线性回归

Herbert002

2021 年 11 月 29 日

摘要

尽量通俗易懂，且不学院气质的线性回归笔记。

目录

1 线性回归的通俗理解

线性回归的通俗理解.

2 线性回归的简要历史

高尔顿爵士关于身高体重的实验.

3 一元线性回归

我们先考虑一元线性回归.

4 多元线性回归

多元线性回归是指回归变量个数大于等于 2 的线性回归.
回归平方和服从卡方分布的证明

$$
\begin{aligned}
SSR &= \sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2 \\
&= [\hat{\boldsymbol{y}}-\mathbf{1}\overline{y}]^T[\hat{\boldsymbol{y}}-\mathbf{1}\overline{y}] \\
&= [\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}-\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T\boldsymbol{y}]^T[\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}-\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T\boldsymbol{y}] \\
&= \boldsymbol{y}^T[\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T-\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T]^T[\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}-\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T]\boldsymbol{y}
\end{aligned} \tag{1}
$$

上式中，$\mathbf{1}\in\mathbb{R}^n$，是一个元素均为 1 的列向量. 响应变量均值可以表示为 $\overline{y}=(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T\boldsymbol{y}$.

接下来证明 $\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T-\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T$ 是对称矩阵.

$$
\begin{aligned}
[\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T-\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T]^T &= [\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T]^T-[\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T]^T \\
&= \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T-\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T
\end{aligned} \tag{2}
$$

得证.

证明 $\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T-\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T$ 是幂等的.

如果说矩阵 A 是幂等的，则有 $\boldsymbol{A}=\boldsymbol{A}\boldsymbol{A}=\cdots=\boldsymbol{A}^n$

证明 $\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T-\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T$ 的幂等性，需要先证明如下结论.

$$
\begin{gathered}
\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\mathbf{1}=\mathbf{1} \\
\mathbf{1}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T=\mathbf{1}^T
\end{gathered} \tag{3}
$$

$\boldsymbol{X}\in\mathbb{R}^{n\times(k+1)}$ 可以分解为 $\boldsymbol{X}=[\mathbf{1}\ \boldsymbol{X}_R]$ ，其中 $\boldsymbol{X}_R\in\mathbb{R}^{n\times k}$ ，则有

$$
\boldsymbol{X}^T\boldsymbol{X}=\begin{bmatrix}\mathbf{1}^T\\ \boldsymbol{X}_R^T\end{bmatrix}\begin{bmatrix}\mathbf{1} & \boldsymbol{X}_R\end{bmatrix}=\begin{bmatrix}\mathbf{1}^T\mathbf{1} & \mathbf{1}^T\boldsymbol{X}_R\\ \boldsymbol{X}_R^T\mathbf{1} & \boldsymbol{X}_R^T\boldsymbol{X}_R\end{bmatrix} \tag{4}
$$

显然，$\boldsymbol{X}^T\boldsymbol{X}$ 是对称矩阵，其逆矩阵也是对称矩阵，假设其为

$$
(\boldsymbol{X}^T\boldsymbol{X})^{-1}=\begin{bmatrix}a & \boldsymbol{b}^T\\ \boldsymbol{b} & \boldsymbol{C}\end{bmatrix} \tag{5}
$$

其中，$a\in\mathbb{R}$ 是个标量，$\boldsymbol{b}\in\mathbb{R}^k$ ，$\boldsymbol{C}\in\mathbb{R}^{k\times k}$ ，则有

$$
\begin{aligned}
\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1} &= \begin{bmatrix}\mathbf{1}^T\mathbf{1} & \mathbf{1}^T\boldsymbol{X}_R\\ \boldsymbol{X}_R^T\mathbf{1} & \boldsymbol{X}_R^T\boldsymbol{X}_R\end{bmatrix}\begin{bmatrix}a & \boldsymbol{b}^T\\ \boldsymbol{b} & \boldsymbol{C}\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{1}^T\mathbf{1}a+\mathbf{1}^T\boldsymbol{X}_R\boldsymbol{b} & \mathbf{1}^T\mathbf{1}\boldsymbol{b}^T+\mathbf{1}^T\boldsymbol{X}_R\boldsymbol{C}\\ \boldsymbol{X}_R^T\mathbf{1}a+\boldsymbol{X}_R^T\boldsymbol{X}_R\boldsymbol{b} & \boldsymbol{X}_R^T\mathbf{1}\boldsymbol{b}^T+\boldsymbol{X}_R^T\boldsymbol{X}_R\boldsymbol{C}\end{bmatrix} \\
&= \begin{bmatrix}1 & \mathbf{0}^T\\ \mathbf{0} & \boldsymbol{I}\end{bmatrix}
\end{aligned} \tag{6}
$$

上式中 $\mathbf{0}\in\mathbb{R}^n$ ，$\boldsymbol{I}\in\mathbb{R}^{k\times k}$ 是单位矩阵. 观察最后一个等号前后矩阵的第一行，显然有

$$
\begin{gathered}
\mathbf{1}^T\mathbf{1}a+\mathbf{1}^T\boldsymbol{X}_R\boldsymbol{b}=1 \\
\mathbf{1}^T\mathbf{1}\boldsymbol{b}^T+\mathbf{1}^T\boldsymbol{X}_R\boldsymbol{C}=\mathbf{0}^T
\end{gathered} \tag{7}
$$

再看 $\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ ，如下

$$
\begin{aligned}
\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\mathbf{1} &= \begin{bmatrix}\mathbf{1} \ \boldsymbol{X}_R\end{bmatrix}\begin{bmatrix}a & \boldsymbol{b}^T \\ \boldsymbol{b} & \boldsymbol{C}\end{bmatrix}\begin{bmatrix}\mathbf{1}^T \\ \boldsymbol{X}_R^T\end{bmatrix}\mathbf{1} \\
&= \begin{bmatrix}a\mathbf{1}+\boldsymbol{X}_R\boldsymbol{b} & \mathbf{1}\boldsymbol{b}^T+\boldsymbol{X}_R\boldsymbol{C}\end{bmatrix}\begin{bmatrix}\mathbf{1}^T \\ \boldsymbol{X}_R^T\end{bmatrix}\mathbf{1} \\
&= \begin{bmatrix}a\mathbf{1}\mathbf{1}^T+\boldsymbol{X}_R\boldsymbol{b}\mathbf{1}^T+\mathbf{1}\boldsymbol{b}^T\boldsymbol{X}_R^T+\boldsymbol{X}_R\boldsymbol{C}\boldsymbol{X}_R^T\end{bmatrix}\mathbf{1} \\
&= a\mathbf{1}\mathbf{1}^T\mathbf{1}+\boldsymbol{X}_R\boldsymbol{b}\mathbf{1}^T\mathbf{1}+\mathbf{1}\boldsymbol{b}^T\boldsymbol{X}_R^T\mathbf{1}+\boldsymbol{X}_R\boldsymbol{C}\boldsymbol{X}_R^T\mathbf{1}
\end{aligned} \tag{8}
$$

观察上式最后一个等号的右边，第一项中 $\mathbf{1}^T\mathbf{1}\in\mathbb{R}$ ，第三项中 $\boldsymbol{b}^T\boldsymbol{X}_R^T\mathbf{1}$ 都是标量，标量的转置是其本身，再结合 (7) 式，第一项和第三项的和

$$
\begin{aligned}
a\mathbf{1}\mathbf{1}^T\mathbf{1}+\mathbf{1}\boldsymbol{b}^T\boldsymbol{X}_R^T\mathbf{1} &= a(\mathbf{1}^T\mathbf{1})\mathbf{1}+(\boldsymbol{b}^T\boldsymbol{X}_R^T\mathbf{1})\mathbf{1} \\
&= a(\mathbf{1}^T\mathbf{1})\mathbf{1}+(\mathbf{1}^T\boldsymbol{X}_R\boldsymbol{b})\mathbf{1} \\
&= [a(\mathbf{1}^T\mathbf{1})+(\mathbf{1}^T\boldsymbol{X}_R\boldsymbol{b})]\mathbf{1} \\
&= \mathbf{1}
\end{aligned} \tag{9}
$$

结合 (7) 式，再看第二项和第四项的和

$$
\begin{aligned}
\boldsymbol{X}_R\boldsymbol{b}\mathbf{1}^T\mathbf{1}+\boldsymbol{X}_R\boldsymbol{C}\boldsymbol{X}_R^T\mathbf{1} &= [[\boldsymbol{X}_R\boldsymbol{b}\mathbf{1}^T\mathbf{1}+\boldsymbol{X}_R\boldsymbol{C}\boldsymbol{X}_R^T\mathbf{1}]^T]^T \\
&= [\mathbf{1}^T\mathbf{1}\boldsymbol{b}^T\boldsymbol{X}_R^T+\mathbf{1}^T\boldsymbol{X}_R\boldsymbol{C}^T\boldsymbol{X}_R^T]^T \\
&= [(\mathbf{1}^T\mathbf{1}\boldsymbol{b}^T+\mathbf{1}^T\boldsymbol{X}_R\boldsymbol{C}^T)\boldsymbol{X}_R^T]^T \\
&= [\mathbf{0}^T\boldsymbol{X}_R^T]^T \\
&= \mathbf{0}
\end{aligned} \tag{10}
$$

回看 (8) 式，可得 $\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\mathbf{1}=\mathbf{1}$ ，同理可得 $\mathbf{1}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T=\mathbf{1}^T$.