统计方法与机器学习

第二章 线性回归分析

倪 葎

Dase@ECNU

SCHOOL OF DATA SCIENCE & ENGINEERING
数据科学与工程学院

什么是线性回归分析？

- 在单因子方差分析模型中，所考虑的因子仅有 a 种不同的取值。

水平	观测到的响应变量			
1	y_{11}	y_{12}	$\cdots$	y_{1m}
2	y_{21}	y_{22}	$\cdots$	y_{2m}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
a	y_{a1}	y_{a2}	$\cdots$	y_{am}

因子	响应变量
1	y_{11}
1	y_{12}
$\vdots$	$\vdots$
1	y_{1m}
$\vdots$	$\vdots$
a	y_{am}

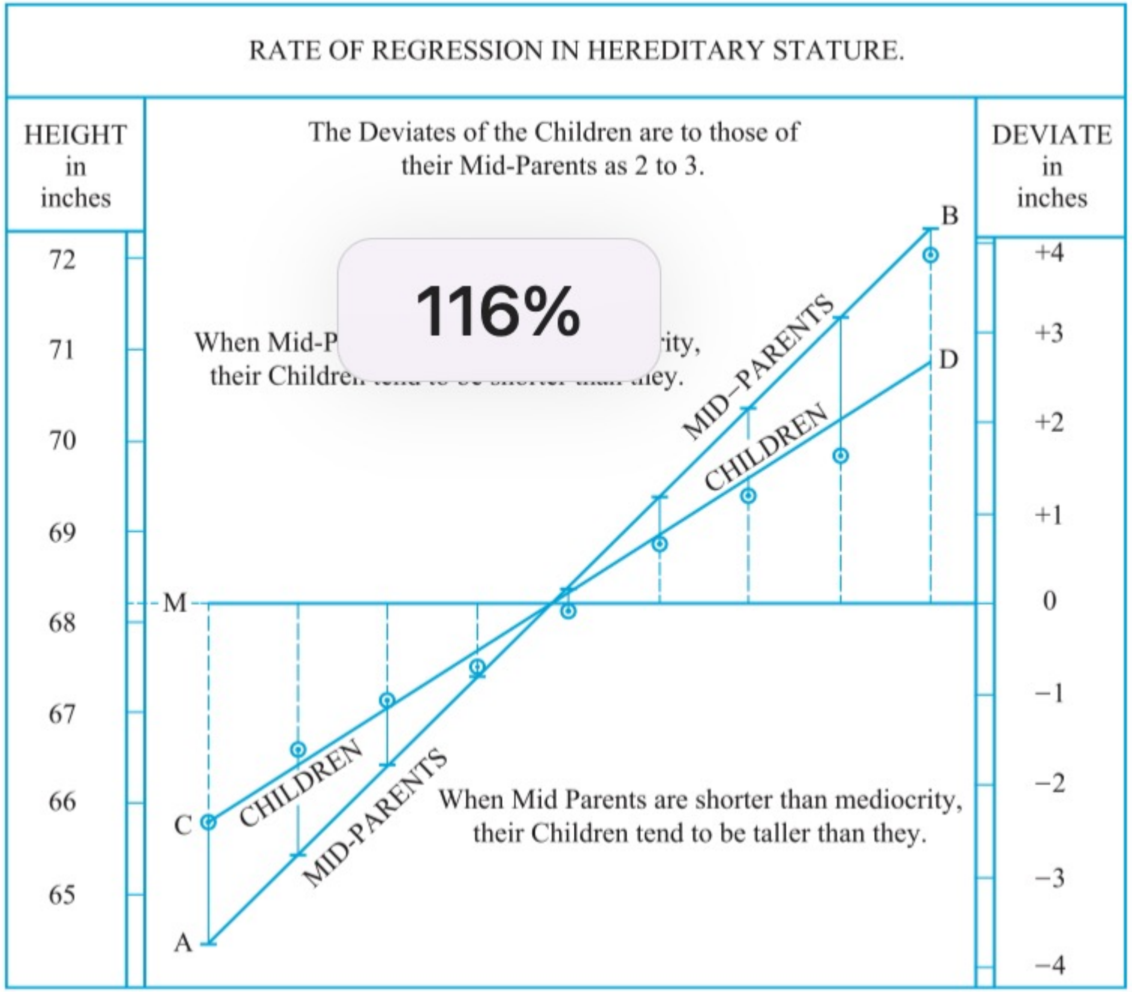
自变量
特征
x

响应变量
标签
y

- 提问：每个样本中因子的取值都不相同，模型将会变成怎样？

什么是线性回归分析？

- **回归**源于生物遗传问题的研究。



什么是线性回归分析？

- **回归**的本质是构建响应变量 y 与自变量 x 的一种统计关系，即

$$f(x) = E(y|x)$$

- 特例：
  - $f(x) = \beta_0 + \beta_1 x$
  - $f(x) = \max(0, \beta_0 + \beta_1 x)$
  - $f(x) = g_1 \circ g_2 \circ \cdots \circ g_n(\beta_0 + \beta_1 x)$
- 应用场景：
  - 信用卡授信金额与申请人年龄、职业、收入、消费偏好有关。
  - 房屋价格与房屋面积、卧室数量、地段、建造时间、交通便利有关。
  - 景区的游览人数与天气情况、是否为节假日、景区类型等有关。

目录

目录

定义

- 线性回归模型为

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon$$

- 称 y 为响应变量或标签，通常认为是一个连续型随机变量。
- 称 x_j 为第 j 个自变量或特征，通常认为是确定性的变量。
- 称 β_j 为回归系数，共有 $p+1$ 个未知参数。
- 称 ε 为随机误差，并假定

$$E(\varepsilon) = 0,$$

$$\mathrm{Var}(\varepsilon) = \sigma^2.$$

数据

- 共有 n 组观测数据

$$\{(x_{i1}, x_{i2}, \cdots, x_{ip}, y_i): i = 1,2, \cdots, n\}$$

- 于是有

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_p x_{1p} + \varepsilon_1 \\ \vdots \\ y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i \\ \vdots \\ y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_p x_{np} + \varepsilon_n \end{cases}$$

数据

$\boldsymbol{y} = (y_1, y_2, \cdots, y_n)^{\top}$

$\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2 \cdots, \beta_p)^{\top}$

$\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)^{\top}$

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_p x_{1p} + \varepsilon_1 \\ \quad\vdots \\ y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i \\ \quad\vdots \\ y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_p x_{np} + \varepsilon_n \end{cases}$$

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

- 矩阵形式为

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

基本假定

- 设计矩阵 $\boldsymbol{X}$ 是确定性的，且 $\operatorname{rank}(\boldsymbol{X}) = p + 1 < n$。
  - 在本课程中， $\boldsymbol{X}$ 是确定性的，而非随机性的；
  - $\boldsymbol{X}$ 是一个列满秩矩阵。
- 随机误差 $\boldsymbol{\varepsilon}$ 是零均值且等方差的，即
  - $E(\varepsilon_i) = 0, i = 1,2,\cdots,n$
  - $\operatorname{Cov}(\varepsilon_i, \varepsilon_j) = \begin{cases} \sigma^2, & i = j \\ 0, & i \neq j \end{cases}$
  - 称为**高斯-马尔可夫条件**。

基本假定

- 进一步，假定随机误差服从正态分布，即

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$$

  - 期望：$E(\boldsymbol{\varepsilon}) = \mathbf{0}$；
  - 方差-协方差矩阵： $\mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}_n$。

- 因为

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

所以，

$$\boldsymbol{y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n)$$

  - 期望：$E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta}$；
  - 方差-协方差矩阵： $\mathrm{Var}(\boldsymbol{y}) = \sigma^2 \boldsymbol{I}_n$。

总结

- 多元线性回归模型为

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

- 基本假定1：$\boldsymbol{X}$ 是确定性的，且 rank($\boldsymbol{X}$)=p+1<n。
- 基本假定2： $\boldsymbol{\varepsilon}$ 是零均值且等方差。
- 基本假定3：$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$

目录

问题定义

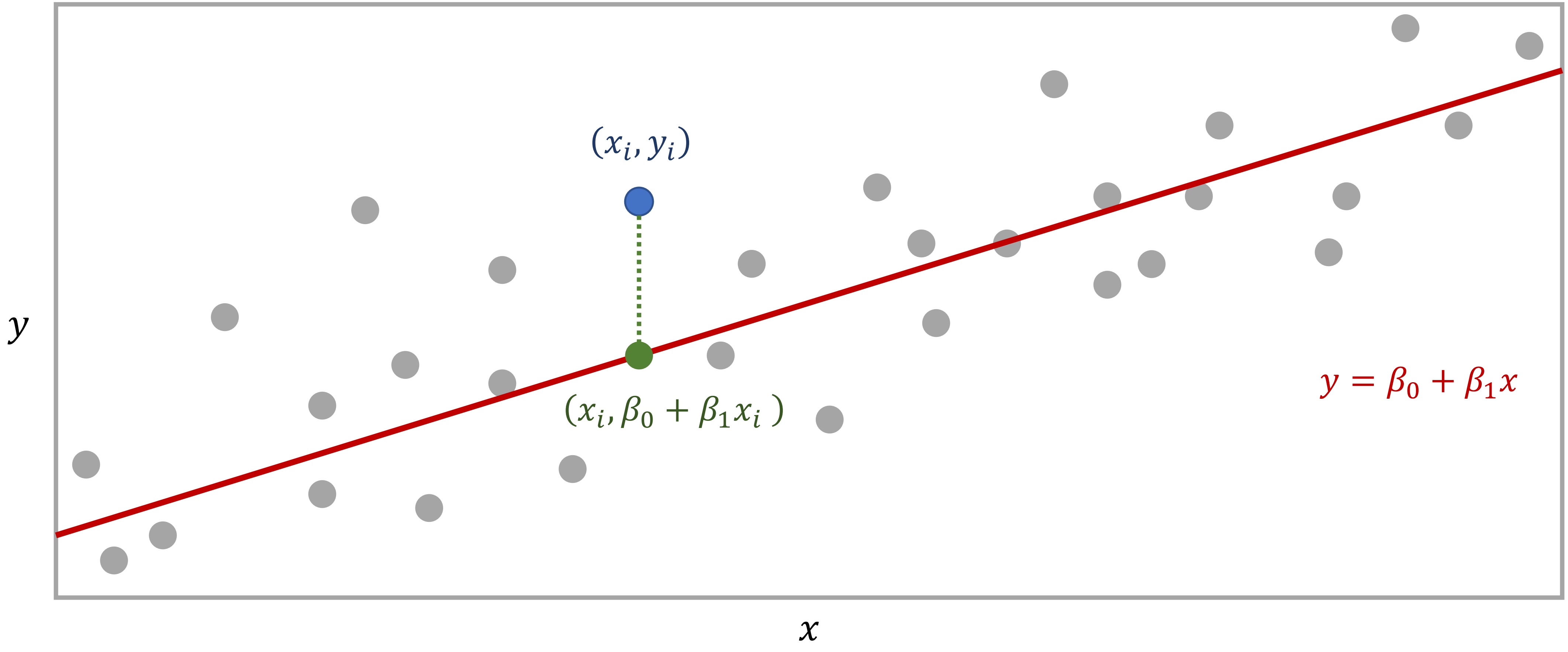
- 多元线性回归模型为

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

- **待估计参数**
  - 回归系数 $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2 \cdots, \beta_p)^{\top}$；
  - 误差方差 σ^2；
- **估计方法**
  - 最小二乘估计
  - 极大似然估计

最小二乘估计

- **基本思想：拟合**



最小二乘估计

- **基本思想：拟合**
- 线性回归模型

$$E(y|\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{\beta} = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

- 第 i 个数据点的实际观测值为 $(\boldsymbol{x}_i, y_i)$
- 第 i 个数据点的拟合值为 $\left(\boldsymbol{x}_i, \boldsymbol{x}_i^\top \boldsymbol{\beta}\right)$
- 差异为

$$y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}$$

最小二乘估计

- 损失函数定义为

$$Q(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}\right)^2$$

- 最小二乘估计为

$$\widehat{\boldsymbol{\beta}}_{\mathrm{LS}} = \mathrm{argmin}_{\boldsymbol{\beta}} Q(\boldsymbol{\beta})$$

最小二乘估计

- **求解过程**
  - 损失函数的另一种形式

$$
\begin{aligned}
Q(\boldsymbol{\beta}) &= \sum_{i=1}^{n}\left(y_i - \boldsymbol{x}_i^{\top}\boldsymbol{\beta}\right)^2 \\
&= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^{\top}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\
&= (\boldsymbol{X}\boldsymbol{\beta})^{\top}(\boldsymbol{X}\boldsymbol{\beta}) - 2(\boldsymbol{X}\boldsymbol{\beta})^{\top}\boldsymbol{y} + \boldsymbol{y}^{\top}\boldsymbol{y} \\
&= \underbrace{\boxed{\boldsymbol{\beta}^{\top}\boldsymbol{X}^{\top}\boldsymbol{X}\boldsymbol{\beta}}}_{\text{二次型}} - \underbrace{\boxed{2\boldsymbol{\beta}^{\top}\boldsymbol{X}^{\top}\boldsymbol{y}}}_{\text{线性}} + \underbrace{\boxed{\boldsymbol{y}^{\top}\boldsymbol{y}}}_{\text{无关}}
\end{aligned}
$$

最小二乘估计

- **求解过程**
  - 求导，即

$$\frac{\partial Q(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{\partial(\boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} - 2\boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{y} + \boldsymbol{y}^\top \boldsymbol{y})}{\partial \boldsymbol{\beta}} = 2\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} - 2\boldsymbol{X}^\top \boldsymbol{y}$$

  - 令 $\frac{\partial Q(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0}$，可得

$$\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}^\top \boldsymbol{y}$$

  - 根据基本假设1可知， $\boldsymbol{X}^\top \boldsymbol{X}$ 是满秩的。
  - 最小二乘估计为

$$\hat{\boldsymbol{\beta}}_{\mathrm{LS}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

知识回顾——向量求导

假定

- $\boldsymbol{x}$ 是一个 p 维向量
- $\boldsymbol{a}$是一个 p 维向量
- $\boldsymbol{B}$ 是 $p \times p$ 矩阵

结论

- 线性：

$$\frac{\partial(\boldsymbol{x}^\top \boldsymbol{a})}{\partial \boldsymbol{x}} = \boldsymbol{a}$$

- 二次型：

$$\frac{\partial(\boldsymbol{x}^\top \boldsymbol{B}\boldsymbol{x})}{\partial \boldsymbol{x}} = (\boldsymbol{B} + \boldsymbol{B}^\top)\boldsymbol{x}$$

最小二乘估计

- 最小二乘估计为

$$\hat{\boldsymbol{\beta}}_{\mathrm{LS}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

- 反思：为什么 基本假定1：$\mathrm{rank}(\boldsymbol{X}) = p + 1 < n$ 是重要的？
- 在最小二乘估计中，矩阵 $\boldsymbol{X}^\top \boldsymbol{X}$ 需要求逆矩阵。
- 因为 $\mathrm{rank}(\boldsymbol{X}^\top \boldsymbol{X}) \le \mathrm{rank}(\boldsymbol{X})$，基本假定1能够保证 $\mathrm{rank}(\boldsymbol{X}^\top \boldsymbol{X}) = p + 1$。
- 基本假定1的直观解释
  - 特征之间不（完全）线性相关；
  - 样本量需要（远）大于特征个数。

最小二乘估计

- 最小二乘估计为

$$\hat{\boldsymbol{\beta}}_{\mathrm{LS}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

- 反思：为什么 基本假定1：$\mathrm{rank}(\boldsymbol{X}) = p + 1 < n$ 是重要的？
- 在最小二乘估计中，矩阵 $\boldsymbol{X}^\top \boldsymbol{X}$ 需要求逆矩阵。
- 因为 $\mathrm{rank}(\boldsymbol{X}^\top \boldsymbol{X}) \leq \mathrm{rank}(\boldsymbol{X})$，基本假定1能够保证 $\mathrm{rank}(\boldsymbol{X}^\top \boldsymbol{X}) = p + 1$。
- 基本假定1的直观解释
  - 特征之间不（完全）线性相关；
  - 样本量需要（远）大于特征个数。

极大似然估计

- **基本思想：似然**

- 极大似然估计依赖于数据分布假定，即

$$\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I}_n)$$

- $\boldsymbol{y}$ 的联合密度函数为

$$p(\boldsymbol{y}; \boldsymbol{\beta}, \sigma^2) = (2\pi)^{-n/2} |\sigma^2 \boldsymbol{I}_n|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\beta})^\top (\sigma^2 \boldsymbol{I}_n)^{-1} (\boldsymbol{y} - \boldsymbol{X\beta})\right\}$$

- $(\boldsymbol{\beta}, \sigma^2)$ 的似然函数为

$$L(\boldsymbol{\beta}, \sigma^2) = (2\pi)^{-p/2} |\sigma^2 \boldsymbol{I}_n|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\beta})^\top (\sigma^2 \boldsymbol{I}_n)^{-1} (\boldsymbol{y} - \boldsymbol{X\beta})\right\}$$

极大似然估计

- **基本思想：似然**
- $(\boldsymbol{\beta}, \sigma^2)$ 的似然函数为

$$L(\boldsymbol{\beta}, \sigma^2) = (2\pi)^{-n/2} |\sigma^2 \boldsymbol{I}_n|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\beta})^\top (\sigma^2 \boldsymbol{I}_n)^{-1} (\boldsymbol{y} - \boldsymbol{X\beta})\right\}$$

$$|\sigma^2 \boldsymbol{I}_n| = \begin{vmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{vmatrix} = (\sigma^2)^n$$

$$(\sigma^2 \boldsymbol{I}_n)^{-1} = \begin{pmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{pmatrix}^{-1} = \sigma^{-2} \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}$$

- 极大似然估计为

$$\begin{aligned} \left(\widehat{\boldsymbol{\beta}}_{\text{MLE}}, \hat{\sigma}^2_{\text{MLE}}\right) &= \operatorname{argmax} L(\boldsymbol{\beta}, \sigma^2) \\ &= \operatorname{argmax} \ \ln(L(\boldsymbol{\beta}, \sigma^2)) \end{aligned}$$

极大似然估计

- **求解过程**
  - 对数似然函数为

$$\ln(L(\boldsymbol{\beta},\sigma^2)) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})$$

  - 对 $\boldsymbol{\beta}$ 和 σ^2 求偏导，即

$$\begin{cases} \dfrac{\partial\ln(L(\boldsymbol{\beta},\sigma^2))}{\partial\boldsymbol{\beta}} = -\dfrac{1}{\sigma^2}(\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta}-\boldsymbol{X}^\top\boldsymbol{y}) \\ \dfrac{\partial\ln(L(\boldsymbol{\beta},\sigma^2))}{\partial\sigma^2} = -\dfrac{n}{2\sigma^2} + \dfrac{n}{2(\sigma^2)^2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) \end{cases}$$

极大似然估计

- **求解过程**
  - 极大似然估计为

$$
\begin{cases}
\widehat{\boldsymbol{\beta}}_{\mathrm{MLE}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y} \\
\widehat{\boldsymbol{\sigma}}^2{}_{\mathrm{MLE}} = \dfrac{1}{n} \left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\mathrm{MLE}}\right)^\top \left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\mathrm{MLE}}\right)
\end{cases}
$$

- 说明
  - $\boldsymbol{\beta}$ 的最小二乘估计与极大似然估计是一致的，记 $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$ 。
  - $\widehat{\boldsymbol{\sigma}}^2{}_{\mathrm{MLE}}$ 不是无偏估计，但是相合估计。

参数估计的性质

定理2.5 在多元线性回归模型中，$\boldsymbol{\beta}$ 的估计为

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

有以下两个重要性质：

- $\widehat{\boldsymbol{\beta}}$ 的期望为

$$E\left(\widehat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta}$$

- $\widehat{\boldsymbol{\beta}}$ 的方差-协方差矩阵为

$$\mathrm{Var}\left(\widehat{\boldsymbol{\beta}}\right) = \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1}$$

参数估计的性质

定理2.5 在多元线性回归模型中，$\boldsymbol{\beta}$ 的估计为

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

有以下两个重要性质：

- $\hat{\boldsymbol{\beta}}$ 的期望为

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

- $\hat{\boldsymbol{\beta}}$ 的方差-协方差矩阵为

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1}$$

参数估计的性质

定理2.5的证明

$$
\begin{aligned}
E(\hat{\boldsymbol{\beta}}) &= E((\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}) \\
&= (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top E(\boldsymbol{y}) \\
&= (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top E(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\
&= (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top (\boldsymbol{X}\boldsymbol{\beta} + E(\boldsymbol{\varepsilon})) \\
&= (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\beta}
\end{aligned}
$$

知识回顾

因为

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

所以，

- 期望：$E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta}$；
- 方差-协方差矩阵：$\mathrm{Var}(\boldsymbol{y}) = \sigma^2 \boldsymbol{I}_n$。

参数估计的性质

定理2.5 在多元线性回归模型中，$\boldsymbol{\beta}$ 的估计为

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}\boldsymbol{X}^{\top}\boldsymbol{y}$$

有以下两个重要性质：

- $\widehat{\boldsymbol{\beta}}$ 的期望为

$$E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

- $\widehat{\boldsymbol{\beta}}$ 的方差-协方差矩阵为

$$\mathrm{Var}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}$$

参数估计的性质

定理2.5的证明

$$
\begin{aligned}
\mathrm{Var}(\hat{\boldsymbol{\beta}}) &= \mathrm{Var}((\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y}) \\
&= (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \mathrm{Var}(\boldsymbol{y})((\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top)^\top \\
&= (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \mathrm{Var}(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})((\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top)^\top \\
&= (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top (\sigma^2 \boldsymbol{I}_n)\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \\
&= \sigma^2(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \\
&= \sigma^2(\boldsymbol{X}^\top \boldsymbol{X})^{-1}
\end{aligned}
$$

知识回顾

因为

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

所以，

- 期望：$E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta}$；
- 方差-协方差矩阵：$\mathrm{Var}(\boldsymbol{y}) = \sigma^2\boldsymbol{I}_n$。

参数估计的几何解释

- $\boldsymbol{y}$ 的拟合值为

$$\begin{aligned}\hat{\boldsymbol{y}} &= \boldsymbol{X}\hat{\boldsymbol{\beta}} \\ &= \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y}\end{aligned}$$

- 定义

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$$

为**帽子矩阵**（hat matrix）。

- 于是，

$$\hat{\boldsymbol{y}} = \boldsymbol{H}\boldsymbol{y}$$

参数估计的几何解释

命题2.1 帽子矩阵

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$$

有以下三个重要性质：

- $\boldsymbol{H}$ 是 n 阶对称矩阵，即 $\boldsymbol{H} = \boldsymbol{H}^\top$；
- $\boldsymbol{H}$ 是幂等矩阵，即 $\boldsymbol{H} = \boldsymbol{H}^2$；
- $\boldsymbol{H}$ 的迹为 $p + 1$，即 $\mathrm{tr}(\boldsymbol{H}) = p + 1$。

参数估计的几何解释

命题2.1 帽子矩阵

$$\boldsymbol{H}=\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$$

有以下三个重要性质：

- $\boldsymbol{H}$ 是 n 阶对称矩阵，即 $\boldsymbol{H}=\boldsymbol{H}^\top$；
- $\boldsymbol{H}$ 是幂等矩阵，即 $\boldsymbol{H}=\boldsymbol{H}^2$；
- $\boldsymbol{H}$ 的迹为 $p+1$，即 $\mathrm{tr}(\boldsymbol{H})=p+1$。

参数估计的性质

命题2.1的证明

$$
\begin{aligned}
\boldsymbol{H}^\top &= (\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)^\top \\
&= (\boldsymbol{X}^\top)^\top\left((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\right)^\top\boldsymbol{X}^\top \\
&= \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top = \boldsymbol{H}
\end{aligned}
$$

参数估计的几何解释

命题2.1 帽子矩阵

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}\boldsymbol{X}^{\top}$$

有以下三个重要性质：

- $\boldsymbol{H}$ 是 n 阶对称矩阵，即 $\boldsymbol{H} = \boldsymbol{H}^{\top}$；
- $\boldsymbol{H}$ 是幂等矩阵，即 $\boldsymbol{H} = \boldsymbol{H}^{2}$；
- $\boldsymbol{H}$ 的迹为 $p+1$，即 $\mathrm{tr}(\boldsymbol{H}) = p+1$。

参数估计的性质

命题2.1的证明

$$
\begin{aligned}
\boldsymbol{H}^2 &= \left(\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\right)^2 \\
&= \left(\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\right)\left(\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\right) \\
&= \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}(\boldsymbol{X}^\top\boldsymbol{X})(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top \\
&= \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top = \boldsymbol{H}
\end{aligned}
$$

参数估计的几何解释

命题2.1 帽子矩阵

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$$

有以下三个重要性质：

- $\boldsymbol{H}$ 是 n 阶对称矩阵，即 $\boldsymbol{H} = \boldsymbol{H}^\top$；
- $\boldsymbol{H}$ 是幂等矩阵，即 $\boldsymbol{H} = \boldsymbol{H}^2$；
- $\boldsymbol{H}$ 的迹为 $p+1$，即 $\mathrm{tr}(\boldsymbol{H}) = p+1$。

参数估计的性质

命题2.1的证明

$$
\begin{aligned}
\operatorname{tr}(\boldsymbol{H}) &= \operatorname{tr}(\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top) \\
&= \operatorname{tr}((\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}) \\
&= \operatorname{tr}(\boldsymbol{I}_{p+1}) = p+1
\end{aligned}
$$

参数估计的几何解释

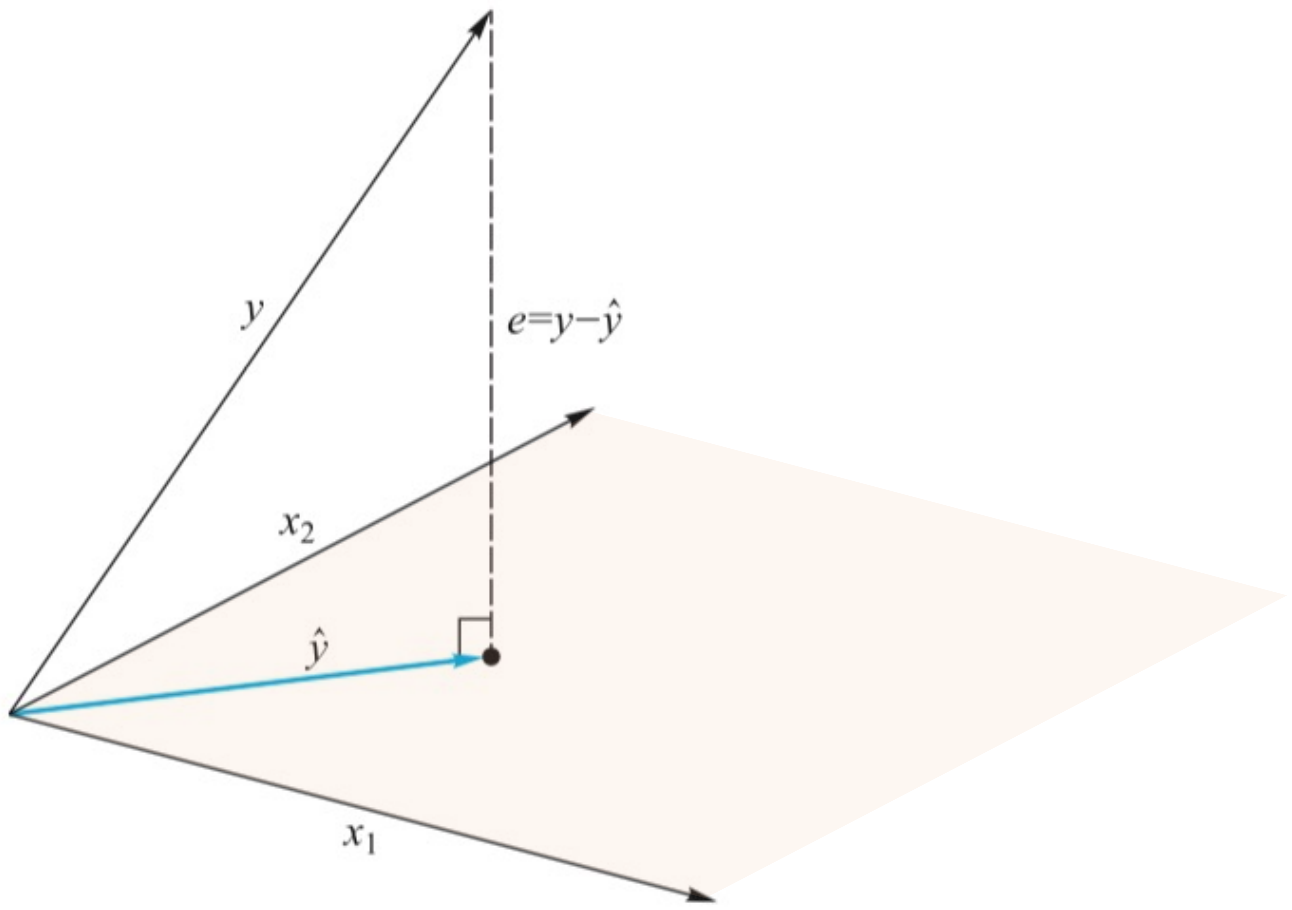
- 残差定义为

$$\begin{aligned} \boldsymbol{e} &= \boldsymbol{y} - \widehat{\boldsymbol{y}} \\ &= \boldsymbol{y} - \boldsymbol{H}\boldsymbol{y} \\ &= (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y} \end{aligned}$$

- 几何关系：拟合值 $\widehat{\boldsymbol{y}}$ 与残差 $\boldsymbol{e}$ **垂直**，即

$$\begin{aligned} \widehat{\boldsymbol{y}}^\top \boldsymbol{e} &= (\boldsymbol{H}\boldsymbol{y})^\top (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y} \\ &= \boldsymbol{y}^\top \boldsymbol{H}^\top (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y} = 0 \end{aligned}$$

参数估计的几何解释



参数估计的几何解释

- **残差的性质**

- 残差的期望

$$E(\boldsymbol{e}) = E\big((\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y}\big) = (\boldsymbol{I} - \boldsymbol{H})E(\boldsymbol{y}) = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{0}$$

- 残差的方差-协方差矩阵

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{e}) &= \mathrm{Cov}(\boldsymbol{e}, \boldsymbol{e}) \\
&= \mathrm{Cov}((\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y}, (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y}) \\
&= (\boldsymbol{I} - \boldsymbol{H})\mathrm{Cov}(\boldsymbol{y}, \boldsymbol{y})(\boldsymbol{I} - \boldsymbol{H})^\top \\
&= \sigma^2(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{I}_n(\boldsymbol{I} - \boldsymbol{H})^\top \\
&= \sigma^2(\boldsymbol{I} - \boldsymbol{H})
\end{aligned}$$

参数估计的几何解释

- 提问：能否得到 σ^2 的无偏估计？

- 一般而言，采用

$$\hat{\sigma}^2 = \frac{1}{n-(p+1)}\boldsymbol{e}^\top\boldsymbol{e}$$

作为 σ^2 的无偏估计。

总结

- 回归系数 $\boldsymbol{\beta}$ 的估计为

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

- 随机误差的方差 σ^2 的估计
  - 无偏估计

$$\hat{\sigma}^2 = \frac{1}{n-(p+1)} \boldsymbol{e}^\top \boldsymbol{e}$$

  - 极大似然估计

$$\hat{\sigma}^2_{\mathrm{MLE}} = \frac{1}{n} \boldsymbol{e}^\top \boldsymbol{e}$$

目录

回顾：最小二乘估计

- 多元线性回归模型

$$y=\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p+\varepsilon$$

- 现有数据 $\{(x_{i1},x_{i2},\cdots,x_{ip},y_i): i=1,2,\cdots,n\}$
- 可以得到回归系数的估计

$$\begin{aligned}\widehat{\boldsymbol{\beta}}&=\left(\hat{\beta}_0,\hat{\beta}_1,\hat{\beta}_2\cdots,\hat{\beta}_p\right)^\top\\&=\left(\hat{\beta}_0,\widehat{\boldsymbol{\beta}}_1^\top\right)^\top\end{aligned}$$

- 定义

$$\boldsymbol{y}=(y_1,\cdots,y_n)^\top,\boldsymbol{X}=(1_n,x_1,x_2,\cdots,x_p)=(1_n,\boldsymbol{X}_o)$$

回顾：最小二乘估计

- 回归系数的估计为

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$$

$$= \left((\mathbf{1}_n, \boldsymbol{X}_o)^\top(\mathbf{1}_n, \boldsymbol{X}_o)\right)^{-1}(\mathbf{1}_n, \boldsymbol{X}_o)^\top\boldsymbol{y}$$

$$= \begin{pmatrix} \mathbf{1}_n^\top\mathbf{1}_n & \mathbf{1}_n^\top\boldsymbol{X}_o \\ \boldsymbol{X}_o^\top\mathbf{1}_n, & \boldsymbol{X}_o^\top\boldsymbol{X}_o \end{pmatrix}^{-1}\begin{pmatrix} \mathbf{1}_n^\top \\ \boldsymbol{X}_o^\top \end{pmatrix}\boldsymbol{y}$$

$$\mathbf{1}_n^\top\mathbf{1}_n = (1 \quad \cdots \quad 1)\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = n$$

$$= \begin{pmatrix} n^{-1} + n^{-2}\mathbf{1}_n^\top\boldsymbol{X}_o\boldsymbol{A}_o\boldsymbol{X}_o^\top\mathbf{1}_n & -n^{-1}\mathbf{1}_n^\top\boldsymbol{X}_o\boldsymbol{A}_o \\ -n^{-1}\boldsymbol{A}_o\boldsymbol{X}_o^\top\mathbf{1}_n & \boldsymbol{A}_o \end{pmatrix}\begin{pmatrix} \mathbf{1}_n^\top \\ \boldsymbol{X}_o^\top \end{pmatrix}\boldsymbol{y}$$

$$= \begin{pmatrix} n^{-1}\mathbf{1}_n^\top + n^{-2}\mathbf{1}_n^\top\boldsymbol{X}_o\boldsymbol{A}_o\boldsymbol{X}_o^\top\mathbf{1}_n\mathbf{1}_n^\top - n^{-1}\mathbf{1}_n^\top\boldsymbol{X}_o\boldsymbol{A}_o\boldsymbol{X}_o^\top \\ -n^{-1}\boldsymbol{A}_o\boldsymbol{X}_o^\top\mathbf{1}_n\mathbf{1}_n^\top + \boldsymbol{A}_o\boldsymbol{X}_o^\top \end{pmatrix}\boldsymbol{y}$$

其中，$\boldsymbol{A}_o = (\boldsymbol{X}_o^\top\boldsymbol{X}_o - n^{-1}\boldsymbol{X}_o^\top\mathbf{1}_n\mathbf{1}_n^\top\boldsymbol{X}_o)^{-1}$

知识回顾——分块矩阵求逆

$\boldsymbol{A}$ 是可逆矩阵，如果 $\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}$ 可逆，那么有

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{E}_{11} & \boldsymbol{E}_{12} \\ \boldsymbol{E}_{21} & \boldsymbol{E}_{22} \end{pmatrix}$$

- $\boldsymbol{E}_{11} = \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{E}_{22}\boldsymbol{C}\boldsymbol{A}^{-1}$
- $\boldsymbol{E}_{12} = -\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{E}_{22}$
- $\boldsymbol{E}_{21} = -\boldsymbol{E}_{22}\,\boldsymbol{C}\boldsymbol{A}^{-1}$
- $\boldsymbol{E}_{22} = (\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}$

问题定义

		原始数据	中心化	标准化
数据变换	响应变量	$\boldsymbol{y}=(y_1,\cdots,y_n)^\top$	$\boldsymbol{y}^*=(y_1^*,\cdots,y_n^*)^\top$ $y_i^*=y_i-\bar{y},\bar{y}=n^{-1}\sum_{i=1}^n y_i$	$\boldsymbol{y}^{**}=(y_1^{**},\cdots,y_n^{**})^\top$ $y_i^{**}=\frac{y_i^*}{\sqrt{l_{yy}}}, l_{yy}=\sum_{i=1}^n (y_i^*)^2$
	自变量	$\boldsymbol{X}=(\boldsymbol{1}_n,\boldsymbol{X}_\mathrm{o})$, $\boldsymbol{X}_\mathrm{o}=\{x_{ij}\}_{n\times p}$	$\boldsymbol{X}^*=(\boldsymbol{1}_n,\boldsymbol{X}_\mathrm{c}),\boldsymbol{X}_\mathrm{c}=\{x_{ij}^*\}_{n\times p}$ $x_{ij}^*=x_{ij}-\bar{x}_j,\bar{x}_j=n^{-1}\sum_{i=1}^n x_{ij}$	$\boldsymbol{X}^{**}=(\boldsymbol{1}_n,\boldsymbol{X}_\mathrm{s}),\boldsymbol{X}_\mathrm{s}=\{x_{ij}^{**}\}_{n\times p}$ $x_{ij}^{**}=\frac{x_{ij}^*}{\sqrt{l_{jj}}}, l_{jj}=\sum_{i=1}^n \left(x_{ij}^*\right)^2$
参数估计	β_0	$\hat{\beta}_0$	$\hat{\beta}_{\mathrm{c},0}=?$	$\hat{\beta}_{\mathrm{s},0}=?$
	$\begin{pmatrix}\beta_1\\ \vdots \\ \beta_p\end{pmatrix}$	$\hat{\boldsymbol{\beta}}_1$	$\hat{\boldsymbol{\beta}}_{\mathrm{c},1}=?$	$\hat{\boldsymbol{\beta}}_{\mathrm{s},1}=?$

中心化

- 回归系数的估计为

$$\widehat{\boldsymbol{\beta}}_{\mathrm{c}} = \left((\boldsymbol{X}^*)^\top \boldsymbol{X}^*\right)^{-1} (\boldsymbol{X}^*)^\top \boldsymbol{y}^*$$

$$= \left((\mathbf{1}_n, \boldsymbol{X}_{\mathrm{c}})^\top (\mathbf{1}_n, \boldsymbol{X}_{\mathrm{c}})\right)^{-1} (\mathbf{1}_n, \boldsymbol{X}_{\mathrm{c}})^\top \boldsymbol{y}^*$$

$$= \begin{pmatrix} \mathbf{1}_n^\top \mathbf{1}_n & \mathbf{1}_n^\top \boldsymbol{X}_{\mathrm{c}} \\ \boldsymbol{X}_{\mathrm{c}}^\top \mathbf{1}_n, & \boldsymbol{X}_{\mathrm{c}}^\top \boldsymbol{X}_{\mathrm{c}} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{1}_n^\top \\ \boldsymbol{X}_{\mathrm{c}}^\top \end{pmatrix} \boldsymbol{y}^*$$

$$= \begin{pmatrix} n^{-1} + n^{-2} \mathbf{1}_n^\top \boldsymbol{X}_{\mathrm{c}} \boldsymbol{A}_{\mathrm{c}} \boldsymbol{X}_{\mathrm{c}}^\top \mathbf{1}_n & -n^{-1} \mathbf{1}_n^\top \boldsymbol{X}_{\mathrm{c}} \boldsymbol{A}_{\mathrm{c}} \\ -n^{-1} \boldsymbol{A}_{\mathrm{c}} \boldsymbol{X}_{\mathrm{c}}^\top \mathbf{1}_n & \boldsymbol{A}_{\mathrm{c}} \end{pmatrix} \begin{pmatrix} \mathbf{1}_n^\top \\ \boldsymbol{X}_{\mathrm{c}}^\top \end{pmatrix} \boldsymbol{y}^*$$

$$= \begin{pmatrix} n^{-1} \mathbf{1}_n^\top + n^{-2} \mathbf{1}_n^\top \boldsymbol{X}_{\mathrm{c}} \boldsymbol{A}_{\mathrm{c}} \boldsymbol{X}_{\mathrm{c}}^\top \mathbf{1}_n \mathbf{1}_n^\top - n^{-1} \mathbf{1}_n^\top \boldsymbol{X}_{\mathrm{c}} \boldsymbol{A}_{\mathrm{c}} \boldsymbol{X}_{\mathrm{c}}^\top \\ -n^{-1} \boldsymbol{A}_{\mathrm{c}} \boldsymbol{X}_{\mathrm{c}}^\top \mathbf{1}_n \mathbf{1}_n^\top + \boldsymbol{A}_{\mathrm{c}} \boldsymbol{X}_{\mathrm{c}}^\top \end{pmatrix} \boldsymbol{y}^*$$

其中，$\boldsymbol{A}_{\mathrm{c}} = (\boldsymbol{X}_{\mathrm{c}}^\top \boldsymbol{X}_{\mathrm{c}} - n^{-1} \boldsymbol{X}_{\mathrm{c}}^\top \mathbf{1}_n \mathbf{1}_n^\top \boldsymbol{X}_{\mathrm{c}})^{-1}$

知识回顾——分块矩阵求逆

$\boldsymbol{A}$ 是可逆矩阵，如果 $\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}$ 可逆，那么有

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{E}_{11} & \boldsymbol{E}_{12} \\ \boldsymbol{E}_{21} & \boldsymbol{E}_{22} \end{pmatrix}$$

- $\boldsymbol{E}_{11} = \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{E}_{22}\boldsymbol{C}\boldsymbol{A}^{-1}$
- $\boldsymbol{E}_{12} = -\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{E}_{22}$
- $\boldsymbol{E}_{21} = -\boldsymbol{E}_{22}\boldsymbol{C}\boldsymbol{A}^{-1}$
- $\boldsymbol{E}_{22} = (\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}$

中心化

- **中心化的本质**
  - 响应变量

$$\boldsymbol{y}^* = \boldsymbol{y} - \mathbf{1}_n(\mathbf{1}_n^\top\mathbf{1}_n)^{-1}\mathbf{1}_n^\top\,\boldsymbol{y} = \left(\boldsymbol{I}_n - \boldsymbol{H}_{\mathbf{1}_n}\right)\boldsymbol{y}$$

  - 自变量

$$\boldsymbol{X}_\mathrm{c} = \boldsymbol{X}_\mathrm{o} - \mathbf{1}_n(\mathbf{1}_n^\top\mathbf{1}_n)^{-1}\mathbf{1}_n^\top\,\boldsymbol{X}_\mathrm{o} = \left(\boldsymbol{I}_n - \boldsymbol{H}_{\mathbf{1}_n}\right)\boldsymbol{X}_\mathrm{o}$$

  - 提问：为什么 $\boldsymbol{H}_{\mathbf{1}_n} = \mathbf{1}_n(\mathbf{1}_n^\top\mathbf{1}_n)^{-1}\mathbf{1}_n^\top$ 是一个对称幂等矩阵？

中心化

- 估计 β_0 的影响

$$\hat{\beta}_{c,0} = (n^{-1}\mathbf{1}_n^\top + n^{-2}\mathbf{1}_n^\top \boldsymbol{X}_c \boldsymbol{A}_c \boldsymbol{X}_c^\top \mathbf{1}_n \mathbf{1}_n^\top - n^{-1}\mathbf{1}_n^\top \boldsymbol{X}_c \boldsymbol{A}_c \boldsymbol{X}_c^\top)\boldsymbol{y}^*$$

$$= \left(n^{-1}\mathbf{1}_n^\top + n^{-2}\mathbf{1}_n^\top\left(\boldsymbol{I}_n - \boldsymbol{H}_{\mathbf{1}_n}\right)\boldsymbol{X}_o \boldsymbol{A}_c \boldsymbol{X}_c^\top \mathbf{1}_n \mathbf{1}_n^\top - n^{-1}\mathbf{1}_n^\top\left(\boldsymbol{I}_n - \boldsymbol{H}_{\mathbf{1}_n}\right)\boldsymbol{X}_o \boldsymbol{A}_c \boldsymbol{X}_c^\top\right)\boldsymbol{y}^*$$

$$= n^{-1}\mathbf{1}_n^\top \boldsymbol{y}^*$$

$$= n^{-1}\mathbf{1}_n^\top\left(\boldsymbol{I}_n - \boldsymbol{H}_{\mathbf{1}_n}\right)\boldsymbol{y}$$

$$= 0$$

计算提示

$$\mathbf{1}_n^\top\left(\boldsymbol{I}_n - \boldsymbol{H}_{\mathbf{1}_n}\right) = \mathbf{1}_n^\top - \mathbf{1}_n^\top \boldsymbol{H}_{\mathbf{1}_n} = \mathbf{1}_n^\top - \mathbf{1}_n^\top \mathbf{1}_n (\mathbf{1}_n^\top \mathbf{1}_n)^{-1} \mathbf{1}_n^\top = \mathbf{1}_n^\top - \mathbf{1}_n^\top = \mathbf{0}$$

中心化

- 估计 $\beta_1,\cdots,\beta_p$ 的影响

$$\hat{\beta}_{c,1} = (-n^{-1}\boldsymbol{A}_c\boldsymbol{X}_c^\top\boldsymbol{1}_n\boldsymbol{1}_n^\top + \boldsymbol{A}_c\boldsymbol{X}_c^\top)\boldsymbol{y}^*$$

$$= \left(-n^{-1}\boldsymbol{A}_c\boldsymbol{X}_o^\top(\boldsymbol{I}_n - \boldsymbol{H}_{\boldsymbol{1}_n})\boldsymbol{1}_n\boldsymbol{1}_n^\top + \boldsymbol{A}_c\boldsymbol{X}_c^\top\right)\boldsymbol{y}^*$$

$$= \boldsymbol{A}_c\boldsymbol{X}_c^\top\boldsymbol{y}^*$$

$$= \boldsymbol{A}_o\boldsymbol{X}_o^\top(\boldsymbol{I}_n - \boldsymbol{H}_{\boldsymbol{1}_n})\boldsymbol{y}$$

$$\hat{\beta}_1 = (-n^{-1}\boldsymbol{A}_o\boldsymbol{X}_o^\top\boldsymbol{1}_n\boldsymbol{1}_n^\top + \boldsymbol{A}_o\boldsymbol{X}_o^\top)\boldsymbol{y}$$

$$= \boldsymbol{A}_o\boldsymbol{X}_o^\top\left(\boldsymbol{I}_n - \boldsymbol{1}_n(\boldsymbol{1}_n^\top\boldsymbol{1}_n)^{-1}\boldsymbol{1}_n^\top\right)\boldsymbol{y}$$

$$= \boldsymbol{A}_o\boldsymbol{X}_o^\top\left(\boldsymbol{I}_n - \boldsymbol{H}_{\boldsymbol{1}_n}\right)\boldsymbol{y}$$

计算提示

$$\boldsymbol{A}_c = (\boldsymbol{X}_c^\top\boldsymbol{X}_c - n^{-1}\boldsymbol{X}_c^\top\boldsymbol{1}_n\boldsymbol{1}_n^\top\boldsymbol{X}_c)^{-1} = \left(\boldsymbol{X}_o^\top(\boldsymbol{I}_n - \boldsymbol{H}_{\boldsymbol{1}_n})\boldsymbol{X}_o - n^{-1}\boldsymbol{X}_o^\top(\boldsymbol{I}_n - \boldsymbol{H}_{\boldsymbol{1}_n})\boldsymbol{1}_n\boldsymbol{1}_n^\top(\boldsymbol{I}_n - \boldsymbol{H}_{\boldsymbol{1}_n})\boldsymbol{X}_o\right)^{-1}$$

$$= \left(\boldsymbol{X}_o^\top(\boldsymbol{I}_n - \boldsymbol{H}_{\boldsymbol{1}_n})\boldsymbol{X}_o\right)^{-1} = \boldsymbol{A}_o$$

小结

		原始数据	中心化	标准化
数据变换	响应变量	$\boldsymbol{y}=(y_1,\cdots,y_n)^{\top}$	$\boldsymbol{y}^*=(y_1^*,\cdots,y_n^*)^{\top}$ $y_i^*=y_i-\bar{y},\bar{y}=n^{-1}\sum_{i=1}^{n}y_i$	$\boldsymbol{y}^{**}=(y_1^{**},\cdots,y_n^{**})^{\top}$ $y_i^{**}=\frac{y_i^*}{\sqrt{l_{yy}}},l_{yy}=\sum_{i=1}^{n}(y_i^*)^2$
	自变量	$\boldsymbol{X}=(\mathbf{1}_n,\boldsymbol{X}_{\mathrm{o}})$, $\boldsymbol{X}_{\mathrm{o}}=\{x_{ij}\}_{n\times p}$	$\boldsymbol{X}^*=(\mathbf{1}_n,\boldsymbol{X}_{\mathrm{c}}),\boldsymbol{X}_{\mathrm{c}}=\{x_{ij}^*\}_{n\times p}$ $x_{ij}^*=x_{ij}-\bar{x}_j,\bar{x}_j=n^{-1}\sum_{i=1}^{n}x_{ij}$	$\boldsymbol{X}^{**}=(\mathbf{1}_n,\boldsymbol{X}_{\mathrm{s}}),\boldsymbol{X}_{\mathrm{s}}=\{x_{ij}^{**}\}_{n\times p}$ $x_{ij}^{**}=\frac{x_{ij}^*}{\sqrt{l_{jj}}},l_{jj}=\sum_{i=1}^{n}(x_{ij}^*)^2$
参数估计	β_0	$\hat{\beta}_0$	$\hat{\beta}_{\mathrm{c},0}=0$	$\hat{\beta}_{\mathrm{s},0}=?$
	$\boldsymbol{\beta}_1$	$\hat{\boldsymbol{\beta}}_1$	$\hat{\boldsymbol{\beta}}_{\mathrm{c},1}=\hat{\boldsymbol{\beta}}_1$	$\hat{\boldsymbol{\beta}}_{\mathrm{s},1}=?$

标准化

- 定义

$$L = \begin{pmatrix} \frac{1}{\sqrt{l_{11}}} & & \\ & \ddots & \\ & & \frac{1}{\sqrt{l_{pp}}} \end{pmatrix}$$

- 于是有

$$\boldsymbol{X}_s = \boldsymbol{X}_c \boldsymbol{L}$$

$$\boldsymbol{y}^{**} = \frac{1}{\sqrt{l_{yy}}} \boldsymbol{y}^{*}$$

标准化

- 估计 $\boldsymbol{\beta}_1 = (\beta_1, \cdots, \beta_p)^\top$ 的影响

$$\widehat{\boldsymbol{\beta}}_{\mathrm{s},1} = (\boldsymbol{X}_\mathrm{s}^\top \boldsymbol{X}_\mathrm{s})^{-1} \boldsymbol{X}_\mathrm{s}^\top \boldsymbol{y}^{**}$$

$$= (\boldsymbol{L}\boldsymbol{X}_\mathrm{c}^\top \boldsymbol{X}_\mathrm{c}\boldsymbol{L})^{-1} \boldsymbol{L}\boldsymbol{X}_\mathrm{c}^\top \left(\frac{1}{\sqrt{l_{yy}}} \boldsymbol{y}^*\right)$$

$$= \boldsymbol{L}^{-1}(\boldsymbol{X}_\mathrm{c}^\top \boldsymbol{X}_\mathrm{c})^{-1} \cancel{\boldsymbol{L}^{-1}\boldsymbol{L}} \boldsymbol{X}_\mathrm{c}^\top \left(\frac{1}{\sqrt{l_{yy}}} \boldsymbol{y}^*\right)$$

$$= \frac{1}{\sqrt{l_{yy}}} \boldsymbol{L}^{-1} (\boldsymbol{X}_\mathrm{c}^\top \boldsymbol{X}_\mathrm{c})^{-1} \boldsymbol{X}_\mathrm{c}^\top \boldsymbol{y}^*$$

$$\widehat{\boldsymbol{\beta}}_{\mathrm{c},1} = (-n^{-1}\boldsymbol{A}_\mathrm{c}\boldsymbol{X}_\mathrm{c}^\top 1_n 1_n^\top + \boldsymbol{A}_\mathrm{c}\boldsymbol{X}_\mathrm{c}^\top)\boldsymbol{y}^*$$

$$= \boldsymbol{A}_\mathrm{c}\boldsymbol{X}_\mathrm{c}^\top \boldsymbol{y}^*$$

$$= (\boldsymbol{X}_\mathrm{c}^\top \boldsymbol{X}_\mathrm{c} - n^{-1}\boldsymbol{X}_\mathrm{c}^\top 1_n 1_n^\top \boldsymbol{X}_\mathrm{c})^{-1}\ \boldsymbol{X}_\mathrm{c}^\top \boldsymbol{y}^*$$

$$= (\boldsymbol{X}_\mathrm{c}^\top \boldsymbol{X}_\mathrm{c})^{-1}\ \boldsymbol{X}_\mathrm{c}^\top \boldsymbol{y}^*$$

标准化

- 估计 $\boldsymbol{\beta}_1 = (\beta_1, \cdots, \beta_p)^\top$ 的影响

$$\widehat{\boldsymbol{\beta}}_{\mathrm{s},1} = \frac{1}{\sqrt{l_{yy}}}\boldsymbol{L}^{-1}\widehat{\boldsymbol{\beta}}_{\mathrm{c},1} = \frac{1}{\sqrt{l_{yy}}}\boldsymbol{L}^{-1}\widehat{\boldsymbol{\beta}}_1$$

其分量为

$$\hat{\beta}_{\mathrm{s}j} = \frac{\sqrt{l_{jj}}}{\sqrt{l_{yy}}}\hat{\beta}_{\mathrm{c}j} = \frac{\sqrt{l_{jj}}}{\sqrt{l_{yy}}}\hat{\beta}_j, j = 1,2,\cdots,p$$

总结

		原始数据	中心化	标准化
数据变换	响应变量	$\boldsymbol{y}=(y_1,\cdots,y_n)^\top$	$\boldsymbol{y}^*=(y_1^*,\cdots,y_n^*)^\top$ $y_i^*=y_i-\bar{y},\bar{y}=n^{-1}\sum_{i=1}^n y_i$	$\boldsymbol{y}^{**}=(y_1^{**},\cdots,y_n^{**})^\top$ $y_i^{**}=\frac{y_i^*}{\sqrt{l_{yy}}},l_{yy}=\sum_{i=1}^n(y_i^*)^2$
数据变换	自变量	$\boldsymbol{X}=(\boldsymbol{1}_n,\boldsymbol{X}_\mathrm{o})$, $\boldsymbol{X}_\mathrm{o}=\{x_{ij}\}_{n\times p}$	$\boldsymbol{X}^*=(\boldsymbol{1}_n,\boldsymbol{X}_\mathrm{c}),\boldsymbol{X}_\mathrm{c}=\{x_{ij}^*\}_{n\times p}$ $x_{ij}^*=x_{ij}-\bar{x}_j,\bar{x}_j=n^{-1}\sum_{i=1}^n x_{ij}$	$\boldsymbol{X}^{**}=(\boldsymbol{1}_n,\boldsymbol{X}_\mathrm{s}),\boldsymbol{X}_\mathrm{s}=\{x_{ij}^{**}\}_{n\times p}$ $x_{ij}^{**}=\frac{x_{ij}^*}{\sqrt{l_{jj}}},l_{jj}=\sum_{i=1}^n(x_{ij}^*)^2$
参数估计	β_0	$\hat{\beta}_0$	$\hat{\beta}_{\mathrm{c},0}=0$	$\hat{\beta}_{\mathrm{s},0}=0$
参数估计	$\begin{pmatrix}\beta_1\\\vdots\\\beta_p\end{pmatrix}$	$\hat{\boldsymbol{\beta}}_1$	$\hat{\boldsymbol{\beta}}_{\mathrm{c},1}=\hat{\boldsymbol{\beta}}_1$	$\hat{\boldsymbol{\beta}}_{\mathrm{s},1}=\frac{1}{\sqrt{l_{yy}}}\begin{pmatrix}\sqrt{l_{11}}&&\\&\ddots&\\&&\sqrt{l_{pp}}\end{pmatrix}\hat{\boldsymbol{\beta}}_1$

谢谢

SCHOOL OF DATA
SCIENCE & ENGINEERING
数据科学与工程学院