4.10 HW 10

作业 10 链接

练习 4.1 对于一般线性模型 $\mathbf{y} = X\beta + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim \left(\mathbf{0}, \sigma^2 I_n\right)$, 我们知道 $\left(I_n - P_X\right) X = 0$, 其中 $P_X = X\left(X^\top X\right)^- X^\top$ 是投影阵, 所以

$$RSS = \mathbf{y}^\top \left(I_n - P_X\right) \mathbf{y} = \boldsymbol{\epsilon}^\top \left(I_n - P_X\right) \boldsymbol{\epsilon},$$

即 RSS 不依赖于 $\boldsymbol{\beta}$ 。基于上述观点重新证明第 6 讲引理 1(2)：对于简单线性模型, 残差平方和 $RSS = s_{\epsilon\epsilon} - s_{\epsilon x}^2 / s_{xx}$ 。

解

简单线性模型

$$y_i = a + bx_i + \varepsilon_i, i = 1, 2, \cdots, n$$

其矩阵-向量形式为 $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, 其中

$$X = \begin{pmatrix} \mathbf{1} & \boldsymbol{x} \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} a \\ b \end{pmatrix}$$

设

$$\tilde{X} = \begin{pmatrix} \mathbf{1} & \frac{1}{\sqrt{s_{xx}}}(\boldsymbol{x} - \mathbf{1}\bar{x}) \end{pmatrix}$$

则 $C(X) = C(\tilde{X})$

$$P_X = P_{\tilde{X}} = \tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T = \left(\frac{1}{n} + \frac{\left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right)}{s_{xx}}\right)_{n \times n}$$

所以

$$\begin{aligned}
RSS &= \varepsilon^T \left(I_n - P_X\right) \varepsilon \\
&= \varepsilon^T \varepsilon - \varepsilon^T P_X \varepsilon \\
&= \sum_{i=1}^n \varepsilon_i^2 - \sum_{i,j=1}^n \varepsilon_i \varepsilon_j \left(\frac{1}{n} + \frac{\left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right)}{s_{xx}}\right) \\
&= \sum_{i=1}^n \varepsilon_i^2 - \sum_{i,j=1}^n \frac{\varepsilon_i \varepsilon_j}{n} - \frac{1}{s_{xx}} \sum_{i,j=1}^n \varepsilon_i \varepsilon_j \left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right) \\
&= \sum_{i=1}^n \varepsilon_i^2 - \frac{1}{n}\left(\sum_{i=1}^n \varepsilon_i\right)^2 - \frac{1}{s_{xx}}\left(\sum_{i=1}^n \varepsilon_i \left(x_i - \bar{x}\right)\right)^2 \\
&= s_{\varepsilon\varepsilon} - \frac{s_{\varepsilon x}^2}{s_{xx}}
\end{aligned}$$

注 注意 $s_{\varepsilon\varepsilon}$ 是中心化的形式!

练习 4.2 假设数据 $\left(y_i, x_i, z_i\right), i = 1, \ldots, n$ 满足线性模型

$$y_i = a + bx_i + cz_i + \epsilon_i, \epsilon_i \sim \left(0, \sigma^2\right), \epsilon_i \text{ 与} x_i, z_i \text{ 独立}$$

或等价地 $\left(\mathbf{x} = \left(x_1, \ldots, x_n\right)^\top, \mathbf{z} = \left(z_1, \ldots, z_n\right)^\top\right)$

$$\mathbf{y} = \mathbb{1}a + \mathbf{x}b + \mathbf{z}c + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \left(0, \sigma^2 I_n\right), \quad \boldsymbol{\epsilon} \perp (\mathbf{x}, \mathbf{z}).$$

(a) 记 $\hat{c}_0 = S_{xz}/S_{zz}, S_{xz} = \frac{1}{n-1}\sum\left(x_i - \bar{x}\right)\left(z_i - \bar{z}\right), S_{zz} = \frac{1}{n-1}\sum\left(z_i - \bar{z}\right)^2$, 试用第 11 讲命题 4 证明 1 的投影方法 (即先将三个向量 $\mathbb{1}, \mathbf{x}, \mathbf{z}$ 正交化, 再投影) 而不是直接应用命题 4 的结论, 证明 b 的 LS 估计

$$\hat{b} = \frac{\sum\left[\left(x_i - \bar{x} - \left(z_i - \bar{z}\right)\hat{c}_0\right] y_i}{\sum\left[\left(x_i - \bar{x} - \left(z_i - \bar{z}\right)\hat{c}_0\right]^2}.$$

(注: 该表达式表明, 为了求解 b 的 LS 估计, 我们可以分两步回归: (1) $\operatorname{lm}(x \sim z): x_i = a_0 + c_0 z_i + \delta_i$, 求得回归系

数 $\hat{c}_0 = s_{xz}/s_{zz}$ 以及 $\hat{a}_0 = \bar{x} - \bar{z}\hat{c}_0$, 得到残差 $x_i^{\perp} = x_i - \bar{x} - (z_i - \bar{z})\,\hat{c}_0$. (2) $y \sim x^{\perp} : y_i = \alpha + \beta x_i^{\perp} +$ error, 由此得到的斜率 LS 估计 $\hat{\beta}$ 就是 $\hat{b}$)

(b) 证明

$$R^2 = \frac{r_{yx}^2 + r_{yz}^2 - 2r_{yx}r_{yz}r_{xz}}{1 - r_{xz}^2},$$

其中 r_{xz} 为两个自变量之间的样本相关系数。

证明 记

$$s_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$$

对 $X = \begin{pmatrix} \mathbf{1} & \boldsymbol{x} & \boldsymbol{z} \end{pmatrix}$ 正交化得到:

于是

$$\hat{\boldsymbol{y}} = P_X \boldsymbol{y} = P_{\tilde{X}} \boldsymbol{y} = \mathbf{1}\hat{a} + \boldsymbol{x}\hat{b} + \boldsymbol{z}\hat{c}$$

关注 x 的系数, 可得

$$\begin{aligned}
\hat{b} &= \left[\frac{\boldsymbol{x}^T - \bar{x}\mathbf{1}^T}{s_{xx}} - \frac{\frac{s_{xz}}{s_{xx}}\left[\boldsymbol{z}^T - \mathbf{1}\bar{z} - \frac{s_{xz}}{s_{xx}}(\boldsymbol{x} - \mathbf{1}\bar{x})^T\right]}{s_{zz} - s_{zx}^2/s_{xx}} \right] \boldsymbol{y} \\
&= \frac{s_{xy}}{s_{xx}} - \frac{s_{xz}\left(s_{zy} - \frac{s_{xz}}{s_{xx}}s_{xy}\right)}{s_{zz}s_{xx} - s_{xz}^2} \\
&= \frac{s_{xy}s_{xx}s_{zz} - s_{xy}s_{xz}^2 - s_{xz}s_{zy}s_{xx} + s_{xy}s_{xz}^2}{s_{xx}\left(s_{zz}s_{xx} - s_{xz}^2\right)} \\
&= \frac{s_{xy}s_{zz} - s_{xz}s_{zy}}{s_{zz}s_{xx} - s_{xz}^2} = \frac{s_{xy} - s_{zy}}{s_{xx} - s_{xz}\hat{c}_0} = \frac{\sum\left[x_i - \bar{x} - (z_i - \bar{z})\,\hat{c}_0\right] y_i}{\sum\left[x_i - \bar{x} - (z_i - \bar{z})\,\hat{c}_0\right]^2}
\end{aligned}$$

同时, x 和 z 地位对等, 交换位置即有

$$\hat{c} = \frac{s_{zy}s_{xx} - s_{xz}s_{xy}}{s_{zz}s_{xx} - s_{xz}^2}$$

考虑 $R^2 = \operatorname{var}(\hat{\boldsymbol{y}})/\operatorname{var}(\boldsymbol{y}) = (n-1)\operatorname{var}(\hat{\boldsymbol{y}})/s_{yy}$,

$$\begin{aligned}
\operatorname{var}(\hat{\boldsymbol{y}}) &= \operatorname{var}(\boldsymbol{x}\hat{b} + \boldsymbol{z}\hat{c}) \\
&= \operatorname{var}(\boldsymbol{x})\hat{b}^2 + \operatorname{var}(\boldsymbol{z})\hat{c}^2 + 2\hat{b}\hat{c}\operatorname{cov}(\boldsymbol{x}, \boldsymbol{z}) \\
&= \frac{1}{n-1}\left(s_{xx}\hat{b}^2 + s_{zz}\hat{c}^2 + 2s_{xz}\hat{b}\hat{c}\right) \\
&= \frac{1}{n-1} \frac{s_{xx}\left(s_{xy}s_{zz} - s_{xz}s_{zy}\right)^2 + s_{zz}\left(s_{zy}s_{xx} - s_{xz}s_{xy}\right)^2 + 2s_{xz}\left(s_{zy}s_{xx} - s_{xz}s_{xy}\right)\left(s_{xy}s_{zz} - s_{xz}s_{zy}\right)}{\left(s_{zz}s_{xx} - s_{xz}^2\right)^2} \\
&= \frac{1}{n-1} \frac{s_{xy}^2 s_{zz} + s_{yz}^2 s_{xx} - 2s_{xy}s_{yz}s_{xz}}{s_{zz}s_{xx} - s_{xz}^2}
\end{aligned}$$

$$R^2 = \frac{s_{xy}^2 s_{zz} + s_{yz}^2 s_{xx} - 2s_{xy}s_{yz}s_{xz}}{\left(s_{zz}s_{xx} - s_{xz}^2\right)s_{yy}}$$

注意到

$$r_{xy}^2 = \frac{s_{xy}^2}{s_{xx}s_{yy}}$$

于是

$$R^2 = \frac{r_{yx}^2 + r_{yz}^2 - 2r_{yx}r_{yz}r_{xz}}{1 - r_{xz}^2}$$

练习 4.3 假设线性模型 $\mathbf{y} = X_1\beta_1 + X_2\beta_2 + \boldsymbol{\epsilon}$, 其中 X_1 为 $n \times (p-q)$ 矩阵, X_2 为 $n \times q$ 矩阵, β_2 为 $q \times 1$ 参数向量。记 $X_2^{\perp} = X_2 - P_{X_1}X_2$ 。假设 W 是任何 $n \times q$ 矩阵使得 $W^{\top}X_1 = 0, W^{\top}X_2$ 可逆, 证明 $\tilde{\beta}_2 = \left(W^{\top}X_2\right)^{-1}W^{\top}\mathbf{y}$ 是 β_2 的无偏估计, 并证明 $\operatorname{var}\left(\tilde{\beta}_2 \mid X\right) \geq \sigma^2\left(X_2^{\perp}X_2^{\perp}\right)^{-1}$, 当 $W = X_2^{\perp}$ 时等号成立。

证明

$$
\begin{aligned}
E\left(\tilde{\beta}_2 \mid X\right) &= \left(W^T X_2\right)^{-1} W^T \left(X_1\beta_1 + X_2\beta_2\right) \\
&= \left(W^T X_2\right)^{-1} W^T X_2\beta_2 = \beta_2
\end{aligned}
$$

因此: $E\left(\tilde{\beta}_2\right) = E\left(E\left(\tilde{\beta}_2 \mid X\right)\right) = \beta_2$.

记 $C = \left(W^T X_2\right)^{-1} W^T$, 则 $CX_2 = I$, 根据投影阵 $P \leqslant I$,

$$
\begin{aligned}
\operatorname{var}\left(\tilde{\beta}_2 \mid X\right) &= \sigma^2 CC^T \\
\left(X_2^{\perp T} X_2^{\perp}\right)^{-1} &= \left(X_2^T \left(I_n - P_{X_1}\right)^T \left(I_n - P_{X_1}\right) X_2\right)^{-1} \\
&= \left(X_2^T \left(I_n - P_{X_1}\right) X_2\right)^{-1} \\
&\leqslant \left(X_2^T X_2\right)^{-1} \\
&= CX_2 \left(X_2^T X_2\right)^{-1} X_2^T C^T \\
&= CP_{X_2} C^T \leqslant CC^T \\
\Rightarrow \operatorname{var}\left(\tilde{\beta}_2 \mid X\right) &\geqslant \sigma^2 \left(X_2^{\perp T} X_2^{\perp}\right)^{-1}
\end{aligned}
$$

✍ **练习 4.4** 假设 A 为任一 $n \times p$ 列满秩矩阵, 其各列为 $A = (\mathbf{a}_1, \ldots, \mathbf{a}_p)$, 我们知道 $A^\top A$ 的 (i,j) 元为 $\mathbf{a}_i^\top \mathbf{a}_j$ 。记 $\mathbf{a}_i^\perp = \mathbf{a}_i - P_{A_{(-i)}}\mathbf{a}_i$, 其中 $A_{(-i)}$ 为 A 删除第 i 列后剩余的列组成的矩阵。证明 $\left(A^\top A\right)^{-1}$ 的 (i,j) 元等于 $\frac{\mathbf{a}_i^{\perp\top}\mathbf{a}_j^\perp}{\|\mathbf{a}_i^\perp\|^2\|\mathbf{a}_j^\perp\|^2}$. 提示: 假设线性模型 $\mathbf{y} = A\beta + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim \left(0, \sigma^2 I_n\right), \beta = (\beta_1, \ldots, \beta_p)^\top$, 注意这里 A 的第一列未必是 $\mathbb{1}$, 但 LS 估计仍具有形式 $\hat{\beta} = \left(A^\top A\right)^{-1} A^\top \mathbf{y}$, 方差矩阵为 $\operatorname{var}(\hat{\beta} \mid A) = \sigma^2 \left(A^\top A\right)^{-1}$, 其 (i,j) 元为 $\hat{\beta}_i, \hat{\beta}_j$ 的协方差, 因此只需计算 $\operatorname{cov}\left(\hat{\beta}_i, \hat{\beta}_j \mid A\right)$ (利用 11 讲命题 4 给出 $\hat{\beta}_i, \hat{\beta}_j$ 的表达).

解

线性模型可以写为: $\boldsymbol{y} = a_1\beta_1 + \cdots + a_p\beta_p + \epsilon$. 由前已知 β_i 的 LS 估计为 $\hat{\beta}_i = \left(a_i^{\perp T} a_i^\perp\right)^{-1} a_i^{\perp T} y$, 于是:

$$
\begin{aligned}
\operatorname{Cov}\left(\hat{\beta}_i, \hat{\beta}_j \mid A\right) &= \left(a_i^{\perp T} a_i^\perp\right)^{-1} a_i^{\perp T} \operatorname{Var}(y \mid A) a_j^\perp \left(a_j^{\perp T} a_j^\perp\right)^{-1} \\
&= \sigma^2 \frac{a_i^{\perp T} a_j^\perp}{\|a_i^\perp\|^2 \, \|a_j^\perp\|^2}
\end{aligned}
$$

所以 $\left(A^T A\right)^{-1}$ 的 (i,j) 元为 $\frac{a_i^{\perp T} a_j^\perp}{\|a_i^\perp\|^2\|a_j^\perp\|^2}$.