线性回归

Herbert002

2021 年 12 月 1 日

摘要

尽量通俗易懂，基于理工科本科数学水平的线性回归笔记.

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

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

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.

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

$\boldsymbol{X} \in \mathbb{R}^{n\times(k+1)}$ 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

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

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

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

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 $\mathbf{1}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T=\mathbf{1}^T$. 推导过程中，需要注意矩阵和向量的维数.

接下来就有

$$
\begin{aligned}
&[\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T-\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T]^2 \\
=& [\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T-\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T][\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T-\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T] \\
=& \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T-\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T \\
&-\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T+\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T \\
=& \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T-\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T-\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T+\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T \\
=& \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T-\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T
\end{aligned} \tag{11}
$$

结合 (2) 式的结论，(1) 式可继续推导为

$$
\begin{aligned}
SSR &= \sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2 \\
&= \boldsymbol{y}^T[\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T-\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T]^T[\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}-\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T]\boldsymbol{y} \\
&= \boldsymbol{y}^T[\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T-\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T][\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T-\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T]\boldsymbol{y} \\
&= \boldsymbol{y}^T[\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T-\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T]\boldsymbol{y}
\end{aligned} \tag{12}
$$

至此，上面已经证明了 $\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T-\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T$ 是一个幂等的实对称矩阵. 接下来，将其简单表示为 $\boldsymbol{Q}$ ，则有

$$SSR=\boldsymbol{y}^T\boldsymbol{Q}\boldsymbol{y} \tag{13}$$

因为 $\boldsymbol{Q}\in\mathbb{R}^{n\times n}$ 是实对称矩阵，所以其可以正交对角化，即存在正交矩阵 $\boldsymbol{P}\in\mathbb{R}^{n\times n}$ ，使得

$$\boldsymbol{P}^{-1}\boldsymbol{Q}\boldsymbol{P}=\boldsymbol{P}^T\boldsymbol{Q}\boldsymbol{P}=\boldsymbol{\Lambda} \tag{14}$$

其中，$\boldsymbol{\Lambda}\in\mathbb{R}^{n\times n}$ 是实对角矩阵，对角线上的元素是 $\boldsymbol{Q}$ 的 n 个特征值. 所以有

$$\begin{aligned}SSR&=\boldsymbol{y}^T\boldsymbol{Q}\boldsymbol{y}\\&=\boldsymbol{y}^T\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T\boldsymbol{y}\\&=(\boldsymbol{P}^T\boldsymbol{y})^T\boldsymbol{\Lambda}(\boldsymbol{P}^T\boldsymbol{y})\end{aligned} \tag{15}$$

又因为 $\boldsymbol{Q}$ 是等幂矩阵，所以有

$$\begin{aligned}\boldsymbol{\Lambda}^2&=\boldsymbol{P}^T\boldsymbol{Q}\boldsymbol{P}\boldsymbol{P}^T\boldsymbol{Q}\boldsymbol{P}\\&=\boldsymbol{P}^T\boldsymbol{Q}^2\boldsymbol{P}\\&=\boldsymbol{P}^T\boldsymbol{Q}\boldsymbol{P}\\&=\boldsymbol{\Lambda}\end{aligned} \tag{16}$$

可见 $\boldsymbol{\Lambda}$ 也是等幂的. 如果一个对角矩阵是等幂的，其对角元素要么是 1，要么是 0（可自行证明）.

用 p_{ij} 表示 $\boldsymbol{P}$ 第 i 行第 j 列的元素. $\boldsymbol{P}^T\boldsymbol{y}\in\mathbb{R}^n$ 的第 i 个元素可以表示为 $\sum_{j=1}^n p_{ji}y_j$ ，带入 (15) 式，可得

$$\begin{aligned}SSR&=(\boldsymbol{P}^T\boldsymbol{y})^T\boldsymbol{\Lambda}(\boldsymbol{P}^T\boldsymbol{y})\\&=\begin{bmatrix}\sum_{j=1}^n p_{j1}y_j & \sum_{j=1}^n p_{j2}y_j \cdots \sum_{j=1}^n p_{jn}y_j\end{bmatrix}\begin{bmatrix}\lambda_1 &&&\\&\lambda_2&&\\&&\ddots&\\&&&\lambda_n\end{bmatrix}\begin{bmatrix}\sum_{j=1}^n p_{j1}y_j\\\sum_{j=1}^n p_{j2}y_j\\\vdots\\\sum_{j=1}^n p_{jn}y_j\end{bmatrix}\\&=\sum_{i=1}^n\lambda_i(\sum_{j=1}^n p_{ji}y_j)^2\\&=\sum_{i=1}^m(\sum_{j=1}^n p_{ji}y_j)^2\end{aligned} \tag{17}$$

上式中，m 是对角矩阵 $\boldsymbol{\Lambda}$ 对角元素为 1 的个数.

当 $\boldsymbol{\beta}=\mathbf{0}$ 时，$\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}=\boldsymbol{\epsilon}$ ，则有

$$SSR=\sum_{i=1}^m(\sum_{j=1}^n p_{ji}y_j)^2 \quad =\sum_{i=1}^m(\sum_{j=1}^n p_{ji}\epsilon_j)^2 \tag{18}$$

则有

$$
\begin{aligned}
&\epsilon_j \sim \mathcal{N}(0, \sigma^2) \\
&\Downarrow \\
&p_{ji}\epsilon_j \sim \mathcal{N}(0, p_{ji}^2\sigma^2) \\
&\Downarrow \\
&\sum_{j=1}^{n} p_{ji}\epsilon_j \sim \mathcal{N}(0, \sum_{i=1}^{n} p_{ji}^2\sigma^2) \\
&\Downarrow \\
&\sum_{j=1}^{n} p_{ji}\epsilon_j \sim \mathcal{N}(0, \sigma^2\sum_{i=1}^{n} p_{ji}^2) \\
&\Downarrow \quad (\sum_{i=1}^{n} p_{ji}^2 = 1) \\
&\sum_{j=1}^{n} p_{ji}\epsilon_j \sim \mathcal{N}(0, \sigma^2) \\
&\Downarrow \\
&\frac{1}{\sigma}\sum_{j=1}^{n} p_{ji}\epsilon_j \sim \mathcal{N}(0, 1) \\
&\Downarrow \\
&\sum_{i=1}^{m} (\frac{1}{\sigma}\sum_{j=1}^{n} p_{ji}\epsilon_j)^2 \sim \chi^2(m) \\
&\Downarrow \\
&\frac{1}{\sigma^2}SSR \sim \chi^2(m)
\end{aligned}
\tag{19}
$$

上面倒数第二个推导 ($\Downarrow$) 需要证明 $\frac{1}{\sigma}\sum_{j=1}^{n} p_{ji}\epsilon_j$ 和 $\frac{1}{\sigma}\sum_{j=1}^{n} p_{jl}\epsilon_j$ ($i \neq l$) 之间的独立性，如下

$$
\begin{aligned}
Cov(\frac{1}{\sigma}\sum_{j=1}^{n} p_{ji}\epsilon_j, \frac{1}{\sigma}\sum_{j=1}^{n} p_{jl}\epsilon_j) &= \frac{1}{\sigma^2}Cov(\sum_{j=1}^{n} p_{ji}\epsilon_j, \sum_{j=1}^{n} p_{jl}\epsilon_j) \\
&= \frac{1}{\sigma^2}\sum_{j=1}^{n}\sum_{h=1}^{n} Cov(p_{ji}\epsilon_j, p_{hl}\epsilon_h) \\
&= \frac{1}{\sigma^2}\sum_{j=1}^{n}\sum_{h=1}^{n} p_{ji}p_{hl}Cov(\epsilon_j, \epsilon_h)
\end{aligned}
\tag{20}
$$

上式中，$j \neq h$ 时，ϵ_j 和 ϵ_h 相互独立，所以 $Cov(\epsilon_j, \epsilon_h) = 0$ ；$j = h$ 时，$Cov(\epsilon_j, \epsilon_j) = Var(\epsilon_j) = \sigma^2$. 此时有

$$
\begin{aligned}
Cov(\frac{1}{\sigma}\sum_{j=1}^{n}p_{ji}\epsilon_j, \frac{1}{\sigma}\sum_{j=1}^{n}p_{jl}\epsilon_j) &= \frac{1}{\sigma^2}\sum_{j=1}^{n}\sum_{h=1}^{n}p_{ji}p_{hl}Cov(\epsilon_j, \epsilon_h) \\
&= \frac{1}{\sigma^2}\sum_{j=1}^{n}p_{ji}p_{jl}Cov(\epsilon_j, \epsilon_j) \\
&= \sum_{j=1}^{n}p_{ji}p_{jl}
\end{aligned} \tag{21}
$$

继续推导，观察上式最后一个等号的后面，它是正交矩阵 $\boldsymbol{\Lambda}$ 第 i 列和第 l 列的点积. 正交矩阵任意两列的点积为 0（可自行证明）. 所以

$$
Cov(\frac{1}{\sigma}\sum_{j=1}^{n}p_{ji}\epsilon_j, \frac{1}{\sigma}\sum_{j=1}^{n}p_{jl}\epsilon_j) = \sum_{j=1}^{n}p_{ji}p_{jl} = 0 \tag{22}
$$

协方差为 0 的两个正态分布，相互独立. 独立性得证.

最后，来看看上面这个卡方分布的自由度 m ，到底是多少. 上面已经说过，是对角矩阵 $\boldsymbol{\Lambda}$ 的对角元素要么是 1，要么是 0，m 是 1 的个数，也就是说，m 也是对角矩阵 $\boldsymbol{\Lambda}$ 的迹，所以

$$
\begin{aligned}
m &= trace(\boldsymbol{\Lambda}) \\
&= trace(\boldsymbol{P}^T\boldsymbol{Q}\boldsymbol{P}) \\
&= trace(\boldsymbol{P}\boldsymbol{P}^T\boldsymbol{Q}) \\
&= trace(\boldsymbol{Q}) \\
&= trace[\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T - \mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T] \\
&= trace[\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T] - trace[\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T] \\
&= trace[\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}] - trace[\mathbf{1}^T\mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}] \\
&= trace(\boldsymbol{I}_{k+1}) - trace(\boldsymbol{I}_1) \\
&= k + 1 - 1 \\
&= k
\end{aligned} \tag{23}
$$

上式中的 k 就是回归模型中回归变量的个数. 推导过程中使用到了矩阵迹的性质。

到此，我们已经基本完整地证明了 $\frac{1}{\sigma^2}SSR$ 服从自由度为 k 的卡方分布，用数学表达式表示为

$$
\frac{1}{\sigma^2}SSR \sim \chi^2(k) \tag{24}
$$

5 附录

定理一

实对称矩阵 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ 必定正交相似于实对角矩阵 $\boldsymbol{\Lambda}$ ，即存在正交矩阵 $\boldsymbol{P}$ ，使得 $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}^T\boldsymbol{A}\boldsymbol{P} = \boldsymbol{\Lambda}$ ，其中 $\boldsymbol{\Lambda}$ 对角线上的元素是 $\boldsymbol{A}$ 的 n 个特征值.

证明可请参考线性代数的教材.

定理二

矩阵迹的运算有如下性质：

$$
\begin{aligned}
&trace(\boldsymbol{AB}) = trace(\boldsymbol{BA}) \\
&trace(\boldsymbol{A}+\boldsymbol{B}) = trace(\boldsymbol{A}) + trace(\boldsymbol{B}) \\
&trace(c\boldsymbol{A}) = c \cdot trace(\boldsymbol{A})
\end{aligned} \tag{25}
$$