Multiple Linear Regression

TOC

    - R_a^2（Adjusted R^2，调整后判定系数）
- Hypothesis Testing（假设检验）
    - $H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$
    - $H_0^* : \beta_i = 0$
- 多因素回归的系数约束检验
    - 交互效应模型转换
        - 一般约束形式 $C\beta = 0_m$
    - Extra Sum of Square(for reduced model)
        - 全模型设定
        - 模型拆分与约简模型
        - 约简模型的平方和与额外平方和推导
    - 回至主线：带约束的模型检验
        - 约束假设与参数拆分
            - 为什么可以这么拆分
        - 约简模型
        - 平方和定义（约简模型与全模型对比）
    - 约束假设下的平方和差异检验
        - 检验逻辑
        - 平方和差异的定义与模型设定
        - 带约束的最小二乘估计推导（拉格朗日乘数法）
    - 约束假设下的 F 检验推导
        - 检验逻辑与前提
        - 平方和差异的分布分析
        - 均方误差与 F 统计量构造
        - 决策规则
        - remark
    - 约束假设检验的实例推导
        - 例 1：单个回归系数的显著性检验
        - 例 2：多个回归系数的联合显著性检验

前面介绍了一元的线性回归模型，现介绍多元模型

模型基础

observations $(y_i, x_{i1}, x_{i2}, \ldots, x_{ik})$，$i = 1, 2, \ldots, n$

x_{ij} 表示第 i 次观测到的第 j 个解释变量

Model：$y_i = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{ik}\beta_k + \varepsilon_i$

其中，$E(\varepsilon_i) = 0$，$\mathrm{Var}(\varepsilon_i) = \sigma^2$，$\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$

用矩阵形式表现则为：

$$Y = \beta_0 \mathbf{1}_n + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$$

$$= [\mathbf{1}_n, X_1, X_2, \ldots, X_k] \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \varepsilon$$

$$= X\beta + \varepsilon$$

其中，$X = [\mathbf{1}_n, X_1, X_2, \ldots, X_k]$

且有 $E(Y) = X\beta$，$\mathrm{Cov}(Y) = \sigma^2 I_n$

可以知道 X 的每一行是一个观测数据，每一列是一个解释变量的取值

Estimator of Parameters

LSE（最小二乘估计）

根据向量的一些基本定理我们可以得到

$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_k x_{ik} = [1, x_{i1}, x_{i2}, \ldots, x_{ik}]\,\hat{\beta} = x_i^T \hat{\beta}$（其中 x_i^T 指 X 中第 i 行）

所以我们有 $\hat{Y} = X\hat{\beta}$

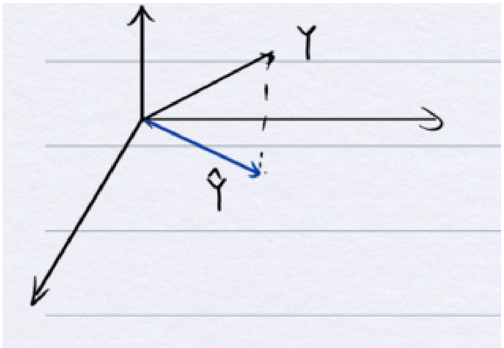
$$\hat{\beta} = \arg\min e_i^2 = \arg\min \|Y - \hat{Y}\|^2 = \arg\min \|Y - X\beta\|^2$$

对 β 求导并令其为 0，有 $-X^T(Y - X\beta) = 0$，则 $\hat{\beta}_{\text{LSE}} = (X^TX)^{-1}X^TY$

$$\hat{Y} = (\mathbf{1}_n, X_1, X_2, \ldots, X_k)\hat{\beta} \in \text{span}\{\mathbf{1}_n, X_1, \ldots, X_k\}$$

从几何投影角度，$\hat{Y} = HY = X(X^TX)^{-1}X^TY = X\hat{\beta}_{\text{LSE}}$（其中 H 为投影矩阵）

投影矩阵 H 的性质



H 称为投影矩阵，性质良好，现说明如下：

① **Symmetric（对称）**
$H^T = H$
验证：

$$\left(X(X^TX)^{-1}X^T\right)^T = X\left((X^TX)^{-1}\right)^T X^T = X\left((X^TX)^T\right)^{-1} X^T = X(X^TX)^{-1}X^T$$

② **Idempotent（幂等）**
$H^2 = H$
验证：

$$\left(X(X^TX)^{-1}X^T\right)^2 = X(X^TX)^{-1}X^TX(X^TX)^{-1}X^T = X(X^TX)^{-1}X^T$$

③ **由对称和幂等性质可得**
H 的特征值只有 0 和 1。
因此，迹 $\text{tr}(H) = \text{tr}\left(X(X^TX)^{-1}X^T\right) = \text{tr}\left((X^TX)^{-1}X^TX\right) = \text{tr}(I_{k+1}) = k + 1$。
故 H 的特征值 1 的几何重数为 $k + 1$，特征值 0 的几何重数为 $n - k - 1$。

④ **投影（后续可能用到）**

$HX = X$，$H1_n = 1_n$

BLUE（最佳线性无偏估计）

下证从 **BLUE** 角度出发，得到的 $\hat{\beta}_{\text{BLUE}}$ 与极大似然估计 $\hat{\beta}_{\text{MLE}}$ 一致。

设 $\tilde{\beta} = \left[(X^TX)^{-1}X^T + \Delta\right]Y$，其中 Δ 是待分析的矩阵。

要使 $\tilde{\beta}$ 是无偏估计，需满足 $E(\tilde{\beta}) = \beta$。

代入 $Y = X\beta + \varepsilon$（其中 $E(\varepsilon) = 0$），得:

$$
\begin{aligned}
E(\tilde{\beta}) &= \left[(X^TX)^{-1}X^T + \Delta\right]E(Y) \\
&= \left[(X^TX)^{-1}X^T + \Delta\right](X\beta) \\
&= \beta + \Delta X\beta
\end{aligned}
$$

令 $E(\tilde{\beta}) = \beta$，需满足 $\Delta X = 0$。

已知 $\text{Cov}(Y) = \sigma^2 I_n$，则:

$$
\begin{aligned}
\text{Cov}(\tilde{\beta}) &= \left[(X^TX)^{-1}X^T + \Delta\right]\text{Cov}(Y)\left[(X^TX)^{-1}X^T + \Delta\right]^T \\
&= \sigma^2\left[(X^TX)^{-1}X^T + \Delta\right]\left[X(X^TX)^{-1} + \Delta^T\right]
\end{aligned}
$$

结合 $\Delta X = 0$ 展开并化简:

$$
\begin{aligned}
\text{Cov}(\tilde{\beta}) &= \sigma^2\left[(X^TX)^{-1} + \Delta\Delta^T + \Delta X(X^TX)^{-1} + (X^TX)^{-1}(X\Delta)^T\right] \\
&= \sigma^2\left[(X^TX)^{-1} + \Delta\Delta^T\right]
\end{aligned}
$$

由于 $\Delta\Delta^T$ 是**半正定矩阵**，根据半正定矩阵的性质，有:

$$
\text{Cov}(\tilde{\beta}) \succeq \sigma^2(X^TX)^{-1}
$$

且等号成立的充要条件是 $\Delta = 0$（若 $\Delta \neq 0$，则存在 i, j 使得 $[\Delta\Delta^T]_{ij} > 0$）。

$\hat{\beta}_{\text{LSE}} = (X^TX)^{-1}X^TY$ 是 β 的**最佳线性无偏估计（BLUE）**；在正态分布假设下，它与极大似然估计 $\hat{\beta}_{\text{MLE}}$ 完全一致。

MLE（极大似然估计）

在多元线性回归模型的正态假设下，极大似然估计（MLE）的推导如下:

多元线性回归模型为 $Y = X\beta + \varepsilon$，其中随机误差 $\varepsilon \sim N(0, \sigma^2 I_n)$，因此条件分布 $Y|X \sim N(X\beta, \sigma^2 I_n)$。

- **概率密度函数（pdf）**：

$$
f(Y)=\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2)^{\frac{n}{2}}}\exp\left\{-\frac{1}{2}(Y-X\beta)^T\frac{1}{\sigma^2}(Y-X\beta)\right\}
$$

- **对数似然函数**:
  对 β 的对数似然 $\ell(\beta)$ 为:

$$
\ell(\beta)=-\frac{n}{2}\ln 2\pi-\frac{n}{2}\ln\sigma^2-\frac{1}{2\sigma^2}(Y-X\beta)^T(Y-X\beta)
$$

最大化 $\ell(\beta)$ 等价于最小化残差平方和 $(Y-X\beta)^T(Y-X\beta)$，因此:

$$
\hat{\beta}_{\text{MLE}}=\arg\min_{\beta}(Y-X\beta)^T(Y-X\beta)=\hat{\beta}_{\text{LSE}}
$$

即 β 的极大似然估计与最小二乘估计（LSE）完全一致。

对 σ^2 的极大似然估计

将对数似然函数关于 σ^2 展开:

$$
\ell(\sigma^2)=-\frac{n}{2}\ln 2\pi-\frac{n}{2}\ln\sigma^2-\frac{1}{2\sigma^2}(Y-X\beta)^T(Y-X\beta)
$$

最大化 $\ell(\sigma^2)$ 等价于最小化 $\frac{n}{2}\ln\sigma^2+\frac{1}{2\sigma^2}(Y-X\beta)^T(Y-X\beta)$。对 σ^2 求导并令导数为 0:

$$
\frac{n}{2}\cdot\frac{1}{\sigma^2}-\frac{1}{2(\sigma^2)^2}(Y-X\hat{\beta})^T(Y-X\hat{\beta})=0
$$

解得 σ^2 的极大似然估计:

$$
\hat{\sigma}^2=\frac{1}{n}\|Y-X\hat{\beta}\|_2^2=\frac{1}{n}\|Y-\hat{Y}\|_2^2
$$

Qualities

β 参数估计

回归系数的估计：$b=(X^TX)^{-1}X^TY=\hat{\beta}$

在正态假设下，$b \sim N\left(\beta, \sigma^2(X^TX)^{-1}\right)$

σ^2 的无偏估计

σ^2 的极大似然估计（MLE）

$$\hat{\sigma}^2_{\text{MLE}} = \frac{1}{n}\|Y - \hat{Y}\|_2^2 = \frac{1}{n}\|Y - Xb\|_2^2$$

$\hat{\sigma}^2_{\text{MLE}}$ 的期望
推导过程：

$$\begin{aligned}
E(\hat{\sigma}^2_{\text{MLE}}) &= \frac{1}{n}E\|Y - Xb\|_2^2 = \frac{1}{n}E\|Y - HY\|_2^2 \quad (\text{因 } \hat{Y} = HY) \\
&= \frac{1}{n}E\left[Y^T(I-H)^T(I-H)Y\right] \quad (\text{因 } I-H \text{幂等}) \\
&= \frac{1}{n}E\left[Y^T(I-H)Y\right] \\
&= \frac{1}{n}E\left[(X\beta+\varepsilon)^T(I-H)(X\beta+\varepsilon)\right] \\
&= \frac{1}{n}E\left[\varepsilon^T(I-H)\varepsilon\right] \quad (\text{因 } (I-H)X = 0) \\
&= \frac{1}{n}E\left[\text{tr}(\varepsilon^T(I-H)\varepsilon)\right] \\
&= \frac{1}{n}\text{tr}\left[(I-H)\text{Cov}(\varepsilon)\right] \\
&= \frac{1}{n}\text{tr}\left[(I-H)\sigma^2 I_n\right] \\
&= \frac{\sigma^2}{n}\text{tr}(I-H) \\
&= \frac{\sigma^2}{n}(n-k-1)
\end{aligned}$$

可见 $E(\hat{\sigma}^2_{\text{MLE}}) \neq \sigma^2$，即 $\hat{\sigma}^2_{\text{MLE}}$ 有偏。

修正后得到无偏估计：

$$s^2 = \frac{\hat{\sigma}^2_{\text{MLE}}}{\frac{n-k-1}{n}} = \frac{1}{n-k-1}\|Y - Xb\|_2^2$$

且 $E(s^2) = \sigma^2$

$s^2 \perp\!\!\!\perp b$

$s^2 = \frac{1}{n-k-1}\|Y - Xb\|^2$，证明 $Y - Xb \perp\!\!\!\perp b$ 即可。

其中，$Y - Xb = Y - HY = (I-H)Y$，$b = (X^TX)^{-1}X^TY$。

向量组 $\begin{pmatrix} Y - Xb \\ b \end{pmatrix} = \begin{pmatrix} I - H \\ (X^TX)^{-1}X^T \end{pmatrix} Y$ 满足**联合正态**。

协方差推导：

$$\begin{aligned} \mathrm{Cov}(Y - Xb, b) &= \mathrm{Cov}\left((I-H)Y, (X^TX)^{-1}X^TY\right) \\ &= (I-H)\mathrm{Cov}(Y)X(X^TX)^{-1} \\ &= 0 \end{aligned}$$

CI: Confidence Interval（置信区间）

（模型：$Y = X\beta + \varepsilon$）

---

① 对 β_i 的置信区间推导

- 由 $b \sim N\left(\beta, \sigma^2(X^TX)^{-1}\right)$，可得第 i 个回归系数的分布：

$$b_i \sim N\left(\beta_i, \sigma^2(X^TX)^{-1}_{[i+1,i+1]}\right)$$

  其中 $(X^TX)^{-1}_{[i+1,i+1]}$ 表示矩阵 $(X^TX)^{-1}$ 的第 $i+1$ 行第 $i+1$ 列元素。

- 构造标准正态变量：

$$\frac{b_i - \beta_i}{\sigma\sqrt{(X^TX)^{-1}_{[i+1,i+1]}}} \sim N(0,1)$$

- 代入 σ^2 的无偏估计 s^2（$s^2 = \frac{1}{n-k-1}\|Y - Xb\|_2^2$），且 $\frac{(n-k-1)s^2}{\sigma^2} \sim \chi^2(n-k-1)$，根据 t 分布定义：

$$\frac{b_i - \beta_i}{\sigma\sqrt{(X^TX)^{-1}_{[i+1,i+1]}}}\sqrt{\frac{\sigma^2}{s^2}} = \frac{b_i - \beta_i}{s\sqrt{(X^TX)^{-1}_{[i+1,i+1]}}} \sim t(n-k-1)$$

- 因此，β_i 的置信区间为：

$$|b_i - \beta_i| \leq t_{\frac{\alpha}{2}}(n-k-1) \cdot s\sqrt{(X^TX)^{-1}_{[i+1,i+1]}}$$

  其中 $t_{\frac{\alpha}{2}}(n-k-1)$ 是自由度为 $n-k-1$ 的 t 分布上 $\frac{\alpha}{2}$ 分位数。

---

② Mean Response（均值响应的置信区间）

目标：构建 $E\left(Y_{n+1} \mid x_{n+1}^T\right) = \mu_{n+1}$ 的置信区间

矩阵与模型设定

设计矩阵 $X \triangleq \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix}$，其中**每行是一个观测，每列是一个特征。**

多元线性回归模型的条件期望：

- 整体条件期望：$E(Y \mid X) = X\beta$
- 单个观测的条件期望：$E(Y_i \mid x_i^T) = x_i^T\beta$

回归系数的最小二乘估计：$b = (X^TX)^{-1}X^TY$

- 均值响应的真实值：$\mu_{n+1} = x_{n+1}^T\beta$
- 均值响应的估计值：$\hat{\mu}_{n+1} = x_{n+1}^Tb$

估计误差的分布：

$$\hat{\mu}_{n+1} - \mu_{n+1} = x_{n+1}^T(b-\beta) \sim N\left(0, \sigma^2 x_{n+1}^T(X^TX)^{-1}x_{n+1}\right)$$

代入 σ^2 的无偏估计 s^2（$s^2 = \frac{1}{n-k-1}\|Y - Xb\|_2^2$），构造 t 统计量：

$$\frac{\hat{\mu}_{n+1} - \mu_{n+1}}{\sigma\sqrt{x_{n+1}^T(X^TX)^{-1}x_{n+1}}}\sqrt{\frac{\sigma^2}{s^2}} = \frac{\hat{\mu}_{n+1} - \mu_{n+1}}{s\sqrt{x_{n+1}^T(X^TX)^{-1}x_{n+1}}} \sim t(n-k-1)$$

因此，μ_{n+1} 的置信区间为：

$$|\mu_{n+1} - \hat{\mu}_{n+1}| < t_{n-k-1}\left(\frac{\alpha}{2}\right) \cdot s\sqrt{x_{n+1}^T (X^T X)^{-1} x_{n+1}}$$

其中 $t_{n-k-1}\left(\frac{\alpha}{2}\right)$ 是自由度为 $n-k-1$ 的 t 分布上 $\frac{\alpha}{2}$ 分位数。

---

③ Y_{n+1}：Actual Response（实际响应的置信区间）

目标：已知 x_{n+1}，构建 Y_{n+1} 的置信区间

实际响应的模型：$Y_{n+1} = x_{n+1}^T \beta + \varepsilon_{n+1}$

估计响应与实际响应的误差：

$$\begin{aligned}
\hat{Y}_{n+1} - Y_{n+1} &= x_{n+1}^T (b - \beta) - \varepsilon_{n+1} \\
&= x_{n+1}^T (X^T X)^{-1} X^T (X\beta + \varepsilon) - \beta - \varepsilon_{n+1} \\
&= x_{n+1}^T (X^T X)^{-1} X^T \varepsilon - \varepsilon_{n+1} \\
&= \left[x_{n+1}^T (X^T X)^{-1} X^T,\ -1\right] \begin{bmatrix} \varepsilon \\ \varepsilon_{n+1} \end{bmatrix}
\end{aligned}$$

由 $\varepsilon \sim N(0, \sigma^2 I_n)$，$\varepsilon_{n+1} \sim N(0, \sigma^2)$ 且与 ε 独立，可知 $\hat{Y}_{n+1} - Y_{n+1} \sim N(0, \mathrm{Var})$，其中方差：

$$\begin{aligned}
\mathrm{Var} &= \sigma^2 \left[x_{n+1}^T (X^T X)^{-1} X^T,\ -1\right] \begin{bmatrix} X(X^T X)^{-1} x_{n+1} \\ -1 \end{bmatrix} \\
&= \sigma^2 \left(x_{n+1}^T (X^T X)^{-1} x_{n+1} + 1\right)
\end{aligned}$$

代入 σ^2 的无偏估计 s^2（$s^2 = \frac{1}{n-k-1} \|Y - Xb\|_2^2$），构造 t 统计量：

$$\frac{|\hat{Y}_{n+1} - Y_{n+1}|}{s\sqrt{x_{n+1}^T (X^T X)^{-1} x_{n+1} + 1}} \sim t(n-k-1)$$

因此，Y_{n+1} 的置信区间为：

$$|\hat{Y}_{n+1} - Y_{n+1}| \le t_{n-k-1}\left(\frac{\alpha}{2}\right) \cdot s\sqrt{x_{n+1}^T (X^T X)^{-1} x_{n+1} + 1}$$

其中 $t_{n-k-1}\left(\frac{\alpha}{2}\right)$ 是自由度为 $n-k-1$ 的 t 分布上 $\frac{\alpha}{2}$ 分位数。

Sum of squares

总平方和（SST）

$$
\begin{aligned}
\mathrm{SST} &= \|Y - \bar{Y}\mathbf{1}_n\|_2^2 \\
&= Y^T Y + n\bar{Y}^2 - 2\bar{Y}\mathbf{1}_n^T Y \\
&= Y^T Y - n\bar{Y}^2 \\
&= Y^T \left(I - \frac{\mathbf{1}_n \mathbf{1}_n^T}{n}\right) Y
\end{aligned}
$$

其自由度：$\mathrm{tr}\left(I - \frac{\mathbf{1}_n \mathbf{1}_n^T}{n}\right) = n - 1$

回归平方和（SSR）

$$
\begin{aligned}
\mathrm{SSR} &= \|\hat{Y} - \bar{Y}\mathbf{1}_n\|_2^2 \\
&= \|HY - \bar{Y}\mathbf{1}_n\|_2^2 \quad (\text{因 } \hat{Y} = HY) \\
&= Y^T HY + \bar{Y}^2 n - 2\bar{Y}\mathbf{1}_n^T HY \quad (\text{因 } \mathbf{1}_n \text{经} H \text{变换后仍为} \mathbf{1}_n) \\
&= Y^T HY - n\bar{Y}^2 \\
&= Y^T \left(H - \frac{\mathbf{1}_n \mathbf{1}_n^T}{n}\right) Y
\end{aligned}
$$

其自由度：$\mathrm{tr}\left(H - \frac{\mathbf{1}_n \mathbf{1}_n^T}{n}\right) = k + 1 - 1 = k$

残差平方和（SSE）

$$
\begin{aligned}
\mathrm{SSE} &= \|\hat{Y} - Y\|_2^2 \\
&= \|HY - Y\|_2^2 \\
&= \|(H - I)Y\|_2^2 \\
&= Y^T (I - H) Y
\end{aligned}
$$

其自由度：$\mathrm{tr}(I - H) = n - k - 1$

平方和分解关系：$\mathrm{SST} = \mathrm{SSR} + \mathrm{SSE}$

Model fitness（模型拟合优度）

R^2

判定系数公式：

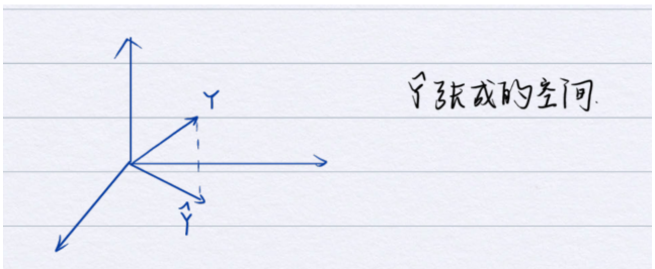
$$R^2=\frac{\mathrm{SSR}}{\mathrm{SST}}=1-\frac{\mathrm{SSE}}{\mathrm{SST}}$$

影响规律：解释变量个数 (k \uparrow) 时，(\text{SSE} \downarrow)，(R^2 \uparrow)

推导与本质：

$$\begin{aligned}
R^2&=\frac{\|\hat{Y}-\bar{Y}\|_2^2}{\|Y-\bar{Y}\|_2^2}=\frac{\|\hat{Y}-\bar{Y}\|_2^2}{\|\hat{Y}-\bar{Y}\|_2^2+\|\hat{Y}-Y\|_2^2}\\
&=\frac{\left[Y^T\left(H-\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\right)Y\right]^2}{\|Y-\bar{Y}\mathbf{1}_n\|_2^4}\\
&=\frac{\left[Y^T\left(H-\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\right)\left(I_n-\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\right)Y\right]^2}{\|Y-\bar{Y}\mathbf{1}_n\|_2^4}\\
&=\frac{\left[(\hat{Y}-\bar{Y})^T(Y-\bar{Y})\right]^2}{\|Y-\bar{Y}\|_2^2\cdot\|\hat{Y}-\bar{Y}\|_2^2}\\
&=\mathrm{Corr}^2(\hat{Y}-\bar{Y},Y-\bar{Y})
\end{aligned}$$

- 几何意义（右侧图形）：
  图中展示了因变量 (Y)、拟合值 (\hat{Y}) 在空间中的投影关系，(\hat{Y}) 是 (Y) 向 (X) 张成的列空间的投影，(R^2) 量化了这种投影与 (Y) 本身的契合程度（“(Y) 张成的空间”为辅助理解的几何背景）。



R_a^2（Adjusted R^2，调整后判定系数）

- 形式一：

$$R_a^2 = 1 - \frac{\text{MSE}}{\text{MST}}$$

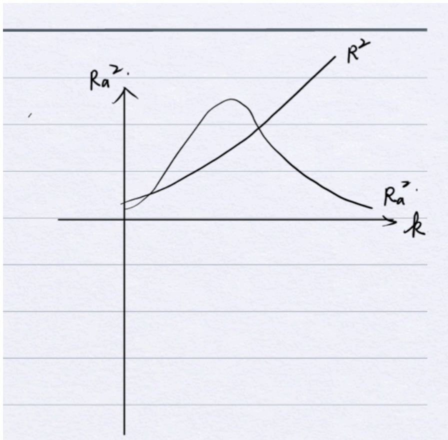
- 形式二（展开推导）：

$$R_a^2 = 1 - \frac{\text{SSE}}{\text{SST}} \cdot \frac{n-1}{n-k-1}$$

变化规律（基于解释变量个数 k 的影响）

- 当 k 较小时：$R_a^2 \approx 1 - \frac{\text{SSE}}{\text{SST}}$，此时 $k\uparrow$ 会带动 $R_a^2\uparrow$。
- 当 k 较大时：$R_a^2 \approx 1 - \frac{n-1}{n-k-1}$，此时 $k\uparrow$ 会导致 $R_a^2\downarrow$。

图中展示了 R^2 和 R_a^2 随解释变量个数 k 变化的趋势：



- R^2 随 k 增大**单调递增**；
- R_a^2 随 k 增大先**上升后下降**，存在一个峰值点，体现了其对解释变量个数的“惩罚性”调整。

Hypothesis Testing（假设检验）

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$

模型设定：$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$

原假设：$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$

F 统计量定义：$F \triangleq \frac{\text{MSR}}{\text{MSE}} \sim F_{k,n-k-1}$

决策规则：若 $F > F_{k,n-k-1}(\alpha)$，则拒绝 H_0

$H_0^*: \beta_i = 0$

t 检验推导:
已知回归系数估计量的分布 $b \sim N\left(\beta, \sigma^2 (X^T X)^{-1}\right)$，因此第 i 个回归系数的分布为:

$$b_i \sim N\left(\beta_i, \sigma^2 (X^T X)^{-1}_{[i+1,i+1]}\right)$$

其中 $(X^T X)^{-1}_{[i+1,i+1]}$ 表示矩阵 $(X^T X)^{-1}$ 的第 $i+1$ 行第 $i+1$ 列元素。

构造 t 统计量:

$$\frac{\frac{b_i-\beta_i}{\sigma\sqrt{(X^T X)^{-1}_{[i+1,i+1]}}}}{\frac{s}{\sigma}} = \frac{b_i-\beta_i}{s\sqrt{(X^T X)^{-1}_{[i+1,i+1]}}} \sim t(n-k-1)$$

当检验 $H_0^*: \beta_i = 0$ 时，统计量简化为:

$$t \triangleq \frac{b_i}{s\sqrt{(X^T X)^{-1}_{[i+1,i+1]}}}$$

- **决策规则**:
  若 $|t| > t_{(n-k-1)}\left(\frac{\alpha}{2}\right)$，则拒绝 H_0^*。

多因素回归的系数约束检验

交互效应模型转换

考虑含两个因素的回归模型:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \quad (1)$$

若需解释 $x_1 - x_2$ 的交互效应，可将模型转换为:

$$y = \beta_0 + \beta_1 (x_1 - x_2) + \varepsilon = \beta_0 + \beta_1 x_1 - \beta_1 x_2 + \varepsilon \quad (2)$$

此时在模型 (1) 中，需满足约束 $\beta_1 + \beta_2 = 0$，即 (1) 可通过该约束转换为 (2)。

一般约束形式 $C\beta = 0_m$

- 回归系数向量:
  $$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1)\times 1}$$
- 约束矩阵:
  C 为 $m \times (k+1)$ 矩阵，用于表示对 β 的线性约束。

示例 1（$\beta_1 + \beta_2 = 0$）

约束矩阵:

$$C = \begin{bmatrix} 0 & 1 & 1 & 0 & \dots & 0 \end{bmatrix}$$

则

$$C\beta = \begin{bmatrix} 0 & 1 & 1 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} = \beta_1 + \beta_2 = 0$$

示例 2（$\beta_i = \beta_j = 0$）

约束矩阵:

$$C = \begin{bmatrix} 0 & \dots & \underset{(i+1)}{1} & \dots & 0 & & \\ 0 & \dots & 0 & \dots & \underset{(j+1)}{1} & \dots & 0 \end{bmatrix}$$

则

$$C\beta = \begin{bmatrix} \beta_i \\ \beta_j \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Extra Sum of Square(for reduced model)

全模型设定

模型形式:

$$Y=\tilde{X}\tilde{\beta}+\varepsilon=\beta_0\mathbf{1}_n+(X_1\ldots X_k)\begin{bmatrix}\beta_1\\\beta_2\\\vdots\\\beta_k\end{bmatrix}+\varepsilon$$

其中，$\tilde{X}=(\mathbf{1}_n\mid X)$ 是 $n\times(k+1)$ 矩阵，$\tilde{\beta}=\begin{bmatrix}\beta_0\\\beta_1\\\vdots\\\beta_k\end{bmatrix}$ 是 $(k+1)\times 1$ 向量，解释变量个数 $\#\text{变量}=k$。

- 全模型残差平方和（SSE $_F$）：

$$\text{SSE}_F=Y^T(I-H_F)Y$$

其中投影矩阵 $H_F=\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T$，$\tilde{X}=(\mathbf{1}_n\mid X)$。

- 全模型回归平方和（SSR $_F$）：

$$\text{SSR}_F=Y^T\left(H_F-\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\right)Y$$

模型拆分与约简模型

将设计矩阵拆分为 $X\triangleq(X_1\mid X_2)$，其中 X_1 是 $n\times k_1$ 矩阵，X_2 是 $n\times k_2$ 矩阵，满足 $k_1+k_2=k$；

回归系数拆分为 $\beta=\begin{bmatrix}\beta_1\\\vdots\\\beta_{k_1}\\\beta_{k_1+1}\\\vdots\\\beta_k\end{bmatrix}\triangleq\begin{bmatrix}\beta_1\\\beta_2\end{bmatrix}$，其中 β_1 是 $k_1\times 1$ 向量，β_2 是 $k_2\times 1$ 向量（将感兴趣的变量放在 X_1 里）。

约简模型（Reduced Model）：

$$Y = \beta_0 \mathbf{1}_n + X_1\beta_1 + \varepsilon$$

其中解释变量个数 $\#\text{变量} = k_1 < k$，且此处的 β_0、β_1 可能与全模型中的系数不同。

约简模型的平方和与额外平方和推导

- 约简模型残差平方和（SSE $_R$）：

  $$\text{SSE}_R = Y^T(I - H_R)Y$$

  其中投影矩阵 $H_R = \mathring{X}(\mathring{X}^T\mathring{X})^{-1}\mathring{X}^T$，$\mathring{X} = (\mathbf{1}_n \mid X_1)$ 是 $n \times (k_1+1)$ 矩阵（表示向约简后的 $\mathring{X}$ 张成的空间投影）。

- 约简模型回归平方和（SSR $_R$）：

  $$\text{SSR}_R = Y^T\left(H_R - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\right)Y$$

- 额外平方和定义：
  全模型与约简模型的残差平方和之差，即 $\text{SSE}_R - \text{SSE}_F = \text{SSR}(X_2 \mid X_1)$，表示在 X_1 已入模的情况下，加入 X_2 带来的**边缘回归平方和**（反映 X_2 对 Y 变异的额外解释能力）。

  直观理解：SSE 可视为“距离”，解释变量个数 $k\uparrow$ 时 $\text{SSE}\downarrow$；由于全模型的 k 大于约简模型的 k_1，故 $\text{SSE}_R \geq \text{SSE}_F$。

- 特殊情况：
  若 X_2 是 X_1 的线性组合，则 X_2 无法带来额外解释能力，此时 $\text{SSE}_R \approx \text{SSE}_F$。

回至主线：带约束的模型检验

约束假设与参数拆分

原假设：$H_0 : C\beta = 0$，其中 C 是 $m \times (k+1)$ 约束矩阵，β 是 $(k+1) \times 1$ 回归系数向量。

- 约束性质：m 个方程，$k+1$ 个参数，且 $m < k+1$。
- 参数拆分：由于有 $k+1-m$ 个参数可变，将 β 拆分为

$$
\beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_{k-m} \\ * \end{bmatrix} \triangleq \begin{bmatrix} \eta \\ A\eta \end{bmatrix}
$$

其中 η 是 $(k-m+1)\times 1$ 向量，A 是 $m\times(k-m+1)$ 矩阵。

为什么可以这么拆分

（1）将回归系数按自由度分块

将系数向量 β 按“自由参数”和“约束参数”分成两部分：

- 前 $(k-m+1)$ 个参数是**自由参数**，记为

$$
\eta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{k-m} \end{bmatrix}
$$

  维度为 $(k-m+1)\times 1$。

- 后 m 个参数是**约束参数**，记为

$$
\gamma = \begin{bmatrix} \beta_{k-m+1} \\ \vdots \\ \beta_k \end{bmatrix}
$$

  维度为 $m\times 1$。

于是整体向量可写成分块形式：

$$
\beta = \begin{bmatrix} \eta \\ \gamma \end{bmatrix}
$$

（2）将约束 $C\beta = 0$ 分块

约束是线性方程：

$$
C\beta = 0
$$

将 C 的列按 η 和 γ 的分块对应拆分：

- C_1：C 的前 $(k-m+1)$ 列，维度为 $m\times(k-m+1)$
- C_2：C 的后 m 列，维度为 $m\times m$，且 $\operatorname{rank}(C_2)=m$（因 C 列满秩）

于是：

$$C\beta = C_1\eta + C_2\gamma = 0$$

（3）由自由参数解出约束参数

由线性方程：

$$C_1\eta + C_2\gamma = 0$$

可解得：

$$\gamma = -C_2^{-1}C_1\eta$$

定义矩阵：

$$A = -C_2^{-1}C_1$$

其维度为 $m\times(k-m+1)$。

于是约束参数 γ 可写为：

$$\gamma = A\eta$$

（4）最终参数表达式

将 γ 代回向量 β：

$$\beta = \begin{bmatrix}\eta \\ A\eta\end{bmatrix}$$

这表明：
在约束 $C\beta=0$ 下，原本的 $(k+1)$ 个参数全部可由 $(k-m+1)$ 维的自由向量 η 唯一表示。

约简模型

全模型：$Y = X\beta+\varepsilon$，其中 $X=(X_1\mid X_2)$（X_1 为 $n\times(k-m+1)$ 矩阵，X_2 为 $n\times m$ 矩阵）。

代入参数拆分式，约简模型为：

$$\begin{aligned} Y &= X_1\eta + X_2A\eta + \varepsilon \\ &= (X_1 + X_2A)\eta + \varepsilon \\ &\triangleq \tilde{X}\eta + \varepsilon \end{aligned}$$

其中 $\tilde{X} = X_1 + X_2A$ 是 $n \times (k-m+1)$ 矩阵，约简模型的解释变量个数 $\#\text{变量} = k - m$。

平方和定义（约简模型与全模型对比）

- 约简模型残差平方和（SSE $_R$）：

  $$\text{SSE}_R = Y^T(I - H_R)Y$$

  其中投影矩阵 $H_R = \tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T$，$\tilde{X} = X_1 + X_2A$。
- 约简模型回归平方和（SSR $_R$）：

  $$\text{SSR}_R = Y^T\left(H_R - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\right)Y$$

- 全模型（无约束）：
  - 回归平方和（SSR $_F$）：$\text{SSR}_F = Y^T\left(H_F - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\right)Y$
  - 残差平方和（SSE $_F$）：$\text{SSE}_F = Y^T(I - H_F)Y$
  - 投影矩阵（H_F）：$H_F = X(X^TX)^{-1}X^T$，X 为全模型设计矩阵。

约束假设下的平方和差异检验

检验逻辑

- 若 H_0 未被拒绝，可认为 $\text{SSE}_R - \text{SSE}_F \approx 0$；
- 若 $\text{SSE}_R - \text{SSE}_F > A$，则拒绝 H_0，其中 $\alpha = P(\text{Reject} \mid H_0)$ 为检验的显著性水平。

平方和差异的定义与模型设定

需明确 $\text{SSE}_R - \text{SSE}_F$ 的推导，定义：

- 约简模型（满足 $H_0: C\beta = 0$）：$Y = X\beta + \varepsilon$ 且 $C\beta = 0$，其残差平方和 $\mathrm{SSE}_R = \|Y - \hat{Y}_R\|_2^2 = \|Y - X\hat{\beta}\|_2^2$，其中 $\hat{\beta} = \arg\min_\beta \|Y - X\beta\|_2^2 \text{ s.t. } C\beta = 0$。
- 全模型：$Y = X\beta + \varepsilon$，其残差平方和 $\mathrm{SSE}_F = \|Y - \hat{Y}_F\|_2^2 = \|Y - Xb\|_2^2$（$b = (X^TX)^{-1}X^TY$ 为无约束最小二乘估计）。

带约束的最小二乘估计推导（拉格朗日乘数法）

构造拉格朗日函数（用于最小化残差平方和 $\|Y - X\hat{\beta}\|_2^2$，约束为 $C\hat{\beta} = 0$）：

$$\mathcal{L}(\hat{\beta}, \lambda) = Y^TY - 2Y^TX\hat{\beta} + \hat{\beta}^TX^TX\hat{\beta} + \lambda^TC\hat{\beta}$$

对 $\hat{\beta}$ 和拉格朗日乘数 λ 求偏导并令其为 0:

- 对 $\hat{\beta}$ 求导:

$$\frac{\partial \mathcal{L}}{\partial \hat{\beta}} = -2X^TY + 2X^TX\hat{\beta} + C^T\lambda^T = 0$$

- 对 λ 求导:

$$\frac{\partial \mathcal{L}}{\partial \lambda} = C\hat{\beta} = 0$$

记无约束最小二乘估计为 $b = (X^TX)^{-1}X^TY$，对 $\hat{\beta}$ 的导数方程左乘 $(X^TX)^{-1}$:

$$-2(X^TX)^{-1}X^TY + 2\hat{\beta} + (X^TX)^{-1}C^T\lambda^T = 0 \tag{③}$$

对约束方程 $C\hat{\beta} = 0$ 左乘 $C(X^TX)^{-1}$ 并结合式③，解出 λ^T:

$$\lambda^T = 2\left[C(X^TX)^{-1}C^T\right]^{-1}C(X^TX)^{-1}X^TY$$

将 λ^T 代入式③，整理得:

$$\begin{aligned}\hat{\beta} &= b - (X^TX)^{-1}C^T\left[C(X^TX)^{-1}C^T\right]^{-1}Cb \\ &\triangleq b - \Delta b\end{aligned}$$

其中 $\Delta = (X^TX)^{-1}C^T\left[C(X^TX)^{-1}C^T\right]^{-1}C$ 为约束影响矩阵。

约简模型残差平方和 $\mathrm{SSE}_R = \|Y - X\hat{\beta}\|_2^2$，全模型残差平方和 $\mathrm{SSE}_F = \|Y - Xb\|_2^2$，两者的差异

为：

$$
\begin{aligned}
\mathrm{SSE}_R-\mathrm{SSE}_F &= \|Y-X(b-\Delta b)\|_2^2-\|Y-Xb\|_2^2\\
&= \|X\Delta b\|_2^2+2(Y-Xb)^TX\Delta b\\
&= \|X\Delta b\|_2^2 \quad (\text{因}(Y-\hat{Y}_F)^TX=Y^T(I-H_F)X=0)\\
&= b^T\Delta^TX^TX\Delta b\\
&= b^TC^T\left[C(X^TX)^{-1}C^T\right]^{-1}Cb
\end{aligned}
$$

其中利用了残差性质：$Y-\hat{Y}_F=(I-H_F)Y$，且 $(I-H_F)X=0$（投影矩阵的正交性），故交叉项 $(Y-Xb)^TX\Delta b=0$。

约束假设下的 F 检验推导

检验逻辑与前提

原假设：\$$H_0: C\beta=0$,\$
其中 C 是 $m\times(k+1)$ 的约束矩阵，β 是 $(k+1)\times 1$ 的回归系数向量。

若 $\mathrm{SSE}_R-\mathrm{SSE}_F>A$ 则拒绝 H_0，其中显著性水平定义为

$$\alpha=P\big(\mathrm{SSE}_R-\mathrm{SSE}_F>A\mid H_0\big).$$

平方和差异的分布分析

记无约束（full）估计为 $b=(X^TX)^{-1}X^TY$，则在高斯误差假设下

$$b\sim N\big(\beta,\ \sigma^2(X^TX)^{-1}\big).$$

考虑线性变换 Cb，由于 C 为常矩阵，

$$Cb\sim N\big(C\beta,\ \sigma^2\,C(X^TX)^{-1}C^T\big).$$

在原假设 H_0 下，$C\beta=0$，因此

$$Cb\sim N\big(0,\ \sigma^2\Sigma\big),\qquad \Sigma\triangleq C(X^TX)^{-1}C^T,$$

且 Σ 为 $m \times m$ 的对称正定矩阵（若 C 列满秩，则 $\operatorname{rank}(\Sigma) = m$）。

于是有标准化的二次型：

$$\frac{1}{\sigma^2}(Cb)^T \Sigma^{-1}(Cb) \sim \chi^2(m).$$

另对平方和差异的矩阵表达式可写成（等价形式）：

$$\mathrm{SSE}_R - \mathrm{SSE}_F = (Cb)^T \left[C(X^T X)^{-1} C^T \right]^{-1} (Cb) = (Cb)^T \Sigma^{-1}(Cb).$$

因此在 H_0 下

$$\frac{1}{\sigma^2}\left(\mathrm{SSE}_R - \mathrm{SSE}_F\right) \sim \chi^2(m).$$

均方误差与 F 统计量构造

记全模型（full model）残差平方和为 SSE_F，其自由度为 $n-(k+1)$（有时记为 $n-k-1$）。定义全模型残差均方：

$$\mathrm{MSE}_F = \frac{\mathrm{SSE}_F}{n-k-1}.$$

已知在模型假设下

$$\frac{1}{\sigma^2}\,\mathrm{SSE}_F \sim \chi^2(n-k-1).$$

并且由投影矩阵的正交分解性质，可证明 $\mathrm{SSE}_R - \mathrm{SSE}_F$ 与 SSE_F 相互独立，因此可以用独立的卡方量比构造 F 统计量：

$$F \triangleq \frac{(\mathrm{SSE}_R - \mathrm{SSE}_F)/m}{\mathrm{MSE}_F} = \frac{\left[(\mathrm{SSE}_R - \mathrm{SSE}_F)/\sigma^2\right]/m}{\left[\mathrm{SSE}_F/\sigma^2\right]/(n-k-1)} \overset{H_0}{\sim} F_{m,\,n-k-1}.$$

决策规则

计算观测到的统计量

$$F_{\text{obs}} = \frac{(\text{SSE}_R - \text{SSE}_F)/m}{\text{SSE}_F/(n-k-1)}.$$

给定显著性水平 α，查表（或用软件）得临界值 $F_{m,n-k-1}(\alpha)$。若

$$F_{\text{obs}} > F_{m,n-k-1}(\alpha),$$

则拒绝 H_0；否则不拒绝 H_0。

---

remark

$\text{SSE}_R - \text{SSE}_F$ 与 SSE_F **独立性说明（关键点）：**

矩阵 $(I-H)$ 与 H 正交，且 $(I-H)X=0$（因为 $HX=X$），因此残差向量 e 与列空间 $\mathcal{C}(X)$ 中的任意向量正交。这一正交性质推导出：

- 由 Cb 可表示为 $C(X^TX)^{-1}X^TY$（属于 X 的线性变换空间），而 $e=(I-H)Y$ 属于与 X 列空间正交的空间；
- 因为 Y 在高斯假设下为多元正态，线性变换的分量如果互相正交且协方差为 0，则它们相互独立；
- 于是 $(\text{SSE}_R - \text{SSE}_F)$（对应于 Cb 的平方型）与 SSE_F（对应于 e 的平方和）独立，从而 F 统计量的推导合法。

约束假设检验的实例推导

例 1：单个回归系数的显著性检验

假设设定：$H_0: \beta_i = 0$，等价于约束矩阵 $C = \left[0,\dots,0,\underset{(i+1)}{1},0,\dots,0\right]$，满足 $C\beta = \beta_i$。

平方和差异与统计量构造：
根据前文推导，平方和差异为：

$$\text{SSE}_R - \text{SSE}_F = (Cb)^T\left[C(X^TX)^{-1}C^T\right]^{-1}Cb$$

其中 $Cb=b_i$（第 i 个回归系数的估计值），$C(X^TX)^{-1}C^T = (X^TX)^{-1}_{[i+1,i+1]}$（矩阵 $(X^TX)^{-1}$ 的第 $i+1$ 行第 $i+1$ 列元素）。

因此，F 统计量为：

$$\frac{\frac{\mathrm{SSE}_R-\mathrm{SSE}_F}{1}}{\frac{\mathrm{SSE}_F}{n-k-1}}=\frac{\frac{b_i^2}{(X^TX)^{-1}_{[i+1,i+1]}}}{s^2}\overset{H_0}{\sim}F_{1,n-k-1}$$

其中 $s^2=\frac{\mathrm{SSE}_F}{n-k-1}$ 是残差均方。

与 t 检验的等价性：
注意到 $F_{1,n-k-1}=t^2_{n-k-1}$，因此该 F 统计量等价于单个回归系数显著性检验的 t 统计量的平方，即：

$$\frac{b_i^2}{s^2(X^TX)^{-1}_{[i+1,i+1]}}=t^2_{n-k-1}$$

例 2：多个回归系数的联合显著性检验

模型与假设：
模型形式：$Y=X\beta+\varepsilon=\beta_0+\beta_1x_1+\cdots+\beta_kx_k+\varepsilon$，其中 $\varepsilon\sim N(0,\sigma^2I_n)$。
原假设：$H_0:\beta_1=\beta_2=0$，等价于约束矩阵

$$C=\begin{bmatrix}0,1,0,\dots,0\\0,0,1,\dots,0\end{bmatrix}$$

满足 $C\beta=\begin{bmatrix}\beta_1\\\beta_2\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}$，约束个数 $m=2$。

平方和差异与 F 统计量：
平方和差异为 $\mathrm{SSE}_R-\mathrm{SSE}_F=(Cb)^T\left[C(X^TX)^{-1}C^T\right]^{-1}Cb$，其中 $Cb=\begin{bmatrix}b_1\\b_2\end{bmatrix}$（$b_1,b_2$ 分别为 β_1,β_2 的估计值）。

构造 F 统计量：

$$\frac{\frac{\mathrm{SSE}_R-\mathrm{SSE}_F}{2}}{\frac{\mathrm{SSE}_F}{n-k-1}}\overset{H_0}{\sim}F_{2,n-k-1}$$

全模型与约简模型的平方和定义：

- 全模型残差平方和：$\mathrm{SSE}_F=Y^T(I-H)Y$，其中投影矩阵 $H=X(X^TX)^{-1}X^T$，$X\triangleq(\mathbf{1}_n,x_1,\dots,x_k)$（含截距项的设计矩阵）。
- 约简模型（不含 x_1,x_2）残差平方和：SSE_R 是该模型的残差平方和，平方和差异 SSE_R-

SSE_F 反映了 x_1, x_2 对 Y 变异的联合解释能力。