Linear Model （一元）

TOC

- proof

- sum of squares
  - Total sum of squares (SST)
  - Regression sum of squares (SSR)
  - Residual sum of squares (SSE)
  - 均方（Mean Square）
  - $\mathrm{SST} = \mathrm{SSR} + \mathrm{SSE}$
  - 判定系数 r^2
    - 意义分析
- Hypothesis Testing
  - 检验假设
  - 假设 H_0 下的平方和分布
    - 总平方和（SST）$\mathrm{SST} \sim \sigma^2\chi^2(n-1).$
    - 回归平方和（SSR）$\mathrm{SSR} \sim \sigma^2\chi^2(1)$
    - 残差平方和（SSE）
  - 基于 F 统计量的检验
  - 基于 t 统计量的检验
    - H_0 成立时 $\hat{\beta}$ 的分布
    - 统计量推导
    - 决策规则
  - remark **（F 与 t 的等价性）**：
- **置信区间（CI: Confidence Interval）**
  - 回归系数 β 的置信区间
    - 统计量分布
    - 置信区间推导
  - 均值 $\mu_0 = \alpha + \beta x_0$ 的置信区间（mean response）
    - 统计量分布
    - 置信区间推导
  - **个体预测值 $y_{n+1} = \alpha + \beta x_{n+1} + \varepsilon_{n+1}$ 的置信区间（预测区间）**

- 统计量分布
- 置信区间（预测区间）推导

基本内容

Assumption

1. $Y = \alpha + \beta x + \varepsilon$
   - α: Intercept（回归直线在纵轴的截距，反映 $x = 0$ 时 Y 的期望水平）
   - β: Slope（x 每变化 1 单位，Y 的平均变化量，体现 x 对 Y 的影响强度与方向）
     （α, β 为模型待估参数）
   - x: Independent variable（自变量，可为给定的固定值，也可为某随机变量的观测值，如“温度”“时间”等）
   - ε: Random Error Term（随机误差项）
     - $E\varepsilon = 0$ / $E(\varepsilon|x) = 0$：误差的期望为 0，说明模型已捕获 x 对 Y 的系统性影响，剩余波动无偏向性
     - $\mathrm{Var}\varepsilon = \sigma^2$ / $\mathrm{Var}(\varepsilon|x) = \sigma^2$：误差的方差恒定（同方差性），即无论 x 取何值，误差波动幅度稳定，σ^2 为待估参数
2. Observation $(x_1, Y_1) \ldots (x_n, Y_n)$（从总体中抽取的 n 组样本观测值）
   - (ε_i) – uncorrelated（不同观测的误差项之间无线性关联，若违背则 OLS 估计有效性下降，如时间序列中误差的“自相关”）
   - (Y_i) – uncorrelated（因变量的不同观测之间无线性关联，由“ε_i 不相关”和模型线性结构推导而来）
   - $E(Y_i) = \alpha + \beta x_i$（$Y_i$ 的期望由 $\alpha + \beta x_i$ 系统性决定）
   - $\mathrm{Var}(Y) = \mathrm{Var}(\varepsilon)$（$Y$ 的波动完全源于误差项 ε 的波动）

Estimates of (α, β)

- **Notation**
  - Sample var of x:

$$\frac{1}{n}\sum(x_i - \bar{x})^2 \triangleq \frac{s_{xx}}{n}$$

（$s_{xx}=\sum(x_i-\bar{x})^2$ *是* x *离均差平方和，反映* x *的样本波动程度*）

- Cov of (x, Y):

$$\frac{1}{n}\sum(x_i-\bar{x})(Y_i-\bar{Y})=\frac{s_{xy}}{n}$$

（$s_{xy}=\sum(x_i-\bar{x})(Y_i-\bar{Y})$ *是* x *与* Y *的离均差乘积和，反映二者的线性关联程度*）

- **We want to get**
  - $(\hat{\alpha}, \hat{\beta})$: Estimates of (α, β)
  - $\hat{\mu}(x)=\hat{\alpha}+\hat{\beta}x$: Fitted regression line
  - $\hat{y}_i=\hat{\alpha}+\hat{\beta}x_i$: Fitted value
  - $e_i=y_i-\hat{y}_i$: Residual

We can get estimates by 3 ways: LSE, BLUE, MLE

LSE 最小二乘法

- **Residual Sum of Squares (RRS)**
  关于直线 $E(Y|x)=\beta x+\alpha$

$$\text{RRS}(\alpha,\beta)=\sum_{i=1}^{n}(y_i-\beta x_i-\alpha)^2=\|\,y-\beta x-\alpha\mathbf{1}_n\,\|_2^2$$

因 $\text{RRS}(\alpha,\beta)$ 关于 (α,β) 均为凸，故对 (α,β) 求偏导并令其为 0:

$$\begin{cases}\frac{\partial \text{RRS}}{\partial \alpha}=-2\,\mathbf{1}_n^T(y-\beta x-\alpha\mathbf{1}_n)=0\\ \frac{\partial \text{RRS}}{\partial \beta}=-2\,x^T(y-\beta x-\alpha\mathbf{1}_n)=0\end{cases}$$

- **求解** $(\hat{\alpha}, \hat{\beta})$

  展开偏导方程:

$$
\begin{cases}
\mathbf{1}_n^T(y-\beta x-\alpha\mathbf{1}_n)=n\bar{y}-\beta n\bar{x}-n\alpha=0\\
x^T(y-\beta x-\alpha\mathbf{1}_n)=x^Ty-\beta x^Tx-n\alpha\bar{x}=0
\end{cases}
$$

由第一个方程得 $\alpha=\bar{y}-\beta\bar{x}$，代入第二个方程：

$$
x^Ty-\beta x^Tx-n(\bar{y}-\beta\bar{x})\bar{x}=0
$$

整理得：

$$
\hat{\beta}=\frac{x^Ty-n\bar{x}\bar{y}}{x^Tx-n\bar{x}^2},\qquad \hat{\alpha}=\bar{y}-\hat{\beta}\bar{x}
$$

BLUE 最佳无偏法

在线性模型中，我们希望估计量 $\hat{\beta},\hat{\alpha}$ 是**线性的**（即能表示为 y 的线性组合 c^Ty），且是**无偏的**（即 $E(\hat{\theta})=\theta$，θ 代表待估参数 β 或 α）。

对 $\hat{\beta}$ 的线性无偏估计推导

令 $\hat{\beta}=c^Ty$，其中
$y=\alpha\mathbf{1}_n+\beta x+\varepsilon$（$\mathbf{1}_n$ 为全1向量，ε 为误差向量）。

- 求期望：

$$
\begin{aligned}
E(\hat{\beta})=E(c^Ty)&=E\big(c^T(\alpha\mathbf{1}_n+\beta x+\varepsilon)\big)\\
&=\alpha\,\mathbf{1}_n^Tc+\beta\,x^Tc\qquad(E(\varepsilon)=0)
\end{aligned}
$$

- 无偏性要求 $E(\hat{\beta})=\beta$，因此必须满足：

$$
\begin{cases}
\mathbf{1}_n^Tc=0 & (\text{消去 }\alpha\text{ 的影响})\\
x^Tc=1 & (\text{保证 }\beta\text{ 系数为1})
\end{cases}
$$

最小方差的优化目标

线性估计的方差为：

$$\mathrm{Var}(\hat{\beta}) = \mathrm{Var}(c^T y) = c^T(\sigma^2 I_n)c = \sigma^2\|c\|_2^2.$$

目标：在满足无偏性约束 $\mathbf{1}_n^T c = 0$，$x^T c = 1$ 下，最小化 $\|c\|_2^2$。

拉格朗日对偶法求解最优 c

构造拉格朗日函数：

$$\mathcal{L} = \|c\|_2^2 + \lambda_1(\mathbf{1}_n^T c) + \lambda_2(x^T c - 1)$$

对 c 求偏导：

$$\frac{\partial \mathcal{L}}{\partial c} = 2c + \lambda_1 \mathbf{1}_n + \lambda_2 x = 0 \tag{1}$$

结合无偏性约束：

$$\begin{cases} \mathbf{1}_n^\top c = 0 & (2) \\ x^\top c = 1 & (3) \end{cases}$$

（1）左乘 $\mathbf{1}_n^T$

$$2\mathbf{1}_n^T c + \lambda_1 n + \lambda_2 n\bar{x} = 0$$

代入 $\mathbf{1}_n^T c = 0$：

$$\lambda_1 = -\lambda_2 \bar{x} \tag{4}$$

（2）左乘 x^T

$$2x^T c + \lambda_1 n\bar{x} + \lambda_2 x^T x = 0$$

代入 $x^T c = 1$ 和 (4)：

$$2 - \lambda_2 \bar{x}\, n\bar{x} + \lambda_2 x^T x = 0$$

整理得：

$$\lambda_2 = -\frac{2}{x^T x - n\bar{x}^2}, \qquad \lambda_1 = \frac{2\bar{x}}{x^T x - n\bar{x}^2}.$$

所以最终得到 c

$$\hat{c} = -\frac{1}{2}(\lambda_1 \mathbf{1}_n + \lambda_2 x) = \frac{-\bar{x}\mathbf{1}_n + x}{x^T x - n\bar{x}^2}.$$

代入 $\hat{\beta} = c^T y$:

$$\hat{\beta} = \frac{x^T y - n\bar{x}\bar{y}}{x^T x - n\bar{x}^2} = \frac{s_{xy}}{s_{xx}}.$$

方差:

$$\mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{s_{xx}}.$$

对 $\hat{\alpha}$ 的 BLUE 推导

线性估计与无偏性要求

令 $\hat{\alpha} = c^T y$（c 为待求系数向量），其中
$y = \alpha\mathbf{1}_n + \beta x + \varepsilon$（$\mathbf{1}_n$ 为全1向量，ε 为误差向量）。

- 求期望:

$$\begin{aligned} E(\hat{\alpha}) &= E(c^T y) = E\left[c^T(\alpha\mathbf{1}_n + \beta x + \varepsilon)\right] \\ &= \alpha\,\mathbf{1}_n^T c + \beta\, x^T c \qquad (\text{因 } E(\varepsilon) = 0) \end{aligned}$$

- 无偏性要求 $E(\hat{\alpha}) = \alpha$，因此约束为:

$$\begin{cases} \mathbf{1}_n^T c = 1 \\ x^T c = 0 \end{cases}$$

即：估计量对 α 的系数为 1，对 β 的影响为 0。

最小方差的目标转化

由于

$$\mathrm{Var}(y) = \sigma^2 I_n$$

故线性估计量方差为:

$$\mathrm{Var}(\hat{\alpha}) = c^T(\sigma^2 I_n)c = \sigma^2 \|c\|_2^2.$$

目标：在约束

$$\mathbf{1}_n^T c = 1, \quad x^T c = 0$$

下**最小化** $\|c\|_2^2$。

拉格朗日法求最优 c

构造拉格朗日函数:

$$\mathcal{L} = \|c\|_2^2 + \lambda_1(\mathbf{1}_n^T c - 1) + \lambda_2(x^T c)$$

对 c 求偏导并令其为零:

$$2c + \lambda_1 \mathbf{1}_n + \lambda_2 x = 0. \tag{1}$$

结合约束条件:

$$\begin{cases} \mathbf{1}_n^\top c = 1 & (2) \\ x^\top c = 0 & (3) \end{cases}$$

(1) 左乘 $\mathbf{1}_n^T$ 得:

$$2\mathbf{1}_n^T c + \lambda_1 n + \lambda_2 n\bar{x} = 0.$$

利用 $\mathbf{1}_n^T c = 1$:

$$2 + n\lambda_1 + n\lambda_2\bar{x} = 0. \tag{4}$$

（2）左乘 x^T 得:

$$2x^T c + \lambda_1 n\bar{x} + \lambda_2 x^T x = 0.$$

利用 $x^T c = 0$:

$$n\lambda_1\bar{x} + \lambda_2 x^T x = 0. \tag{5}$$

解线性方程组 (4)-(5)

由 (4):

$$\lambda_1 = -\frac{2 + n\lambda_2\bar{x}}{n}.$$

代入 (5):

$$n\bar{x}\left(-\frac{2 + n\lambda_2\bar{x}}{n}\right) + \lambda_2 x^T x = 0$$

化简:

$$-2\bar{x} - n\lambda_2\bar{x}^2 + \lambda_2 x^T x = 0$$

令 $s_{xx} = x^T x - n\bar{x}^2$，解得:

$$\lambda_2 = \frac{2\bar{x}}{s_{xx}}.$$

代回 (4) 得:

$$\lambda_1 = -\frac{2x^T x}{n s_{xx}}.$$

$$\hat{c} = -\frac{1}{2}(\lambda_1 \mathbf{1}_n + \lambda_2 x) = \frac{x^T x\,\mathbf{1}_n - n\bar{x}\,x}{n s_{xx}}.$$

利用

$$\hat{\alpha} = \hat{c}^T y, \qquad x^T y = n\bar{x}\bar{y} + s_{xy}, \qquad s_{xx} = x^T x - n\bar{x}^2,$$

代入得:

$$\begin{aligned}\hat{\alpha} &= \frac{x^T x \cdot n\bar{y} - n\bar{x}(n\bar{x}\bar{y} + s_{xy})}{n s_{xx}} \\ &= \frac{\bar{y}(x^T x - n\bar{x}^2) - \bar{x} s_{xy}}{s_{xx}} \\ &= \bar{y} - \bar{x}\frac{s_{xy}}{s_{xx}} \\ &= \bar{y} - \hat{\beta}\bar{x}.\end{aligned}$$

即:

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}.$$

$$\mathrm{Var}(\hat{\alpha}) = \sigma^2 \|\hat{c}\|_2^2.$$

带入 $\hat{c} = \frac{x^T x \mathbf{1}_n - n\bar{x}x}{n s_{xx}}$ 得:

$$\mathrm{Var}(\hat{\alpha}) = \frac{\sigma^2 x^T x}{n s_{xx}}.$$

所以

- $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$
  是满足无偏性的**线性估计量**中方差最小者，即 **BLUE**。
- 它与最小二乘法 LSE 得到的截距完全一致。
- 方差:

$$\mathrm{Var}(\hat{\alpha}) = \frac{\sigma^2 x^T x}{n(x^T x - n\bar{x}^2)}.$$

极大似然估计（MLE）：加入对σ^2的估计

一、α, β 的极大似然估计

假设线性模型为 $Y = \alpha\mathbf{1}_n + \beta x + \varepsilon$，其中 $\varepsilon \sim N(0, \sigma^2 I_n)$（即误差服从正态分布），则

$$Y \sim N(\alpha\mathbf{1}_n + \beta x, \sigma^2 I_n).$$

似然函数与对数似然函数

联合概率密度（似然函数）为：

$$L(\alpha,\beta,\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y_i-\alpha-\beta x_i)^2}{2\sigma^2}\right\} = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\|y-\alpha\mathbf{1}_n-\beta x\|_2^2\right\}.$$

取对数得对数似然函数：

$$\ell(\alpha,\beta,\sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\|y-\alpha\mathbf{1}_n-\beta x\|_2^2.$$

对 α, β 的极大化

最大化 $\ell(\alpha,\beta,\sigma^2)$ 关于 α, β 的部分，等价于最小化残差平方和：

$$\mathrm{RSS}(\alpha,\beta) = \|y-\alpha\mathbf{1}_n-\beta x\|_2^2 = \sum_{i=1}^{n}(y_i-\alpha-\beta x_i)^2.$$

这与**最小二乘法（LSE）**的目标完全一致，因此：

$$(\hat\alpha,\hat\beta)_{\mathrm{MLE}} = (\hat\alpha,\hat\beta)_{\mathrm{LSE}} = \left(\bar y - \hat\beta\bar x, \frac{s_{xy}}{s_{xx}}\right),$$

其中

$$s_{xx} = \sum_{i=1}^{n}(x_i-\bar x)^2, \quad s_{xy} = \sum_{i=1}^{n}(x_i-\bar x)(y_i-\bar y), \quad \bar x = \frac{1}{n}\sum x_i, \quad \bar y = \frac{1}{n}\sum y_i.$$

二、σ^2 的极大似然估计与无偏修正

σ^2 的 MLE

对对数似然函数关于 σ^2 求偏导并令其为 0：

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\|y - \alpha \mathbf{1}_n - \beta x\|_2^2 = 0.$$

代入 $\hat{\alpha}, \hat{\beta}$，解得：

$$\hat{\sigma}^2_{\mathrm{MLE}} = \frac{1}{n}\|y - \hat{\alpha}\mathbf{1}_n - \hat{\beta}x\|_2^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \hat{y}_i)^2,$$

其中 $\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$ 为拟合值。

1.3.2.2 偏差分析与无偏估计

计算 $\hat{\sigma}^2_{\mathrm{MLE}}$ 的期望：

$$E(\hat{\sigma}^2_{\mathrm{MLE}}) = E\left(\frac{1}{n}\sum_{i=1}^n (y_i - \hat{y}_i)^2\right).$$

利用 $\mathrm{Var}(y_i) = \sigma^2$ 和 $\mathrm{Var}(\hat{y}_i) = \mathrm{Var}(\hat{\alpha} + \hat{\beta}x_i) = \frac{\sigma^2}{n} + \frac{(x_i - \bar{x})^2}{s_{xx}}\sigma^2$，推导得：

$$E(\hat{\sigma}^2_{\mathrm{MLE}}) = \frac{n-2}{n}\sigma^2.$$

为得到无偏估计，对其修正：

$$s^2 = \frac{n}{n-2}\hat{\sigma}^2_{\mathrm{MLE}} = \frac{1}{n-2}\sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

此时 $E(s^2) = \sigma^2$，即 s^2 是 σ^2 的无偏估计。

估计量的性质

$\hat{\beta}$

$$
\begin{aligned}
\hat{\beta} = \frac{s_{xy}}{s_{xx}} = \frac{x^T y - n\bar{x}\bar{y}}{s_{xx}} = \frac{x^T y - \bar{x}\mathbf{1}_n^T y}{s_{xx}} = \frac{(x^T - \bar{x}\mathbf{1}_n^T)y}{s_{xx}} \\
&= \frac{(x^T - \bar{x}\mathbf{1}_n^T)(\alpha\mathbf{1}_n + \beta x + \varepsilon)}{s_{xx}} \\
&= \frac{1}{s_{xx}}\left[(x^T - \bar{x}\mathbf{1}_n^T)\alpha\mathbf{1}_n + (x^T - \bar{x}\mathbf{1}_n^T)\beta x + (x^T - \bar{x}\mathbf{1}_n^T)\varepsilon\right] \\
&= \frac{1}{s_{xx}}\left[n\alpha\bar{x} - \alpha\bar{x}n + \beta x^T x - \beta n\bar{x}^2 + (x^T - \bar{x}\mathbf{1}_n^T)\varepsilon\right] \\
&= \frac{1}{s_{xx}}\left[\beta s_{xx} + (x^T - \bar{x}\mathbf{1}_n^T)\varepsilon\right] \\
&= \beta + \frac{x^T - \bar{x}\mathbf{1}_n^T}{s_{xx}}\varepsilon
\end{aligned}
$$

- **无偏性**：$E(\hat{\beta}) = \beta + \frac{x^T - \bar{x}\mathbf{1}_n^T}{s_{xx}}E(\varepsilon) = \beta$（因 $E(\varepsilon) = 0$）。
- **方差**：

$$
\mathrm{Var}(\hat{\beta}) = \frac{(x^T - \bar{x}\mathbf{1}_n^T)(x - \bar{x}\mathbf{1}_n)}{s_{xx}^2}\sigma^2 = \frac{\sigma^2}{s_{xx}}
$$

2$\hat{\alpha}$

$$
\begin{aligned}
\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = \frac{1}{n}\mathbf{1}_n^T y - \bar{x}\left(\beta + \frac{x^T - \bar{x}\mathbf{1}_n^T}{s_{xx}}\varepsilon\right) \\
&= \frac{1}{n}\mathbf{1}_n^T(\alpha\mathbf{1}_n + \beta x + \varepsilon) - \beta\bar{x} - \frac{\bar{x}(x^T - \bar{x}\mathbf{1}_n^T)}{s_{xx}}\varepsilon \\
&= \alpha + \beta\bar{x} + \frac{1}{n}\mathbf{1}_n^T\varepsilon - \beta\bar{x} - \frac{\bar{x}x^T - \bar{x}^2\mathbf{1}_n^T}{s_{xx}}\varepsilon \\
&= \alpha + \left(\frac{1}{n}\mathbf{1}_n^T - \frac{\bar{x}x^T - \bar{x}^2\mathbf{1}_n^T}{s_{xx}}\right)\varepsilon
\end{aligned}
$$

- **无偏性**：$E(\hat{\alpha}) = \alpha$。
- **方差**：

$$
\begin{aligned}
\mathrm{Var}(\hat{\alpha}) &= \left(\frac{1}{n}\mathbf{1}_n^T - \frac{\bar{x}x^T - \bar{x}^2\mathbf{1}_n^T}{s_{xx}}\right)\sigma^2 I_n \left(\frac{1}{n}\mathbf{1}_n - \frac{x\bar{x} - \bar{x}^2\mathbf{1}_n}{s_{xx}}\right) \\
&= \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2 s_{xx}}{s_{xx}^2}\right) \\
&= \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}}\right)
\end{aligned}
$$

$\hat{\alpha}$ 与 $\hat{\beta}$ 的协方差

$$
\begin{aligned}
\mathrm{Cov}(\hat{\alpha}, \hat{\beta}) &= \mathrm{Cov}\left(\alpha + \left(\frac{1}{n}\mathbf{1}_n^T - \frac{\bar{x}x^T - \bar{x}^2\mathbf{1}_n^T}{s_{xx}}\right)\varepsilon, \beta + \frac{x^T - \bar{x}\mathbf{1}_n^T}{s_{xx}}\varepsilon\right) \\
&= \left(\frac{1}{n}\mathbf{1}_n^T - \frac{\bar{x}x^T - \bar{x}^2\mathbf{1}_n^T}{s_{xx}}\right)\sigma^2 I_n \frac{x - \bar{x}\mathbf{1}_n}{s_{xx}} \\
&= -\frac{\bar{x}\sigma^2}{s_{xx}}
\end{aligned}
$$

联合正态分布(joint distribution)

已知 $\varepsilon \sim N(0, \sigma^2 I_n)$，且 $\hat{\alpha}, \hat{\beta}$ 是 ε 的线性组合，故

$$
\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} \sim N\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \sigma^2 \begin{bmatrix} \frac{1}{n} + \frac{\bar{x}^2}{s_{xx}} & -\frac{\bar{x}}{s_{xx}} \\ -\frac{\bar{x}}{s_{xx}} & \frac{1}{s_{xx}} \end{bmatrix}\right)
$$

残差与估计量的独立性及卡方分布

- $(n-2)s^2 \sim \sigma^2\chi^2(n-2)$
- $\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} \perp\!\!\!\perp s^2$

$$
(n-2)s^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = e_1^2 + e_2^2 + \cdots + e_n^2
$$

其中 $e_i = y_i - \hat{y}_i$ 为残差。

> **proof:** $\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}\end{bmatrix} \perp\!\!\!\perp s^2$

$$(n-2)s^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = e_1^2 + e_2^2 + \cdots + e_n^2$$

证明 $\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}\end{bmatrix} \perp\!\!\!\perp s^2$，只需证 $e_j \perp\!\!\!\perp \begin{bmatrix}\hat{\alpha}\\ \hat{\beta}\end{bmatrix}$ $(j = 1 \to n)$

对残差 e_j 展开：

$$\begin{aligned}
e_j &= y_j - \hat{y}_j = y_j - (\hat{\alpha} + \hat{\beta}x_j) \\
&= \sum_{i=1}^{n} y_i\delta_{ij} - \left(\frac{1}{n}\sum_{i=1}^{n} y_i + \frac{(x_i - \bar{x})(x_j - \bar{x})}{s_{xx}}\sum_{i=1}^{n} y_i\right) \\
&= \sum_{i=1}^{n}\left[\delta_{ij} - \frac{1}{n} - \frac{(x_i - \bar{x})(x_j - \bar{x})}{s_{xx}}\right] y_i
\end{aligned}$$

因 $\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}\end{bmatrix}$ 是 y 的线性组合，且 $y \sim N(*,*)$，故 $\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}\\ e_j\end{bmatrix}$ 服从**联合正态分布。**

协方差为0的推导（以 $\mathrm{Cov}(\hat{\beta}, e_j)$ 为例）

$$\begin{aligned}
\mathrm{Cov}(\hat{\beta}, e_j) &= \mathrm{Cov}\left(\sum_{i=1}^{n}\frac{x_i - \bar{x}}{s_{xx}}y_i, \sum_{i=1}^{n}\left[\delta_{ij} - \frac{1}{n} - \frac{(x_i - \bar{x})(x_j - \bar{x})}{s_{xx}}\right] y_i\right) \\
&= \sum_{i=1}^{n}\frac{x_i - \bar{x}}{s_{xx}}\left(\delta_{ij} - \frac{1}{n} - \frac{(x_i - \bar{x})(x_j - \bar{x})}{s_{xx}}\right)\sigma^2 \\
&= \left[\frac{x_j - \bar{x}}{s_{xx}} - \frac{x_j - \bar{x}}{s_{xx}} \cdot \frac{1}{s_{xx}}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]\sigma^2 \\
&= \left[\frac{x_j - \bar{x}}{s_{xx}} - \frac{x_j - \bar{x}}{s_{xx}}\right]\sigma^2 = 0
\end{aligned}$$

同理可证 $\mathrm{Cov}(\hat{\alpha}, e_j) = 0$。

因联合正态分布中“协方差为0等价于独立”，故每个 e_j 均独立于 $\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}\end{bmatrix}$。

因此，

$$(n-2)s^2=\sum_{i=1}^{n}e_i^2 \quad \text{独立于} \quad \begin{bmatrix}\hat{\alpha}\\ \hat{\beta}\end{bmatrix}$$

---

proof: $(n-2)s^2\sim\sigma^2\chi^2(n-2)$

证明这个结论之前我们先证明一个基础的

单变量正态样本的方差卡方分布与独立性（基础结论）

若 $X_1,X_2,\ldots,X_n\overset{iid}{\sim}N(\mu,\sigma^2)$，令

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2,$$

则有以下结论：

1. $(n-1)s^2\sim\sigma^2\chi^2(n-1)$
2. $s^2\perp\!\!\!\perp\bar{X}$

proof:

令 $Z_i=X_i-\bar{X}$，则 $E(Z_i)=0$。
构造向量

$$Z=\begin{bmatrix}Z_1\\ Z_2\\ \vdots\\ Z_n\end{bmatrix},\quad F=\begin{bmatrix}X_1-\mu\\ X_2-\mu\\ \vdots\\ X_n-\mu\end{bmatrix}.$$

因为 $X_i\sim N(\mu,\sigma^2)$，所以 $F\sim N(0,\sigma^2I_n)$。

由 $\bar{X}=\mu+\bar{F}$（其中 $\bar{F}=\frac{1}{n}\sum_{i=1}^{n}F_i\sim N(0,\frac{\sigma^2}{n})$），可得：

$$Z=F-\bar{F}\mathbf{1}_n=F-\frac{\mathbf{1}_n\mathbf{1}_n^T}{n}F=\left(I_n-\frac{\mathbf{1}_n\mathbf{1}_n^T}{n}\right)F$$

令

$$A = I_n - \frac{\mathbf{1}_n \mathbf{1}_n^T}{n},$$

则 A 是**对称幂等矩阵**（$A^T = A$，$A^2 = A$）。

- 迹：$\mathrm{tr}(A) = \mathrm{tr}(I_n) - \mathrm{tr}\left(\frac{\mathbf{1}_n \mathbf{1}_n^T}{n}\right) = n - 1$
- 特征值：对称幂等矩阵的特征值仅为0或1，且非零特征值的个数等于迹，故 A 有 $n-1$ 个特征值1，1个特征值0。

对 A 进行正交特征分解 $A = Q^T \Lambda Q$（Λ 为对角矩阵，对角元为特征值），则：

$$Z^T Z = F^T A^T A F = F^T A F = (QF)^T \Lambda (QF)$$

因为 $QF \sim N(0, \sigma^2 I_n)$，故

$$(n-1)s^2 = Z^T Z = \sigma^2 \sum_{i=1}^{n-1} Y_i^2$$

（$Y_i \sim N(0,1)$ 且独立），即 $(n-1)s^2 \sim \sigma^2 \chi^2(n-1)$。

构造正交矩阵

$$A = \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{2}} & \cdots & \frac{-(n-1)}{\sqrt{n(n-1)}} \\ \vdots & \vdots & & \vdots \\ \frac{1}{\sqrt{n}} & -\frac{1}{\sqrt{2}} & \cdots & \frac{1}{\sqrt{n(n-1)}} \end{bmatrix},$$

令 $Y = AX$，则 $Y \sim N(0, \sigma^2 I_n)$。

其中 $Y_1 = \sqrt{n}\bar{X}$，且

$$s^2 = \frac{1}{n-1} \sum_{i=2}^{n} Y_i^2$$

由于 $Y_1, Y_2, \ldots, Y_n$ 相互独立，故

$$s^2 \perp\!\!\!\perp Y_1 \implies s^2 \perp\!\!\!\perp \bar{X}.$$

下面类似的我们给出$(n-2)s^2 \sim \sigma^2\chi^2(n-2)$的证明

proof

令残差 $e_j = y_j - \hat{y}_j$，则有：

$$\begin{aligned} e_j &= y_j - (\hat{\alpha} + \hat{\beta} x_j) \\ &= \sum_{i=1}^{n} y_i \delta_{ij} - \left(\frac{1}{n} \sum_{i=1}^{n} y_i + \frac{(x_i-\bar{x})(x_j-\bar{x})}{s_{xx}} \sum_{i=1}^{n} y_i \right) \\ &= \sum_{i=1}^{n} \left[\delta_{ij} - \frac{1}{n} - \frac{(x_i-\bar{x})(x_j-\bar{x})}{s_{xx}} \right] y_i = A_j^T y \end{aligned}$$

其中

$$A = \begin{bmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_n^T \end{bmatrix} = I_n - B, \quad B_{ij} = \frac{1}{n} + \frac{(x_i-\bar{x})(x_j-\bar{x})}{s_{xx}}.$$

remark 矩阵 B 与 A 的性质

- B 是**对称幂等矩阵**：$B^T = B$，且 $B^2 = B$（可通过展开验证：$B_{ij}^2 = \sum_{k=1}^{n} B_{ik} B_{kj} = \frac{1}{n} + \frac{(x_i-\bar{x})(x_j-\bar{x})}{s_{xx}} = B_{ij}$）。
- A 是**对称幂等矩阵**：因为 $A = I_n - B$，故 $A^T = A$ 且 $A^2 = A$。

$$\operatorname{tr}(A) = \operatorname{tr}(I_n) - \operatorname{tr}(B) = n - \left(1 + \frac{s_{xx}}{s_{xx}}\right) = n - 2$$

残差平方和：

$$(n-2)s^2 = \sum_{i=1}^{n} e_i^2 = e^T e = y^T A^T A y = \varepsilon^T A^T A \varepsilon = \varepsilon^T A \varepsilon$$

（因 A 幂等）。

对 A 进行正交特征分解 $A = Q^T \Lambda Q$，则：

$$(n-2)s^2 = \sigma^2 \sum_{i=1}^{n-2} Z_i^2, \quad Z_i \sim N(0,1) \text{ 且独立}$$

故 $(n-2)s^2 \sim \sigma^2\chi^2(n-2)$。

sum of squares

Total sum of squares (SST)

$$\text{SST} = \sum_{i=1}^{n}(y_i-\bar{y})^2$$

- $\dfrac{\text{SST}}{n-1}$: Sample variance of y_i
- $df_{\text{SST}} = n-1$

Regression sum of squares (SSR)

$$\text{SSR} = \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$$

- 由 $\sum_{i=1}^{n}(y_i-\hat{y}_i) = \sum_{i=1}^{n}(y_i-\hat{\alpha}-\hat{\beta}x_i) = 0$，可得 $\text{SSR} = \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$。
- $\dfrac{\text{SSR}}{1}$: Mean square for regression (since $df_{\text{SSR}} = 1$)
- 进一步推导:

  $$\text{SSR} = \sum_{i=1}^{n}(\hat{\alpha}+\hat{\beta}x_i-\bar{y})^2 = \sum_{i=1}^{n}(\hat{\beta}(x_i-\bar{x}))^2 = \hat{\beta}^2\sum_{i=1}^{n}(x_i-\bar{x})^2 = \hat{\beta}^2 s_{xx},$$

  其中 $s_{xx} = \sum_{i=1}^{n}(x_i-\bar{x})^2$（固定值）。
- $df_{\text{SSR}} = 1$

Residual sum of squares (SSE)

$$\mathrm{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

- 定义 $s^2 = \dfrac{\mathrm{SSE}}{n-2}$，其中 s^2 为 Mean Square Error (MSE)
- 性质：$E(s^2) = \sigma^2$（s^2 为 σ^2 的无偏估计）
- $df_{\mathrm{SSE}} = n - 2$

均方（Mean Square）

$$\text{Mean Square} = \frac{\text{Sum of Squares}}{df}$$

$\mathrm{SST} = \mathrm{SSR} + \mathrm{SSE}$

展开总平方和：

$$\begin{aligned}
\mathrm{SST} &= \sum_{i=1}^{n} (y_i - \bar{y})^2 \\
&= \sum_{i=1}^{n} \left((y_i - \hat{y}_i) + (\hat{y}_i - \bar{y}) \right)^2 \\
&= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 + 2 \sum_{i=1}^{n} (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) \\
&= \mathrm{SSE} + \mathrm{SSR} + 2 \sum_{i=1}^{n} (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}).
\end{aligned}$$

交叉项为 0 的证明

需证

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = 0.$$

从

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i, \qquad \bar{y} = \bar{\hat{y}},$$

开始：

$$\begin{aligned}
\sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) &= \sum_{i=1}^{n}(y_i - \hat{y}_i)\hat{y}_i \\
&= \sum_{i=1}^{n}(y_i - \hat{y}_i)(\hat{\alpha} + \hat{\beta} x_i) \\
&= \hat{\alpha}\sum_{i=1}^{n}(y_i - \hat{y}_i) + \hat{\beta}\sum_{i=1}^{n}(y_i - \hat{y}_i)x_i.
\end{aligned}$$

- 第一项：残差和为 0

$$\sum_{i=1}^{n}(y_i - \hat{y}_i) = 0.$$

- 第二项：展开并化简

$$\begin{aligned}
\sum_{i=1}^{n}(y_i - \hat{y}_i)x_i &= \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i(\hat{\alpha} + \hat{\beta} x_i) \\
&= \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i(\bar{y} - \hat{\beta}\bar{x} + \hat{\beta} x_i) \\
&= \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} + n\hat{\beta}\bar{x}^2 - \hat{\beta}\sum_{i=1}^{n} x_i^2 \\
&= x^T y - n\bar{x}\bar{y} - \hat{\beta} s_{xx} \\
&= s_{xy} - \frac{s_{xy}}{s_{xx}} \cdot s_{xx} = 0.
\end{aligned}$$

因此交叉项为 0，平方和分解成立：

$$\mathrm{SST} = \mathrm{SSR} + \mathrm{SSE}.$$

判定系数 r^2

$$r^2 = \frac{\mathrm{SSR}}{\mathrm{SST}} = \frac{\mathrm{SSR}}{\mathrm{SSR} + \mathrm{SSE}} \in [0, 1].$$

意义分析

- **若** $r^2 = 0$:
  - $\mathrm{SSR} = 0$，$\mathrm{SST} = \mathrm{SSE}$
  - 回归模型未能解释 y 的任何波动，拟合完全失败。
- **若** $r^2 = 1$:
  - $\mathrm{SSE} = 0$
  - 对所有 i，$\hat{y}_i = y_i$，模型完美拟合数据，拟合效果极佳。

Hypothesis Testing

检验假设

$H_0 : \beta = 0 \quad \text{vs} \quad H_1 : \beta \neq 0$

假设 H_0 下的平方和分布

总平方和（SST）$\mathrm{SST} \sim \sigma^2\chi^2(n-1).$

$$\mathrm{SST} = \|y - \bar{y}\mathbf{1}_n\|_2^2 = s_{yy}$$

其中

$$s_{yy} = y^T y - n\bar{y}^2 = y^T \left(I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T \right) y.$$

在 H_0 下，$y = \alpha\mathbf{1}_n + \varepsilon$，令

$$H = I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$$

（对称幂等矩阵），则

$$s_{yy} = \varepsilon^T H \varepsilon.$$

对 H 进行谱分解 $H = U^T \Lambda U$。因为 $\mathrm{tr}(H) = n-1$，所以 Λ 有 $n-1$ 个特征值 1 和 1 个特征值 0。又 $\varepsilon \sim N(0, \sigma^2 I_n)$，因此 $U\varepsilon \sim N(0, \sigma^2 I_n)$，从而

$$\mathrm{SST} \sim \sigma^2 \chi^2(n-1).$$

回归平方和（SSR）$\mathrm{SSR} \sim \sigma^2\chi^2(1)$

$$\mathrm{SSR} = \|\hat{y} - \bar{y}\mathbf{1}_n\|_2^2 = \hat{\beta}^2 s_{xx} = \frac{s_{xy}^2}{s_{xx}}.$$

在 H_0 下，$\hat{\beta} \sim N\left(0, \frac{\sigma^2}{s_{xx}}\right)$，因此

$$\frac{\mathrm{SSR}}{\sigma^2} = \left(\frac{\hat{\beta}}{\sigma/\sqrt{s_{xx}}}\right)^2 \sim \chi^2(1),$$

即 $\mathrm{SSR} \sim \sigma^2\chi^2(1)$。

残差平方和（SSE）

$$\mathrm{SSE} = \|y - \hat{y}\|_2^2 = (n-2)s^2,$$

其中 $s^2 = \dfrac{1}{n-2}\sum_{i=1}^n (y_i - \hat{y}_i)^2$。此前已证在线性模型假设下

$$\mathrm{SSE} \sim \sigma^2\chi^2(n-2)$$

且 SSE 与 $\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}\end{bmatrix}$ 独立。

基于 F 统计量的检验

当 $H_0 : \beta = 0$ 成立时，回归平方和 SSR 倾向于较小值（模型无解释力），而残差方差 s^2 反映误差变异。

已知：

- $\mathrm{SSR} \sim \sigma^2\chi^2(1)$;
- $(n-2)s^2 \sim \sigma^2\chi^2(n-2)$，即 $s^2 \sim \sigma^2\frac{\chi^2(n-2)}{n-2}$。

构造 F 统计量：

$$F = \frac{\mathrm{SSR}/1}{s^2} \sim F(1, n-2).$$

若 $F > F_\alpha(1, n-2)$（$F_\alpha(1, n-2)$ 为 $F(1, n-2)$ 的上 α 分位数），则**拒绝** H_0（认为 $\beta \neq 0$）；否则不拒绝 H_0。

基于 t 统计量的检验

H_0 成立时 $\hat{\beta}$ 的分布

在 $H_0 : \beta = 0$ 下，

$$\hat{\beta} \sim N\left(0, \frac{\sigma^2}{s_{xx}}\right), \qquad s_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2.$$

统计量推导

标准化：

$$\frac{\hat{\beta}}{\sqrt{\sigma^2/s_{xx}}} \sim N(0,1).$$

用 s^2 替代 σ^2（s^2 为 σ^2 的无偏估计），并结合 $(n-2)s^2/\sigma^2 \sim \chi^2(n-2)$，得到 Student t 统计量:

$$t = \frac{\hat{\beta}}{\sqrt{s^2/s_{xx}}} = \frac{\hat{\beta}}{\sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{(n-2)\,s_{xx}}}} \sim t(n-2).$$

决策规则

若 $|t| > t_{\alpha/2}(n-2)$（$t_{\alpha/2}(n-2)$ 为 t_{n-2} 的上 $\alpha/2$ 分位数），则 **拒绝** H_0；否则不拒绝。

---

remark（F 与 t 的等价性）：

在单个回归系数的检验中有如下关系（分布级别）：

$$t^2 \sim F(1, n-2),$$

因此 $|t| > t_{\alpha/2}(n-2)$ 等价于 $F > F_\alpha(1, n-2)$，两检验等价。

置信区间（CI: Confidence Interval）

统计学习里我们很关注我们对于估计量的真实值的范围，即我们很关注在得出估计量以后实际值的范围。

回归系数 β 的置信区间

统计量分布

已知 $\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{s_{xx}}\right)$，且 $\frac{(n-2)s^2}{\sigma^2} \sim \chi^2(n-2)$，构造 t 统计量:

$$\frac{(\hat{\beta}-\beta)\sqrt{s_{xx}}}{s} \sim t(n-2)$$

置信区间推导

由置信水平 $1-\alpha$，有:

$$P\left(\left|\frac{(\hat{\beta}-\beta)\sqrt{s_{xx}}}{s}\right|<t_{\alpha/2}(n-2)\right)=1-\alpha$$

整理得 β 的置信区间:

$$\hat{\beta}\pm t_{\alpha/2}(n-2)\cdot\frac{s}{\sqrt{s_{xx}}}$$

其中 $s^2=\frac{1}{n-2}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$，$s_{xx}=\sum_{i=1}^{n}(x_i-\bar{x})^2$。

均值 $\mu_0=\alpha+\beta x_0$ 的置信区间（mean response）

统计量分布

令 $\hat{\mu}_0=\hat{\alpha}+\hat{\beta}x_0$，则:

$$\hat{\mu}_0\sim N\left(\alpha+\beta x_0,\ \frac{\sigma^2}{n}+\frac{(x_0-\bar{x})^2}{s_{xx}}\sigma^2\right)$$

构造 t 统计量（用 s 替代 σ）：

$$\frac{\hat{\mu}_0-\mu_0}{\sqrt{\frac{s^2}{n}+\frac{(x_0-\bar{x})^2}{s_{xx}}s^2}}\sim t(n-2)$$

置信区间推导

由置信水平 $1-\alpha$，有:

$$P\left(\left|\frac{\hat{\mu}_0-\mu_0}{s\sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{s_{xx}}}}\right|<t_{\alpha/2}(n-2)\right)=1-\alpha$$

整理得 μ_0 的置信区间:

$$\hat{\mu}_0 \pm t_{\alpha/2}(n-2)\cdot s\sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{s_{xx}}}$$

个体预测值 $y_{n+1}=\alpha+\beta x_{n+1}+\varepsilon_{n+1}$ 的置信区间（预测区间）

统计量分布

令 $\hat{y}_{n+1}=\hat{\alpha}+\hat{\beta}x_{n+1}$，则预测误差为:

$$y_{n+1}-\hat{y}_{n+1}=(\alpha+\beta x_{n+1}+\varepsilon_{n+1})-(\hat{\alpha}+\hat{\beta}x_{n+1})$$

因 ε_{n+1} 与 $\hat{\alpha},\hat{\beta}$ 独立，且

$$\hat{\alpha}+\hat{\beta}x_{n+1}\sim N\left(\alpha+\beta x_{n+1},\ \frac{\sigma^2}{n}+\frac{(x_{n+1}-\bar{x})^2}{s_{xx}}\sigma^2\right)$$

且 $\varepsilon_{n+1}\sim N(0,\sigma^2)$，故:

$$y_{n+1}-\hat{y}_{n+1}\sim N\left(0,\ \sigma^2\left(1+\frac{1}{n}+\frac{(x_{n+1}-\bar{x})^2}{s_{xx}}\right)\right)$$

构造 t 统计量（用 s 替代 σ）：

$$\frac{y_{n+1}-\hat{y}_{n+1}}{s\sqrt{1+\frac{1}{n}+\frac{(x_{n+1}-\bar{x})^2}{s_{xx}}}}\sim t(n-2)$$

置信区间（预测区间）推导

由置信水平 $1-\alpha$，有:

$$P\left(\left|\frac{y_{n+1}-\hat{y}_{n+1}}{s\sqrt{1+\frac{1}{n}+\frac{(x_{n+1}-\bar{x})^2}{s_{xx}}}}\right|<t_{\alpha/2}(n-2)\right)=1-\alpha$$

整理得 y_{n+1} 的预测区间：

$$\hat{y}_{n+1} \pm t_{\alpha/2}(n-2) \cdot s\sqrt{1+\frac{1}{n}+\frac{(x_{n+1}-\bar{x})^2}{s_{xx}}}$$

说明：预测区间比均值的置信区间更宽，因为它不仅包含了估计 μ_0 的误差，还包含了个体误差 ε_{n+1} 的波动。