* + \boldsymbol{\varepsilon}, \quad X^* = (\mathbf{1}, Z_c), \beta_0^* = \beta_0 + \bar{\mathbf{x}}^{\mathsf{T}}\mathbf{b}$$

应用LS估计公式：

$$\hat{\boldsymbol{\beta}}^* = \begin{pmatrix} \hat{\beta}_0^* \\ \hat{\mathbf{b}} \end{pmatrix} = (X^{*\mathsf{T}}X^*)^{-1}X^{*\mathsf{T}}\mathbf{y} = \begin{pmatrix} n & 0 \\ 0 & Z_c{}^{\mathsf{T}}Z_c \end{pmatrix}^{-1}\begin{pmatrix} \mathbf{1}^{\mathsf{T}}\mathbf{y} \\ Z_c{}^{\mathsf{T}}\mathbf{y} \end{pmatrix} = \begin{pmatrix} \bar{y} \\ (Z_c{}^{\mathsf{T}}Z_c)^{-1}Z_c{}^{\mathsf{T}}\mathbf{y} \end{pmatrix}$$

$\Rightarrow \hat{\mathbf{b}} = (Z_c{}^{\mathsf{T}}Z_c)^{-1}Z_c{}^{\mathsf{T}}\mathbf{y}$，$\hat{\beta}_0^* = \bar{y}$。

注意$Z_c{}^{\mathsf T}Z_c=\sum(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^{\mathsf T}=(n-1)S_{\mathbf{xx}}$，$Z_c{}^{\mathsf T}\mathbf{y}=(n-1)S_{\mathbf{x}y}\Rightarrow$

$$\hat{\mathbf{b}}=(Z_cZ_c{}^{\mathsf T})^{-}Z_c{}^{\mathsf T}\mathbf{y}=S_{\mathbf{xx}}^{-1}S_{\mathbf{x}y}$$

（2）显然$\mathrm{var}(\hat{\mathbf{b}}\mid X)=\sigma^2(Z_c{}^{\mathsf T}Z_c)^{-1}$，所以

$$\mathrm{var}(\hat{\mathbf{b}}\mid X)=\sigma^2(Z_c{}^{\mathsf T}Z_c)^{-1}=\sigma^2S_{\mathbf{xx}}^{-1}/(n-1)\text{。}$$

（3）由$\hat{\mathbf{y}}=P_{\mathbf{1}}\mathbf{y}+P_{Z_c}\mathbf{y}=\mathbf{1}\bar{y}+P_{Z_c}\mathbf{y}$，知$\hat{\mathbf{y}}-\mathbf{1}\bar{y}=P_{Z_c}\mathbf{y}$，

$Z_c{}^{\mathsf T}\mathbf{y}=Z_c{}^{\mathsf T}\mathbf{y}_c$

所以

$$\mathrm{SS}_{回}=\|\hat{\mathbf{y}}-\mathbf{1}\bar{y}\|^2=\|P_{Z_c}\mathbf{y}\|^2=\mathbf{y}^{\mathsf T}Z_c(Z_c{}^{\mathsf T}Z_c)^{-1}Z_c{}^{\mathsf T}\mathbf{y}=(n-1)S_{y\mathbf{x}}S_{\mathbf{xx}}^{-1}S_{\mathbf{x}y}$$

而$\mathrm{SS}_{总}=\|\mathbf{y}-\mathbf{1}\bar{y}\|^2=(n-1)S_{yy}$，所以

$$RSS=\mathrm{SS}_{总}-\mathrm{SS}_{回}=(n-1)\left[S_{yy}-S_{y\mathbf{x}}S_{\mathbf{xx}}^{-1}S_{\mathbf{x}y}\right]=(n-1)S_{yy\bullet\mathbf{x}},$$

所以

$$\hat{\sigma}^2=\mathrm{RSS}/(n-p)=\frac{n-1}{n-p}S_{yy\bullet\mathbf{x}},\quad R^2=\frac{S_{y\mathbf{x}}S_{\mathbf{xx}}^{-1}S_{\mathbf{x}y}}{S_{yy}}.$$

其它更直接但更复杂的证明（可略过）：

证2. (应用分块矩阵的逆，求$\hat{\boldsymbol\beta}$的分量) $\mathbf{y}=X\boldsymbol\beta+\boldsymbol\varepsilon=(\mathbf{1},\mathrm{Z})\begin{pmatrix}\beta_0\\ \mathbf{b}\end{pmatrix}+\boldsymbol\varepsilon$

由定理1，已知LS估计$\hat{\boldsymbol\beta}=(X^\top X)^{-1}X^\top\mathbf{y}$, 下面求分量$\hat{\mathbf{b}}$。

分块：$X^\top X=\begin{pmatrix}\mathbf{1}^\top\\ Z^\top\end{pmatrix}(\mathbf{1},\mathrm{Z})=\begin{pmatrix}\mathbf{1}^\top\mathbf{1} & \mathbf{1}^\top Z\\ Z^\top\mathbf{1} & Z^\top Z\end{pmatrix}=\begin{pmatrix}n & n\bar{\mathbf{x}}^\top\\ n\bar{\mathbf{x}} & \sum\mathbf{x}_i\mathbf{x}_i^\top\end{pmatrix}$, $X^\top\mathbf{y}=\begin{pmatrix}\mathbf{1}^\top\\ Z^\top\end{pmatrix}\mathbf{y}=\begin{pmatrix}n\bar{y}\\ \sum\mathbf{x}_i y_i\end{pmatrix}$,

由分块矩阵求逆公式（第4讲命题1）

$$(X^\top X)^{-1}=\begin{pmatrix}a & -\bar{\mathbf{x}}^\top A^{-1}\\ -A^{-1}\bar{\mathbf{x}} & A^{-1}\end{pmatrix},\text{其中}A=(n-1)S_{\mathbf{xx}},\quad a=1/n+\bar{\mathbf{x}}^\top A^{-1}\bar{\mathbf{x}}$$

$$\Rightarrow\hat{\boldsymbol\beta}=\begin{pmatrix}\hat\beta_0\\ \hat{\mathbf{b}}\end{pmatrix}=\begin{pmatrix}n & n\bar{\mathbf{x}}^\top\\ n\bar{\mathbf{x}} & \sum\mathbf{x}_i\mathbf{x}_i^\top\end{pmatrix}^{-1}\begin{pmatrix}n\bar{y}\\ \sum\mathbf{x}_i y_i\end{pmatrix}=\begin{pmatrix}a & -\bar{\mathbf{x}}^\top A^{-1}\\ -A^{-1}\bar{\mathbf{x}} & A^{-1}\end{pmatrix}\begin{pmatrix}n\bar{y}\\ \sum\mathbf{x}_i y_i\end{pmatrix}$$

$$=\begin{pmatrix}na\bar{y}-\bar{\mathbf{x}}^\top A^{-1}\sum\mathbf{x}_i y_i\\ A^{-1}\sum\mathbf{x}_i y_i-nA^{-1}\overline{\mathbf{x}y}\end{pmatrix}=\begin{pmatrix}\bar{y}-\bar{\mathbf{x}}^\top S_{\mathbf{xx}}^{-1}S_{\mathbf{x}y}\\ S_{\mathbf{xx}}^{-1}S_{\mathbf{x}y}\end{pmatrix}$$

所以$\hat{\mathbf{b}}=S_{\mathbf{xx}}^{-1}S_{\mathbf{x}y}$，$\hat\beta_0=\bar{y}-\bar{\mathbf{x}}^\top\hat{\mathbf{b}}$.

第4讲命题1：$\Sigma>0$,

$$\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}^{-1}=\begin{pmatrix}\Sigma_{11\bullet2}^{-1} & -\Sigma_{11\bullet2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\\ -\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11\bullet2}^{-1} & \Sigma_{22\bullet1}^{-1}\end{pmatrix}$$

证3（误差平方和分别对β_0和$\mathbf{b}$求导） $\boldsymbol{\beta}=(\beta_0,\mathbf{b}^{\mathsf{T}})^{\mathsf{T}}$

$$Q(\boldsymbol{\beta})=\|\mathbf{y}-X\boldsymbol{\beta}\|^2=\sum_{i=1}^{n}\left(y_i-\tilde{\mathbf{x}}_i^{\mathsf{T}}\boldsymbol{\beta}\right)^2=\sum_{i=1}^{n}\left(y_i-\beta_0-\mathbf{x}_i^{\mathsf{T}}\mathbf{b}\right)^2$$

$$\frac{\partial Q}{\partial \boldsymbol{\beta}}=0\Leftrightarrow\begin{cases}\dfrac{\partial Q}{\partial \beta_0}=\sum\limits_{i=1}^{n}\left(y_i-\beta_0-\mathbf{x}_i^{\mathsf{T}}\mathbf{b}\right)=0\\ \dfrac{\partial Q}{\partial \mathbf{b}}=\sum\limits_{i=1}^{n}\mathbf{x}_i\left(y_i-\beta_0-\mathbf{x}_i^{\mathsf{T}}\mathbf{b}\right)=0\end{cases}\Leftrightarrow\begin{cases}\sum y_i-n\beta_0-\left(\sum \mathbf{x}_i\right)^{\mathsf{T}}\mathbf{b}=0\\ \sum \mathbf{x}_i y_i-\left(\sum \mathbf{x}_i\right)\beta_0-\left(\sum \mathbf{x}_i\mathbf{x}_i^{\mathsf{T}}\right)\mathbf{b}=0\end{cases}$$

$$\Leftrightarrow\begin{cases}\bar{y}-\beta_0-\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{b}=0\\ \sum \mathbf{x}_i y_i-n\bar{\mathbf{x}}\beta_0-\left(\sum \mathbf{x}_i\mathbf{x}_i^{\mathsf{T}}\right)\mathbf{b}=0\end{cases}$$

第一式左乘$n\bar{\mathbf{x}}$减去第二式，消去β_0，得:

消去β_0的过程⇔中心化

$$\left(\sum \mathbf{x}_i\mathbf{x}_i^{\mathsf{T}}-n\bar{\mathbf{x}}\bar{\mathbf{x}}^{\mathsf{T}}\right)\mathbf{b}=\sum \mathbf{x}_i y_i-n\overline{\mathbf{x}}\overline{y}$$

$$\Rightarrow\hat{\mathbf{b}}=\left(\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^{\mathsf{T}}\right)^{-1}\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})y_i=S_{\mathbf{xx}}^{-1}S_{\mathbf{x}y},\hat{\beta}_0=\bar{y}-\bar{\mathbf{x}}^{\mathsf{T}}\hat{\mathbf{b}},$$

为了简化论证过程，常常“不妨假设数据已经中心化或中心标准化”

我们已知：

- 如果$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{1}\beta_0 + Z\mathbf{b} + \boldsymbol{\varepsilon}$中$Z$已经中心化，那么

$$\hat{\mathbf{b}} = (Z^{\top}Z)^{-1}Z^{\top}\mathbf{y} = S_{\mathbf{xx}}^{-1}S_{\mathbf{xy}}，\ \hat{\beta}_0 = \bar{y}$$

- 如果$\mathbf{y}$，X都已经中心化，那么

$$\hat{\mathbf{b}} = (Z^{\top}Z)^{-1}Z^{\top}\mathbf{y} = S_{\mathbf{xx}}^{-1}S_{\mathbf{xy}}，\ \hat{\beta}_0 \equiv 0，\text{此时}\beta_0\text{无需估计。}$$

假设$\mathbf{y}, Z$都已经中心化，则截距项LS估计恒为0，故假设无截距项：

- $\mathbf{y} = Z\mathbf{b} + \boldsymbol{\varepsilon}$（不含截距项），假设$\mathbf{y}$，$Z$都已经中心化，

$$\hat{\mathbf{b}} = (Z^{\top}Z)^{-1}Z^{\top}\mathbf{y} = S_{\mathbf{xx}}^{-1}S_{\mathbf{xy}}，\ S:\text{样本方差矩阵}$$

- $\mathbf{y} = Z\mathbf{b} + \boldsymbol{\varepsilon}$（不含截距项），假设$\mathbf{y}$，$Z$都已经中心标准化，

$$\hat{\mathbf{b}} = (Z^{\top}Z)^{-1}Z^{\top}\mathbf{y} = R_{\mathbf{xx}}^{-1}R_{\mathbf{xy}}，\ R:\text{样本相关系数矩阵。}$$

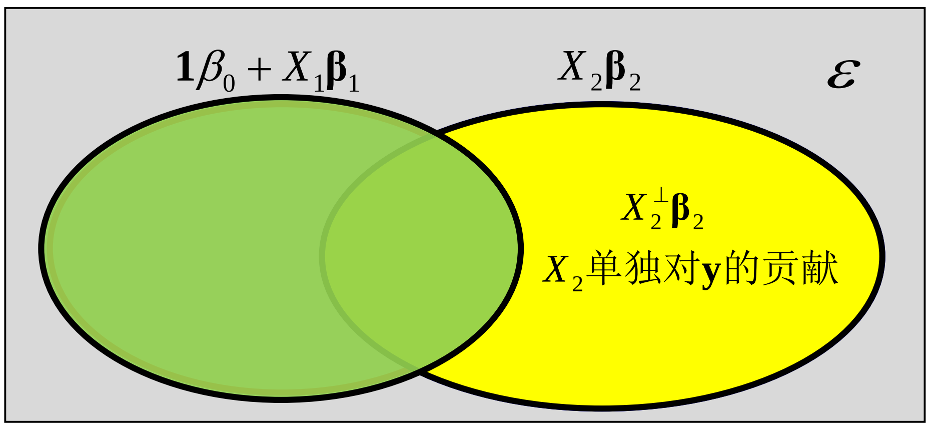
部分回归系数的LS估计

假设设计阵X按列划分为$X=(\mathbf{1},X_1,X_2)$，模型为

$$\mathbf{y}=X\boldsymbol{\beta}+\boldsymbol{\varepsilon}=\mathbf{1}\beta_0+X_1\boldsymbol{\beta}_1+X_2\boldsymbol{\beta}_2+\boldsymbol{\varepsilon}$$

其中X_1的列是干扰因素，X_2的列是感兴趣的变量。LS估计$\hat{\boldsymbol{\beta}}_2$是否有效地消除了$X_1$的影响？

$\mathbf{y}=$



$$X_2^{\perp}=X_2-P_{\mathbf{1},X_1}X_2$$

$$\hat{\boldsymbol{\beta}}_2=(X_2^{\perp\mathsf{T}}X_2^{\perp})^{-1}X_2^{\perp\mathsf{T}}\mathbf{y}$$

记号：记所有数据$(\mathbf{y}, X_1, X_2)$的样本方差-协方差矩阵为

$$S = \begin{array}{c|ccc} & \mathbf{x}_1 & \mathbf{x}_2 & y \\ \hline \mathbf{x}_1 & S_{11} & S_{12} & S_{1y} \\ \mathbf{x}_2 & S_{21} & S_{22} & S_{2y} \\ y & S_{y1} & S_{y2} & S_{yy} \end{array}, \quad \begin{pmatrix} S_{22\bullet1} & S_{2y\bullet1} \\ S_{y2\bullet1} & S_{yy\bullet1} \end{pmatrix} = \begin{pmatrix} S_{22} & S_{2y} \\ S_{y2} & S_{yy} \end{pmatrix} - \begin{pmatrix} S_{21} \\ S_{y1} \end{pmatrix} S_{11}^{-1} (S_{12}, S_{1y})$$

LS估计的分量

命题4：假设线性模型：

$$\mathbf{y}_{n\times1} = X_{n\times p}\boldsymbol{\beta}_{p\times1} + \boldsymbol{\varepsilon}_{n\times1} = \mathbf{1}\beta_0 + X_1\boldsymbol{\beta}_1 + X_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim (\mathbf{0}, \sigma^2 I_n), \ \boldsymbol{\varepsilon} \perp\!\!\!\perp X$$

其中$X = (\mathbf{1}, X_1, X_2)$列满秩,$\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1^{\top}, \boldsymbol{\beta}_2^{\top})^{\top}$，令$X_2^{\perp} = X_2 - P_{\mathbf{1}, X_1} X_2$，则

$$\hat{\boldsymbol{\beta}}_2 = (X_2^{\perp\top} X_2^{\perp})^{-1} X_2^{\perp\top}\mathbf{y} = S_{22\bullet1}^{-1} S_{2y\bullet1},$$

$$\mathrm{var}(\hat{\boldsymbol{\beta}}_2 \mid X) = \sigma^2 (X_2^{\perp\top} X_2^{\perp})^{-1} = \sigma^2 S_{22\bullet1}^{-1}/(n-1),$$

对比第5讲命题3(总体)

$\boldsymbol{\beta}_2 = \Sigma_{22\bullet1}^{-1}\Sigma_{2y\bullet1}$

这里给出$\hat{\boldsymbol{\beta}}$的分量的数学表达是为了理解参数估计的含义。实际数据计算时，无需应用命题4，只需要套用一般公式$\hat{\boldsymbol{\beta}} = (X^{\top}X)^{-1}X^{\top}\mathbf{y}$即可。

$\hat{\beta}_2$的评注

注1. $\boldsymbol{\beta}$的整体LS估计$\hat{\boldsymbol{\beta}}=\left(X^{\mathrm{T}}X\right)^{-1}X^{\mathrm{T}}\mathbf{y}$和它的一部分$\hat{\boldsymbol{\beta}}_2=\left(X_2^{\perp\mathrm{T}}X_2^{\perp}\right)^{-1}X_2^{\perp\mathrm{T}}\mathbf{y}$形式完全相同。

$$\hat{\boldsymbol{\beta}}_2=\left(X_2^{\perp\mathrm{T}}X_2^{\perp}\right)^{-1}X_2^{\perp\mathrm{T}}\mathbf{y}=\left(X_2^{\perp\mathrm{T}}X_2^{\perp}\right)^{-1}X_2^{\perp\mathrm{T}}\mathbf{y}^{\perp},\mathbf{y}^{\perp}=\mathbf{y}-P_{X_1}\mathbf{y}$$

表明$\hat{\boldsymbol{\beta}}_2$是$X_2$和/或$\mathbf{y}$消除了$X_1$影响之后的估计，即$\mathbf{y}\sim X_2^{\perp}$或$\mathbf{y}^{\perp}\sim X_2^{\perp}\Rightarrow\hat{\boldsymbol{\beta}}_2$。

注2. 特别地，当X_2是向量时，$\hat{\beta}_2=S_{22\bullet1}^{-1}S_{2y\bullet1}\propto r_{2y\bullet1}$(样本偏相关系数)。即线性回归与偏相关系数本质上相同，都是消除了X_1的影响之后得到的$\mathbf{y}$和X_2之间的关系。

注3: 由注1的观察，$\hat{\boldsymbol{\beta}}_2$可由两步回归得到：

(1) $X_2\sim X_1$（多元回归，X_2的每一列对X_1回归）所有残差按列排列得：

$$X_2^{\perp}=X_2-P_{1,X_1}X_2;$$

(2) $\mathbf{y}\sim X_2^{\perp}$：$\mathbf{y}=X_2^{\perp}\boldsymbol{\beta}_2+\boldsymbol{\delta}$得$\hat{\boldsymbol{\beta}}_2=(X_2^{\perp\mathrm{T}}X_2^{\perp})^{-1}X_2^{\perp\mathrm{T}}\mathbf{y}$。

注4. 由命题4所给的表达式，我们可以如下方式简单求解$\hat{\boldsymbol{\beta}}_2$:

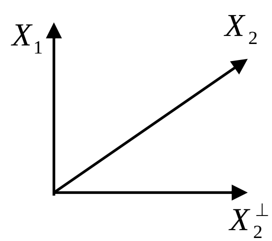
模型 $\mathbf{y} = X_1\boldsymbol{\beta}_1 + X_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$ 两端同时左乘$X_2^{\perp\top}$，并令$X_2^{\perp\top}\boldsymbol{\varepsilon} = \mathbf{0}$，得

$$X_2^{\perp\top}\mathbf{y} = X_2^{\perp\top}X_2\boldsymbol{\beta}_2 \Rightarrow \hat{\boldsymbol{\beta}}_2 = \left(X_2^{\perp\top}X_2^{\perp}\right)^{-1}X_2^{\perp\top}\mathbf{y}.$$

为什么方程两端同时左乘$X_2^{\perp\top}$？

(1) $X_2^{\perp}$与X_1列正交（消除$\boldsymbol{\beta}_1$），

(2) $X_2^{\perp}$与X_2最接近(投影的LS性质)。



一般地，假设W是任一$n\times q$矩阵，$W^\top X_1 = 0, W^\top X_2$可逆，

模型方程两端同时左乘$W^\top$并令$W^\top\boldsymbol{\varepsilon} = 0$，得

$$W^\top\mathbf{y} = W^\top X_2\boldsymbol{\beta}_2 \Rightarrow \tilde{\boldsymbol{\beta}}_2 = (W^\top X_2)^{-1}W^\top\mathbf{y},$$

$\tilde{\boldsymbol{\beta}}_2$是无偏估计，由GM定理，当$W = X_2^{\perp} = X_2 - P_{X_1}X_2$时方差最小。

命题4的证明1:

⑴ 不妨设 X_1, X_2已经中心化，所以$\mathbf{1}, X_1, X_2^{\perp}$三部分相互列正交.
下面我们将投影 $\hat{\mathbf{y}} = P_X\mathbf{y} = P_{(\mathbf{1},X_1,X_2)}\mathbf{y}$表示成$\mathbf{1}$，$X_1$，$X_2$的组合形式
$\mathbf{1}\{\ \} + X_1\{\ \} + X_2\{\ \}$，则三个组合系数分别是$\hat{\beta}_0$，$\hat{\boldsymbol{\beta}}_1$，$\hat{\boldsymbol{\beta}}_2$.

$$
\begin{aligned}
\hat{\mathbf{y}} &= P_{(\mathbf{1},X_1,X_2)}\mathbf{y} = P_{(\mathbf{1},X_1,X_2^{\perp})}\mathbf{y} = P_{\mathbf{1}}\mathbf{y} + P_{X_1}\mathbf{y} + P_{X_2^{\perp}}\mathbf{y} \\
&= \mathbf{1}\bar{y} + X_1\left[(X_1^{\top}X_1)^{-1}X_1^{\top}\mathbf{y}\right] + X_2^{\perp}\left[(X_2^{\perp\top}X_2^{\perp})^{-1}X_2^{\perp\top}\mathbf{y}\right] \\
&\overset{\Delta}{=} \mathbf{1}\bar{y} + X_1[\mathbf{u}] + X_2^{\perp}[\mathbf{v}] \\
&= \mathbf{1}\bar{y} + X_1\mathbf{u} + \left(X_2 - X_1(X_1^{\top}X_1)^{-1}X_1^{\top}X_2\right)\mathbf{v} \\
&= \mathbf{1}\{\bar{y}\} + X_1\left\{\mathbf{u} - (X_1^{\top}X_1)^{-1}X_1^{\top}X_2\mathbf{v}\right\} + X_2\{\mathbf{v}\}
\end{aligned}
$$

$$\Rightarrow \hat{\beta}_0 = \bar{y}，\ \hat{\boldsymbol{\beta}}_1 = \mathbf{u} - (X_1^{\top}X_1)^{-1}X_1^{\top}X_2\mathbf{v}，\ \hat{\boldsymbol{\beta}}_2 = \mathbf{v} = (X_2^{\perp\top}X_2^{\perp})^{-1}X_2^{\perp\top}\mathbf{y}$$

由对称性$\hat{\boldsymbol{\beta}}_1$必定等于$(X_1^{\perp\top}X_1^{\perp})^{-1}X_1^{\perp\top}\mathbf{y}$，$X_1^{\perp} = X_1 - P_{X_2}X_1$

另外，$X_2^{\perp} = X_2 - P_{X_1} X_2 \Rightarrow X_2^{\perp\top} X_2^{\perp} = X_2^{\top} X_2 - X_2^{\top} X_1 (X_1^{\top} X_1)^{-1} X_1^{\top} X_2$

$= (n-1)S_{22} - (n-1)S_{21}S_{11}^{-1}S_{12} = (n-1)S_{22\bullet 1}$，同样$X_2^{\perp\top}\mathbf{y} = (n-1)S_{2y\bullet 1}$，

$\Rightarrow \hat{\boldsymbol{\beta}}_2 = (X_2^{\perp\top} X_2^{\perp})^{-1} X_2^{\perp\top}\mathbf{y} = S_{22\bullet 1}^{-1} S_{2y\bullet 1}$，

$\mathrm{var}(\hat{\boldsymbol{\beta}}_2 \mid X) = \sigma^2 (X_2^{\perp\top} X_2^{\perp})^{-1} = \sigma^2 S_{22\bullet 1}^{-1}/(n-1)$. 证毕。

证明2: (X_1, X_2先正交化，再应用LS估计一般公式) 不妨假设X_1, X_2已经中心化，这不影响$\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$的LS估计。$X_2^{\perp} = X_2 - P_{X_1} X_2$，改写模型：

$$\begin{aligned}
\mathbf{y} &= \mathbf{1}\beta_0 + X_1\boldsymbol{\beta}_1 + X_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} = \mathbf{1}\beta_0 + X_1\boldsymbol{\beta}_1 + (X_2^{\perp} + P_{X_1} X_2)\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} \\
&= \mathbf{1}\beta_0 + X_1\left[\boldsymbol{\beta}_1 + (X_1^{\top} X_1)^{-1} X_1^{\top} X_2 \boldsymbol{\beta}_2\right] + X_2^{\perp}\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} \\
&\triangleq \mathbf{1}\beta_0 + X_1\boldsymbol{\beta}_1^* + X_2^{\perp}\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}, \text{其中 } \boldsymbol{\beta}_1^* = \boldsymbol{\beta}_1 + (X_1^{\top} X_1)^{-1} X_1^{\top} X_2 \boldsymbol{\beta}_2
\end{aligned}$$

该模型的设计阵的三部分$\mathbf{1}, X_1, X_2^{\perp}$相互正交，$X^{\top}X$分块对角，所以由LS估计公式

$$\text{易得}\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\boldsymbol{\beta}}_1^* \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{1}^{\top}\mathbf{1} & 0 & 0 \\ 0 & X_1^{\top} X_1 & 0 \\ 0 & 0 & X_2^{\perp\top} X_2^{\perp} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{1}^{\top}\mathbf{y} \\ X_1^{\top}\mathbf{y} \\ X_2^{\perp\top}\mathbf{y} \end{pmatrix} = \begin{pmatrix} \bar{y} \\ (X_1^{\top} X_1)^{-1} X_1^{\top}\mathbf{y} \\ (X_2^{\perp\top} X_2^{\perp})^{-1} X_2^{\perp\top}\mathbf{y} \end{pmatrix} \Rightarrow \hat{\boldsymbol{\beta}}_2 = (X_2^{\perp\top} X_2^{\perp})^{-1} X_2^{\perp\top}\mathbf{y}。$$

证明3：（求导，不假设X_1，X_2已经中心化）

$Q=\|\mathbf{y}-\mathbf{1}\beta_0-X_1\boldsymbol{\beta}_1-X_2\boldsymbol{\beta}_2\|^2$ 求导得正则方程$X^\top(\mathbf{y}-X\boldsymbol{\beta})=\mathbf{0}$:

$$\begin{cases}\partial Q/\partial\beta_0=-\mathbf{1}^\top(\mathbf{y}-\mathbf{1}\beta_0-X_1\boldsymbol{\beta}_1-X_2\boldsymbol{\beta}_2)=0\\ \partial Q/\partial\boldsymbol{\beta}_1=-X_1{}^\top(\mathbf{y}-\mathbf{1}\beta_0-X_1\boldsymbol{\beta}_1-X_2\boldsymbol{\beta}_2)=\mathbf{0}\\ \partial Q/\partial\boldsymbol{\beta}_2=-X_2{}^\top(\mathbf{y}-\mathbf{1}\beta_0-X_1\boldsymbol{\beta}_1-X_2\boldsymbol{\beta}_2)=\mathbf{0}\end{cases}$$

下面应用初等消去法解正则方程(即求$X^\top X$的逆）

第一式整理得：$\beta_0=\bar{y}-\bar{\mathbf{x}}_1{}^\top\boldsymbol{\beta}_1-\bar{\mathbf{x}}_2{}^\top\boldsymbol{\beta}_2$，代入二、三式

$$\begin{cases}X_1{}^\top\left[\mathbf{y}-\mathbf{1}(\bar{y}-\bar{\mathbf{x}}_1{}^\top\boldsymbol{\beta}_1-\bar{\mathbf{x}}_2{}^\top\boldsymbol{\beta}_2)-X_1\boldsymbol{\beta}_1-X_2\boldsymbol{\beta}_2\right]=0\\ X_2{}^\top\left[\mathbf{y}-\mathbf{1}(\bar{y}-\bar{\mathbf{x}}_1{}^\top\boldsymbol{\beta}_1-\bar{\mathbf{x}}_2{}^\top\boldsymbol{\beta}_2)-X_1\boldsymbol{\beta}_1-X_2\boldsymbol{\beta}_2\right]=0\end{cases}$$

整理得：

$$\begin{cases}X_1{}^\top(X_1-\mathbf{1}\bar{\mathbf{x}}_1{}^\top)\boldsymbol{\beta}_1+X_1{}^\top(X_2-\mathbf{1}\bar{\mathbf{x}}_2{}^\top)\boldsymbol{\beta}_2=X_1{}^\top(\mathbf{y}-\mathbf{1}\bar{y})=(X_1-\mathbf{1}\bar{\mathbf{x}}_1{}^\top)^\top\mathbf{y}\\ X_2{}^\top(X_1-\mathbf{1}\bar{\mathbf{x}}_1{}^\top)\boldsymbol{\beta}_1+X_2{}^\top(X_2-\mathbf{1}\bar{\mathbf{x}}_2{}^\top)\boldsymbol{\beta}_2=X_2{}^\top(\mathbf{y}-\mathbf{1}\bar{y})\end{cases}$$

记（中心化）$X_{1c}=X_1-\mathbf{1}\bar{\mathbf{x}}_1^\top, X_{2c}=X_2-\mathbf{1}\bar{\mathbf{x}}_2^\top, \mathbf{y}_c=\mathbf{y}-\mathbf{1}\bar{y}$，

前述方程组简写为

$$\begin{cases} X_{1c}{}^\top X_{1c}\boldsymbol{\beta}_1+X_{1c}{}^\top X_{2c}\boldsymbol{\beta}_2=X_{1c}{}^\top\mathbf{y} \\ X_{2c}{}^\top X_{1c}\boldsymbol{\beta}_1+X_{2c}{}^\top X_{2c}\boldsymbol{\beta}_2=X_{2c}{}^\top\mathbf{y} \end{cases}$$

第一式左乘$X_{2c}{}^\top X_{1c}\left(X_{1c}{}^\top X_{1c}\right)^{-1}$得：

$$X_{2c}{}^\top X_{1c}\boldsymbol{\beta}_1+X_{2c}{}^\top X_{1c}\left(X_{1c}{}^\top X_{1c}\right)^{-1}X_{1c}{}^\top X_{2c}\boldsymbol{\beta}_2=X_{2c}{}^\top X_{1c}\left(X_{1c}{}^\top X_{1c}\right)^{-1}X_{1c}{}^\top\mathbf{y}$$

即$X_{2c}{}^\top X_{1c}\boldsymbol{\beta}_1+X_{2c}{}^\top P_{X_{1c}}X_{2c}\boldsymbol{\beta}_2=X_{2c}{}^\top P_{X_{1c}}\mathbf{y}$，与第二式相减(消去$\boldsymbol{\beta}_1$)得：

$$X_{2c}{}^\top(I-P_{X_{1c}})X_{2c}\boldsymbol{\beta}_2=X_{2c}{}^\top(I-P_{X_{1c}})\mathbf{y},$$

$$\Leftrightarrow X_2^{\perp\top}X_2^{\perp}\boldsymbol{\beta}_2=X_2^{\perp\top}\mathbf{y}\Rightarrow\hat{\boldsymbol{\beta}}_2=\left(X_2^{\perp\top}X_2^{\perp}\right)^{-1}X_2^{\perp\top}\mathbf{y}$$

证明4: 不用中心化、正交化，应用分块矩阵的逆（略）

例2.（命题4的$p=3$情形）假设数据$(y_i, x_i, z_i), i=1,2,\ldots,n$,满足模型：

$$y_i = a + bx_i + cz_i + \varepsilon_i,\quad \varepsilon_i \text{ iid} \sim (0,\sigma^2),\ \varepsilon_i \text{ 与 } x_i, z_i \text{ 独立。}$$

由命题4,

$$\hat{b} = \frac{S_{xy\bullet z}}{S_{xx\bullet z}} = \frac{S_{xy} - S_{xz}S_{yz}/S_{zz}}{S_{xx} - S_{xz}^2/S_{zz}} = \sqrt{\frac{S_{yy}}{S_{xx}}}\left(\frac{r_{yx} - r_{yz}r_{xz}}{1-r_{xz}^2}\right) \propto r_{yx\bullet z},$$

$$\mathrm{var}(\hat{b}\mid x,z) = \frac{\sigma^2}{(n-1)(S_{xx} - S_{xz}^2/S_{zz})} = \frac{\sigma^2}{(n-1)S_{xx}} \times \frac{1}{1-r_{xz}^2},$$

注意$r_{xz}=0$时，方差$\mathrm{var}(\hat{b}\mid x,z) = \dfrac{\sigma^2}{(n-1)S_{xx}}$,

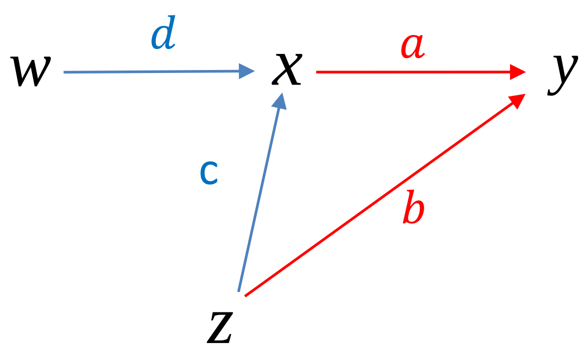
故$VIF \stackrel{\Delta}{=} \dfrac{1}{1-r_{xz}^2}$称为方差膨胀因子(VIF: variance inflation factor)。

路径模型（path model）

路径模型

路径模型(Path model, Sewall Wright, 1921) 是若干个有联系的线性回归模型, 表示变量之间的因果链。 通常路径模型用有向树图表示，箭头旁边的数字是回归系数/效应。

为了避免截距项的干扰，假设所有变量都已中心化或标准化。



$y = ax + bz + \varepsilon$，$\varepsilon \sim (0, \sigma^2)$
$x = cz + dw + \eta$，$\eta \sim (0, \tau^2)$

x是第一个模型的自变量，
是第二个模型的响应变量

基于所有变量的协方差矩阵，可以求出路径模型中所有的参数。

路径模型的指定通常是非常困难的问题。对于有时间次序的问题，路径模型相对容易指定。

例3. 美国社会学家布劳和邓肯(Blau & Duncan,1967；Freedman,P81-86） 研究了美国职业结构和社会阶层分化过程(social stratification)。调查对象是20000名年龄在20-64的男性的职业(分值0-96)和教育程度(分值0-8)。变量包括（按时间次序）：

当前职业Y，第一份工作W，学历U，父亲的职业X，父亲的学历 V

5个变量的样本相关系数如下 。

		Y Son's occ	W Son's 1^{st} job	U Son's ed	X Dad's occ	V Dad's ed
Y	Son's occ	1.000	.541	.596	.405	.322
W	Son's 1^{st} job	.541	1.000	.538	.417	.332
U	Son's ed	.596	.538	1.000	.438	.453
X	Dad's occ	.405	.417	.438	1.000	.516
V	Dad's ed	.322	.332	.453	.516	1.000

按照时间次序，假设中心标准化的5个变量满足Blau-Duncan status attainment路径模型：

学历U 受到父亲的职业X和学历V的影响：$U \sim X + V$
第一份工作W受其学历和父亲的职业影响：$W \sim U + X$
职业Y受W, U, X的影响：$Y \sim W + U + X$

(1) 假设Blau−Duncan模型正确。考虑第二个方程方程

$$W = dU + eX + \delta, \delta\sim(0,\boldsymbol{\tau}^2)$$

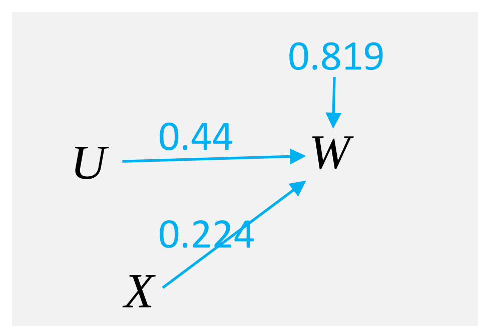
该模型仅涉及响应W, 自变量$\mathbf{z} = (U,X)$, 我们仅需要考虑原始相关系数矩阵的W，U，X对应的行和列

$$\begin{array}{cc} & \begin{array}{ccc} W & U & X \end{array} \\ S = & \begin{pmatrix} 1 & 0.538 & 0.417 \\ 0.538 & 1 & 0.438 \\ 0.417 & 0.438 & 1 \end{pmatrix} \end{array} = \begin{pmatrix} S_{ww} & S_{w\mathbf{z}} \\ S_{\mathbf{z}w} & S_{\mathbf{zz}} \end{pmatrix}, \quad \mathbf{z} = (UX)$$

$$\begin{pmatrix} \hat{d} \\ \hat{e} \end{pmatrix} = S_{\mathbf{zz}}^{-1} S_{\mathbf{z}w} = \begin{pmatrix} 1 & 0.438 \\ 0.438 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0.538 \\ 0.417 \end{pmatrix} = \begin{pmatrix} 0.440 \\ 0.224 \end{pmatrix}$$

$$\hat{\tau}^2 \approx S_{ww\bullet\mathbf{z}} = S_{ww} - S_{w\mathbf{z}} S_{\mathbf{zz}}^{-1} S_{\mathbf{z}w} = 1 - (0.538, 0.417) \begin{pmatrix} 0.440 \\ 0.224 \end{pmatrix} = 0.670, \quad \widehat{\boldsymbol{\tau}} = 0.819$$

$$R^2 = SS_{w\mathbf{z}} S_{\mathbf{zz}}^{-1} S_{\mathbf{z}w} / S_{ww} = 1 - \hat{\sigma}^2 = 0.330$$



(2) $Y = aW + bU + cX + \varepsilon, \varepsilon \sim (0, \sigma^2)$，该模型仅涉及响应$Y$, 自变量$\mathbf{x} = (W, U, X)$, 我们仅需要相关系数矩阵的前4行、列.

$$S = \begin{array}{c} \\ \\ \end{array}\begin{array}{cccc} Y & W & U & X \end{array}$$

$$S = \begin{pmatrix} 1 & 0.541 & 0.596 & 0.405 \\ 0.541 & 1 & 0.538 & 0.417 \\ 0.596 & 0.538 & 1 & 0.438 \\ 0.405 & 0.417 & 0.438 & 1 \end{pmatrix} = \begin{array}{c} \begin{array}{cc} Y & \mathbf{x} = (WUX) \end{array} \\ \begin{pmatrix} S_{yy} & S_{y\mathbf{x}} \\ S_{\mathbf{x}y} & S_{\mathbf{xx}} \end{pmatrix} \end{array}$$

$$\begin{pmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \end{pmatrix} = S_{\mathbf{xx}}^{-1} S_{\mathbf{x}y} = \begin{pmatrix} 1 & 0.538 & 0.417 \\ 0.538 & 1 & 0.438 \\ 0.417 & 0.438 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0.541 \\ 0.596 \\ 0.405 \end{pmatrix} = \begin{pmatrix} 0.281 \\ 0.394 \\ 0.115 \end{pmatrix}$$

$$\hat{\sigma}^2 \approx S_{yy\bullet\mathbf{x}} = S_{yy} - S_{y\mathbf{x}} S_{\mathbf{xx}}^{-1} S_{\mathbf{x}y} = 0.566, \ \hat{\sigma} = 0.753$$

$$R^2 = S_{y\mathbf{x}} S_{\mathbf{xx}}^{-1} S_{\mathbf{x}y} / S_{yy} = 1 - \hat{\sigma}^2 = 0.434$$