# 《高级计量经济学》2024-2025学年期中作业参考答案

1. $\boldsymbol{X}$ 为 $n \times k$ 矩阵，$n > k$。证明 $\text{rank}(\boldsymbol{X}) = k \Leftrightarrow \text{rank}(\boldsymbol{X}^\top \boldsymbol{X}) = k$ 。[30分]

解：

令 $\boldsymbol{Q} = \boldsymbol{X}^\top \boldsymbol{X}$。令 $\boldsymbol{d}$ 为任意 $k \times 1$ 向量，并令 $\boldsymbol{v} = \boldsymbol{X}\boldsymbol{d}$。因此 $\boldsymbol{v}$ 是 $\boldsymbol{X}$ 的列向量的线性结合。$\boldsymbol{v}^\top \boldsymbol{v} = \boldsymbol{d}^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{d} = \boldsymbol{d}^\top \boldsymbol{Q}\boldsymbol{d}$ 是二次型。

(a)

假设 $\text{rank}(\boldsymbol{X}) < k$，则 $\boldsymbol{X}\boldsymbol{d} = \boldsymbol{0}$ 存在非零解 $\boldsymbol{d} \neq \boldsymbol{0}$，此时 $\boldsymbol{v} = \boldsymbol{0}$，$\boldsymbol{Q}\boldsymbol{d} = \boldsymbol{X}^\top \boldsymbol{v} = \boldsymbol{0}$。因此 $\boldsymbol{d}$ 也是 $\boldsymbol{Q}\boldsymbol{d} = \boldsymbol{0}$ 的非零解，因此 $\text{rank}(Q) < k$。

(b)

假设 $\text{rank}(Q) < k$，则 $\boldsymbol{Q}\boldsymbol{d} = \boldsymbol{0}$ 存在非零解 $\boldsymbol{d} \neq \boldsymbol{0}$，此时 $\boldsymbol{v}^\top \boldsymbol{v} = \boldsymbol{d}^\top \boldsymbol{Q}\boldsymbol{d} = \boldsymbol{d}^\top \boldsymbol{0} = 0$。因此 $\boldsymbol{v} = \boldsymbol{0}$，即 $\boldsymbol{d}$ 也是 $\boldsymbol{X}\boldsymbol{d} = \boldsymbol{0}$ 的非零解。因此 $\text{rank}(\boldsymbol{X}) < k$。

由 (a) 和 (b) 可得 $\text{rank}(\boldsymbol{X}) < k \Leftrightarrow \text{rank}(\boldsymbol{X}^\top \boldsymbol{X}) < k$。因为 $\text{rank}(\boldsymbol{X}) \leq k$，$\text{rank}(\boldsymbol{Q}) \leq k$，可得 $\text{rank}(\boldsymbol{X}) = k \Leftrightarrow \text{rank}(\boldsymbol{X}^\top \boldsymbol{X}) = k$。

2. 证明下面的命题 [40分]

a. 若 $\boldsymbol{A}$ 是 $n \times n$ 矩阵，$\lambda_1, \cdots, \lambda_n$ 是 $\boldsymbol{A}$ 的 $n$ 个特征值，则 $\text{tr}(A) = \sum_{i=1}^{n} \lambda_i$
b. 投影矩阵 $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$ 的 $n$ 个特征值中有 $k$ 个为 $1$，$n - k$ 个为 $0$。

解：

a. $\boldsymbol{A}$ 的特征多项式是

$$\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = \begin{vmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{12} - \lambda & \dots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{12} & \dots & a_{1n} - \lambda \end{vmatrix}$$

沿第一列展开，可得

$$\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = (a_{11} - \lambda) \det(\boldsymbol{M}_{11}) + \sum_{i=2}^{n} a_{i1}(-1)^{i+1} \det(\boldsymbol{M}_{i1})$$

且余子式 $\det(\boldsymbol{M}_{i1}), i=2,\ldots,n$ 中不包含 $(a_{11}-\lambda)$ 和 $(a_{ii}-\lambda)$。如果我们用同样的方法展开 $\det(\boldsymbol{M}_{11})$ 并重复，可知 $\det(\boldsymbol{A}-\lambda\boldsymbol{I})$ 的展开式中唯一包含大于 $n-2$ 个对角成分的项是

$$(a_{11}-\lambda)(a_{22}-\lambda)\cdot\cdots\cdot(a_{nn}-\lambda) \qquad \text{......} (\#)$$

将 (#) 式展开可知 $\lambda^n$ 项的系数是 $(-1)^n$。因此，$\det(\boldsymbol{A}-\lambda\boldsymbol{I})$ 的最高次项的系数是 $(-1)^n$。当 $\boldsymbol{A}$ 的特征值为 $\lambda_1,\cdots,\lambda_n$ 时，

$$\begin{aligned}\det(\boldsymbol{A}-\lambda\boldsymbol{I}) &= (-1)^n(\lambda-\lambda_1)(\lambda-\lambda_2)\cdot\cdots\cdot(\lambda-\lambda_n) \\ &= (\lambda_1-\lambda)(\lambda_2-\lambda)\cdot\cdots\cdot(\lambda_n-\lambda)\end{aligned} \qquad \text{......} (*)$$

由 (#) 式还可知 $(-\lambda)^{n-1}$ 项的系数是 $\sum_{i=1}^n a_{ii}$。若将 (*) 式展开，则可得 $(-\lambda)^{n-1}$ 项的系数是 $\sum_{i=1}^n \lambda_i$。因此 $\mathrm{tr}(\boldsymbol{A})=\sum_{i=1}^n \lambda_i$ 。

b. $\boldsymbol{P}$ 是幂等矩阵，根据谱分解可知 $\boldsymbol{P}=\boldsymbol{H\Lambda H}^\top$，由此可得

$$\boldsymbol{P}=\boldsymbol{PP}=\boldsymbol{H\Lambda H}^\top\boldsymbol{H\Lambda H}^\top=\boldsymbol{H\Lambda}^2\boldsymbol{H}^\top$$

因此，$\boldsymbol{\Lambda}^2=\boldsymbol{\Lambda}$，即 $\lambda_i^2=\lambda_i, i=1,\ldots,n$。因此 $\lambda_i$ 的值只能是 1 或 0。由 (a) 的结果可知 $\mathrm{tr}(\boldsymbol{P})=\sum_{i=1}^n \lambda_i$。同时，

$$\mathrm{tr}(\boldsymbol{P})=\mathrm{tr}(\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)=\mathrm{tr}((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X})=\mathrm{tr}(\boldsymbol{I}_k)=k$$

因此 $\sum_{i=1}^n \lambda_i=k$。由于特征值只能是 1 或者 0，可得 $\boldsymbol{P}$ 有 $k$ 个特征值为 $1$，$n-k$ 个特征值为 $0$。

3. 什么是“多重共线性问题”？多重共线性对线性模型系数的 OLS 估计量有什么影响？为什么说样本量过小也会带来相同的问题？ [30分]

参考答案提纲：

需要区分准确多重共线性（exact multicollinearity）和近似多重共线性（approximate/near multicollinearity）。存在准确多重共线性时不存在 OLS 解，即 $\boldsymbol{X}$ 不满秩，因此 $(\boldsymbol{X}^\top\boldsymbol{X})$ 不可逆。近似多重共线性会导致系数估计量的方差过大，估计精确度降低。

考虑回归模型 $\boldsymbol{y}=\boldsymbol{X\beta}+\boldsymbol{u}=\boldsymbol{x}_1\beta_1+\boldsymbol{X}_2\boldsymbol{\beta}_2+\boldsymbol{u}$，$E(\boldsymbol{uu}^\top)=\sigma^2\boldsymbol{I}$。根据 FWL 定理，$\widehat{\beta}_1$ 可以从下面的残差回归模型中获得，

$$\boldsymbol{M}_2\boldsymbol{y}=\boldsymbol{M}_2\boldsymbol{x}_1\beta_1+\boldsymbol{v}$$

因此，

$$\widehat{\beta}_1=\frac{\boldsymbol{x}_1^\top\boldsymbol{M}_2\boldsymbol{y}}{\boldsymbol{x}_1^\top\boldsymbol{M}_2\boldsymbol{x}_1}$$

参照 $\widehat{\boldsymbol{\beta}}$ 的方差公式 $\mathrm{var}(\widehat{\boldsymbol{\beta}})=\sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$，$\widehat{\beta}_1$ 的方差可以表达为

$$
\operatorname{var}(\widehat{\beta}_1) = \frac{\sigma^2}{\boldsymbol{x}_1^\top \boldsymbol{M}_2 \boldsymbol{x}_1} = \frac{\sigma^2}{(\boldsymbol{M}_2 \boldsymbol{x}_1)^\top (\boldsymbol{M}_2 \boldsymbol{x}_1)}
$$

分母为向量 $\boldsymbol{M}_2 \boldsymbol{x}_1$ 的长度的平方，也是回归模型 $\boldsymbol{x}_1 = \boldsymbol{X}_2 \boldsymbol{b}_2 + \boldsymbol{w}$ 的 OLS 残差平方和。该残差平方和越小，意味着所有的残差项都很小，即 $\boldsymbol{X}_2$ 中的变量对 $\boldsymbol{x}_1$ 的解释能力很强，也就意味着存在近似多重共线性。

存在近似多重共线性（程度高），或者样本量很小时，$\operatorname{var}(\widehat{\beta}_1)$ 的取值会变大，代表 $\widehat{\beta}_1$ 的估计精确度变小；不存在近似多重共线性（程度低），或者样本量很大时，$\operatorname{var}(\widehat{\beta}_1)$ 的取值会变小，代表 $\widehat{\beta}_1$ 的估计精确度变大。

注：由以上讨论可知，多重共线性问题并不是回归分析中的主要问题，其影响和小样本的影响相似，属于样本自身的特征。