统计方法与机器学习

第二章 线性回归分析

倪 葎

Dase@ECNU

Da E

Data Science
& Engineering

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

目录

目录

问题定义

- 线性回归模型

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

- 线性回归模型中的显著性水平检验：
  - F 检验：用于检验**回归方程**的显著性；
  - t 检验：用于检验**回归系数**的显著性；
- 此外，还需要评估线性回归的**拟合优度**。

F 检验

- 线性回归模型

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

- **检验的本质**：判断自变量 $x_1, x_2, \cdots, x_p$ 从整体上对响应变量 y 是否有明显的影响。
- 假设问题

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0 \text{ vs } H_1: \text{存在} \beta_j \text{ 不为零，} j = 1,2, \cdots, p.$$

- 当 H_0 为真，则表明响应变量 y 与 $x_1, x_2, \cdots, x_p$ 之间的关系采用多元线性回归模型来刻画并不合适的。

F 检验

- 多元线性回归模型的方差分解公式

拟合值 $\hat{y}_i = \boldsymbol{x}_i^{\top}\hat{\boldsymbol{\beta}}$

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

残差 $e_i = y_i - \hat{y}_i$

记为

$$SS_T = SS_R + SS_E$$

总偏差平方和

回归平方和

残差平方和

- 提问：如何构造该检验统计量？

F 检验

定理2.7 在正态分布假定下，即

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

有

1. 残差平方和

$$\frac{SS_E}{\sigma^2} \sim \chi^2(n-p-1)\text{；}$$

2. 残差平方和 SS_E 和回归平方和 SS_R 相互独立；
3. 在原假设 H_0 成立下，

$$\frac{SS_R}{\sigma^2} \sim \chi^2(p).$$

F 检验

定理2.7 在正态分布假定下，即

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

有

1. 残差平方和

$$\frac{SS_E}{\sigma^2} \sim \chi^2(n-p-1)\text{；}$$

2. 残差平方和 SS_E 和回归平方和 SS_R 相互独立；
3. 在原假设 H_0 成立下，

$$\frac{SS_R}{\sigma^2} \sim \chi^2(p).$$

F 检验

- **定理2.7的证明**

由于残差

$$\boldsymbol{e} = \boldsymbol{y} - \widehat{\boldsymbol{y}} = \boldsymbol{y} - \boldsymbol{H}\boldsymbol{y} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y}$$

于是残差平方和为

$$SS_E = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \boldsymbol{e}^\top\boldsymbol{e} = \boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y}$$

F 检验

- **定理2.7的证明**

由于

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

并且 $\boldsymbol{I}-\boldsymbol{H}$ 是对称幂等矩阵，其秩 $\mathrm{rank}(\boldsymbol{I}-\boldsymbol{H}) = n-(p+1)$。

因此，

$$\frac{SS_E}{\sigma^2} = \frac{\boldsymbol{y}^\top(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{y}}{\sigma^2} \sim \chi^2(n-p-1)$$

注意到，

$$\delta = \frac{(\boldsymbol{X\beta})^\top(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{X\beta}}{\sigma^2} = 0$$

知识回顾

设 $\boldsymbol{x} \sim N_n(\boldsymbol{\mu}, \sigma^2 \boldsymbol{I}_n)$。

如果 $\boldsymbol{C}$ 是一个对称矩阵

且 $\mathrm{rank}(\boldsymbol{C}) = r \le n$，

那么

$$\frac{\boldsymbol{x}^\top \boldsymbol{C}\boldsymbol{x}}{\sigma^2} \sim \chi^2(r, \delta)$$

其中，

$$\delta = \frac{\boldsymbol{\mu}^\top \boldsymbol{C}\boldsymbol{\mu}}{\sigma^2}$$

当且仅当

$$\boldsymbol{C}^2 = \boldsymbol{C}$$

F 检验

定理2.7 在正态分布假定下，即

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

有

1. 残差平方和

$$\frac{SS_E}{\sigma^2} \sim \chi^2(n-p-1)\,;$$

2. 残差平方和 SS_E 和回归平方和 SS_R 相互独立；
3. 在原假设 H_0 成立下，

$$\frac{SS_R}{\sigma^2} \sim \chi^2(p).$$

F 检验

- **定理2.7的证明**

由于回归平方和

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

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

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \begin{array}{l} \to \boldsymbol{x}_1^\top \\ \to \boldsymbol{x}_2^\top \\ \quad\vdots \\ \to \boldsymbol{x}_n^\top \end{array}$$

$$\frac{\partial Q}{\partial \beta_0} = -2\sum_{i=1}^{n}\left(y_i - \boldsymbol{x}_i^\top\boldsymbol{\beta}\right) = 0$$

$$\Downarrow$$

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

F 检验

- **定理2.7的证明**

由于回归平方和

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

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

$$= \widehat{\boldsymbol{\beta}}^\top \sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top \widehat{\boldsymbol{\beta}}$$

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

$$\boldsymbol{x}_i^\top\widehat{\boldsymbol{\beta}} - \overline{\boldsymbol{x}}^\top\widehat{\boldsymbol{\beta}} = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}^\top(\boldsymbol{x}_i - \overline{\boldsymbol{x}})$$

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

其中

$$\boldsymbol{A} = \sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top$$

F 检验

- **定理2.7的证明**

由于回归平方和

$$SS_R = \boldsymbol{y}^\top \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{A} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

又因为残差平方和

$$SS_E = \boldsymbol{y}^\top (\boldsymbol{I} - \boldsymbol{H}) \boldsymbol{y}$$

因此，

$$SS_E \text{ 与 } SS_R \text{ 独立。}$$

知识回顾

设 $\boldsymbol{x} \sim N_n(\boldsymbol{\mu}, \sigma^2 \boldsymbol{I}_n)$。

如果 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 均是对称矩阵，那么

$$\boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} \text{ 和 } \boldsymbol{x}^\top \boldsymbol{B} \boldsymbol{x} \text{ 独立}$$

当且仅当

$$\boldsymbol{A}\boldsymbol{B} = \boldsymbol{0}_{n\times n}$$

F 检验

定理2.7 在正态分布假定下，即

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

有

1. 残差平方和

$$\frac{SS_E}{\sigma^2} \sim \chi^2(n-p-1)\text{；}$$

2. 残差平方和 SS_E 和回归平方和 SS_R 相互独立；

3. 在原假设 H_0 成立下，

$$\frac{SS_R}{\sigma^2} \sim \chi^2(p).$$

F 检验

- **定理2.7的证明**

因为

$$\overline{\boldsymbol{x}}^\top = (\mathbf{1}_n^\top \mathbf{1}_n)^{-1} \mathbf{1}_n^\top \boldsymbol{X}$$

我们有

$$\begin{aligned}
\boldsymbol{A} &= \sum_{i=1}^{n} (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^\top \\
&= \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top - n\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^\top \\
&= \boldsymbol{X}^\top \boldsymbol{X} - (\mathbf{1}_n^\top \mathbf{1}_n) \boldsymbol{X}^\top \mathbf{1}_n (\mathbf{1}_n^\top \mathbf{1}_n)^{-1} (\mathbf{1}_n^\top \mathbf{1}_n)^{-1} \mathbf{1}_n^\top \boldsymbol{X} \\
&= \boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{X}^\top \mathbf{1}_n (\mathbf{1}_n^\top \mathbf{1}_n)^{-1} \mathbf{1}_n^\top \boldsymbol{X} \\
&= \boldsymbol{X}^\top (\boldsymbol{I}_n - \boldsymbol{H}_{\mathbf{1}_n}) \boldsymbol{X}
\end{aligned}$$

F 检验

- **定理2.7的证明**

所以

$$SS_R = \boldsymbol{y}^\top \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{A}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y}$$

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

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

F 检验

- **定理2.7的证明**

由于 $\boldsymbol{X} = (\mathbf{1}_n \quad \boldsymbol{X}_\text{o})$，因此，注意到

$$\boldsymbol{H}\boldsymbol{H}_{\mathbf{1}_n} = (\mathbf{1}_n \quad \boldsymbol{X}_\text{o})\begin{pmatrix} n^{-1} + n^{-2}\mathbf{1}_n^\top \boldsymbol{X}_\text{o}\boldsymbol{A}_\text{o}\boldsymbol{X}_\text{o}^\top \mathbf{1}_n & -n^{-1}\mathbf{1}_n^\top \boldsymbol{X}_\text{o}\boldsymbol{A}_\text{o} \\ -n^{-1}\boldsymbol{A}_\text{o}\boldsymbol{X}_\text{o}^\top \mathbf{1}_n & \boldsymbol{A}_\text{o} \end{pmatrix}\begin{pmatrix} \mathbf{1}_n^\top \\ \boldsymbol{X}_\text{o}^\top \end{pmatrix}\boldsymbol{H}_{\mathbf{1}_n}$$

$$= (n^{-1}\mathbf{1}_n\mathbf{1}_n^\top + n^{-2}\mathbf{1}_n\mathbf{1}_n^\top \boldsymbol{X}_\text{o}\boldsymbol{A}_\text{o}\boldsymbol{X}_\text{o}^\top \mathbf{1}_n\mathbf{1}_n^\top - n^{-1}\mathbf{1}_n\mathbf{1}_n^\top \boldsymbol{X}_\text{o}\boldsymbol{A}_\text{o}\boldsymbol{X}_\text{o}^\top - n^{-1}\boldsymbol{X}_\text{o}\boldsymbol{A}_\text{o}\boldsymbol{X}_\text{o}^\top \mathbf{1}_n\mathbf{1}_n^\top + \boldsymbol{X}_\text{o}\boldsymbol{A}_\text{o}\boldsymbol{X}_\text{o}^\top)\boldsymbol{H}_{\mathbf{1}_n}$$

$$= \left(n^{-1}\mathbf{1}_n\mathbf{1}_n^\top + \boldsymbol{H}_{\mathbf{1}_n}\boldsymbol{X}_\text{o}\boldsymbol{A}_\text{o}\boldsymbol{X}_\text{o}^\top \boldsymbol{H}_{\mathbf{1}_n} - \boldsymbol{H}_{\mathbf{1}_n}\boldsymbol{X}_\text{o}\boldsymbol{A}_\text{o}\boldsymbol{X}_\text{o}^\top - \boldsymbol{X}_\text{o}\boldsymbol{A}_\text{o}\boldsymbol{X}_\text{o}^\top \boldsymbol{H}_{\mathbf{1}_n} + \boldsymbol{X}_\text{o}\boldsymbol{A}_\text{o}\boldsymbol{X}_\text{o}^\top\right)\boldsymbol{H}_{\mathbf{1}_n}$$

$$= (n^{-1}\mathbf{1}_n\mathbf{1}_n^\top)^2 = \boldsymbol{H}_{\mathbf{1}_n}$$

计算提示

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

F 检验

- **定理2.7的证明**

所以

$$
\begin{aligned}
SS_R &= \boldsymbol{y}^\top \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{A} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y} \\
&= \boldsymbol{y}^\top \underbrace{\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top}_{\boldsymbol{H}} \left(\boldsymbol{I} - \boldsymbol{H}_{\boldsymbol{1}_n}\right) \underbrace{\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top}_{\boldsymbol{H}} \boldsymbol{y} \\
&= \boldsymbol{y}^\top \boldsymbol{H} \left(\boldsymbol{I} - \boldsymbol{H}_{\boldsymbol{1}_n}\right) \boldsymbol{H} \boldsymbol{y} \\
&= \boldsymbol{y}^\top \left(\boldsymbol{H} - \boldsymbol{H}_{\boldsymbol{1}_n}\right) \boldsymbol{y}
\end{aligned}
$$

F 检验

- **定理2.7的证明**

令

$$\boldsymbol{B} = \left(\boldsymbol{H} - \boldsymbol{H}_{\mathbf{1}_n}\right)$$

易证

- $\boldsymbol{B}$ 是对称矩阵；
- $\boldsymbol{B}$ 是幂等矩阵；
- $\mathrm{rank}(\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{H}) - \mathrm{tr}\left(\boldsymbol{H}_{\mathbf{1}_n}\right) = (p+1) - 1 = p.$

F 检验

- **定理2.7的证明**

在 H_0 成立时，$\beta_1=\beta_2=\cdots=\beta_p=0$，有

$$\boldsymbol{y}\sim N(\beta_0\mathbf{1}_n,\sigma^2\boldsymbol{I}_n)$$

所以，

$$\frac{SS_R}{\sigma^2}=\frac{\boldsymbol{y}^\top\boldsymbol{B}\boldsymbol{y}}{\sigma^2}\sim\chi^2(p)$$

其中，

$$\delta=\frac{(\beta_0\mathbf{1}_n)^\top\boldsymbol{B}(\beta_0\mathbf{1}_n)}{\sigma^2}=\frac{\beta_0^2}{\sigma^2}\mathbf{1}_n^\top\boldsymbol{B}\mathbf{1}_n=\frac{\beta_0^2}{\sigma^2}\mathrm{tr}(\mathbf{1}_n^\top\boldsymbol{B}\mathbf{1}_n)=\frac{\beta_0^2}{\sigma^2}\mathrm{tr}(\boldsymbol{B}\mathbf{1}_n\mathbf{1}_n^\top)$$

$$=\frac{\beta_0^2}{\sigma^2}\mathrm{tr}\left(n\boldsymbol{B}\boldsymbol{H}_{\mathbf{1}_n}\right)=\frac{n\beta_0^2}{\sigma^2}\mathrm{tr}\left((\boldsymbol{H}-\boldsymbol{H}_{\mathbf{1}_n})\boldsymbol{H}_{\mathbf{1}_n}\right)=0$$

知识回顾

设 $\boldsymbol{x}\sim N_n(\boldsymbol{\mu},\sigma^2\boldsymbol{I}_n)$。

如果 $\boldsymbol{C}$ 是一个对称矩阵

且 $\mathrm{rank}(\boldsymbol{C})=r\le n$，

那么

$$\frac{\boldsymbol{x}^\top\boldsymbol{C}\boldsymbol{x}}{\sigma^2}\sim\chi^2(r,\delta)$$

其中，

$$\delta=\frac{\boldsymbol{\mu}^\top\boldsymbol{C}\boldsymbol{\mu}}{\sigma^2}$$

当且仅当

$$\boldsymbol{C}^2=\boldsymbol{C}$$

F 检验

定理2.7 在正态分布假定下，即

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

有

1. 残差平方和

$$\frac{SS_E}{\sigma^2} \sim \chi^2(n-p-1)\text{；}$$

2. 残差平方和 SS_E 和回归平方和 SS_R 相互独立；
3. 在原假设 H_0 成立下，

$$\frac{SS_R}{\sigma^2} \sim \chi^2(p).$$

F 检验

- 检验统计量为

$$F_0 = \frac{SS_R/p}{SS_E/(n-p-1)} \sim F(p, n-p-1)$$

取显著性水平为 α

- **临界值法**
  - 如果 $F_0 > F_{1-\alpha}(p, n-p-1)$，那么拒绝原假设；反之，接受原假设。
- **p 值法**
  - 如果 $p_0 = P(F \geq F_0) < \alpha$，那么拒绝原假设；反之，接受原假设。
- 拒绝原假设，认为回归方程时显著的。

F 检验

方差分析表

来源	平方和	自由度	均方	F 值	p 值
回归	SS_R	p	$\frac{SS_R}{p}$	$F_0 = \frac{SS_R/p}{SS_E/(n-p-1)}$	$p_0 = P(F \geq F_0)$
误差	SS_E	$n-p-1$	$\frac{SS_E}{n-p-1}$		
总和	SS_T	$n-1$			

t 检验

- **目的**：需要进一步判断每一个变量是否显著。
- 检验问题

$$H_{0j}: \beta_j = 0, j = 1,2,\cdots,p$$

  - 如果拒绝原假设，认为第 j 个特征是显著的；
  - 如果接受原假设，认为第 j 个特征是不显著的；

回顾

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

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

F 检验

定理2.7 在正态分布假定下，即

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

有

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})\boldsymbol{X}^\top \boldsymbol{y} \sim N(\boldsymbol{\beta}, \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1})$$

- **重要结论**

$$\hat{\beta}_j \sim N\left(\beta_j, c_{jj}\sigma^2\right)$$

其中，c_{jj} 是矩阵 $(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$ 第 $(j+1)$ 个对角线元素，称为方差扩大因子。

知识回顾

设随机向量

$$\boldsymbol{x} = (x_1, x_2, \cdots, x_p)^\top$$

服从正态分布 $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，其中

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix}$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{pmatrix}$$

则每一个分量

$$x_j \sim N\left(\mu_j, \sigma_{jj}\right)$$

t 检验

- 检验统计量为

$$\frac{\hat{\beta}_j}{\sqrt{c_{jj}}\hat{\sigma}}$$

其中，

$$\hat{\sigma}^2 = SS_E/(n-p-1)$$

- 在 H_0 成立时，

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{c_{jj}}\hat{\sigma}} \sim t(n-p-1)$$

t 检验

取显著性水平为 α

- **临界值法**
  - 如果 $|t_j| > t_{1-\alpha/2}(n-p-1)$，那么拒绝原假设；反之，接受原假设。
- **p 值法**
  - 如果 $p_0 = 2P\left(t \geq |t_j|\right) < \alpha$，那么拒绝原假设；反之，接受原假设。
- 拒绝原假设，认为回归系数时显著的。

复相关系数

- **拟合优度**是用于度量回归方程对观测值的拟合程度。
- 定义**样本决定系数**为

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T}$$

- 可证明： 在多元线性回归模型中，$0 \leq R^2 \leq 1$。
  - R^2 越接近 1，回归拟合效果越好；
  - R^2 越接近 0，回归拟合效果越差；
- 虽然 R^2 常用于回归任务的评价指标，但是， R^2 并不是严格的显著性检验。

复相关系数

- 称

$$R = \sqrt{R^2}$$

为 y 关于 $x_1, x_2, \cdots, x_p$ 的样本复相关系数。

- **样本复相关系数与样本相关系数的比较**

	度量对象	取值范围	用途
样本复相关系数 R	y 和 $x_1, x_2, \cdots, x_p$	$[0,1]$	衡量 $x_1, x_2, \cdots, x_p$ 与 y 的线性关系
样本相关系数 r	y 和 x_j	$[-1,1]$	衡量 x_j 与 y 的线性关系

总结

- **回归方程的检验**

$$F_0 = \frac{SS_R/p}{SS_E/(n-p-1)} \sim F(p, n-p-1)$$

- **回归系数的检验**

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{c_{jj}}\hat{\sigma}} \sim t(n-p-1)$$

- **回归的拟合优度**

$$R^2 = \frac{SS_R}{SS_T}$$

目录

问题定义

- 提问：在确定参数估计 $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_p)$ 之后，我们可以做些什么？
- 对于一个新数据

$$\boldsymbol{x}_0 = \left(1, x_{01}, x_{02}, \cdots, x_{0p}\right)^{\top}$$

其所对应的响应变量 y_0 是未知的。

- 自然可以假设 y_0 是一个随机变量。
- 由此衍生出两类问题：
  - 对 y_0 的猜测，统称为**预测**问题；
  - 对 $\mu_0 = E(y_0)$ 的猜测，统称为**估计**问题。

问题定义

- **前提**：参数估计为 $\widehat{\boldsymbol{\beta}}=(\hat{\beta}_0,\hat{\beta}_1,\cdots,\hat{\beta}_p)$，新数据为 $\boldsymbol{x}_0=(1,x_{01},x_{02},\cdots,x_{0p})^\top$。
- 如果要用一个“点”来猜测 y_0 或 $E(y_0)$，一个自然的想法是

$$\hat{y}_0=\hat{\beta}_0+\hat{\beta}_1x_{01}+\cdots+\hat{\beta}_px_{0p}=\boldsymbol{x}_0^\top\widehat{\boldsymbol{\beta}}$$

- 提问：$\hat{y}_0$ 作为一个猜测，它的效果如何？
- 显然，$\hat{y}_0$ 是 $\widehat{\boldsymbol{\beta}}$ 的线性函数。

点预测

定理2.8 在多元线性回归模型中，$\hat{y}_0$ 是 y_0 的无偏预测，即 $E(\hat{y}_0) = E(y_0)$。

- **定理2.8的证明**
- 根据定理2.5可知，

$$E(\hat{y}_0) = E\left(\boldsymbol{x}_0^\top \hat{\boldsymbol{\beta}}\right) = \boldsymbol{x}_0^\top E\left(\hat{\boldsymbol{\beta}}\right) = \boldsymbol{x}_0^\top \boldsymbol{\beta}$$

而

$$y_0 = \boldsymbol{x}_0^\top \boldsymbol{\beta} + \varepsilon_0$$

其中 $E(\varepsilon_0) = 0$ 和 $\mathrm{Var}(\varepsilon_0) = \sigma^2$

所以，

$$E(y_0) = \boldsymbol{x}_0^\top \boldsymbol{\beta}$$

知识回顾

$\boldsymbol{\beta}$ 的估计为

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

其期望为

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

点预测

- 点预测为

$$\hat{y}_0 = \boldsymbol{x}_0^\top \widehat{\boldsymbol{\beta}} = \boldsymbol{x}_0^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

是 $\boldsymbol{y} = (y_1, y_2, \cdots, y_n)^\top$ 的线性函数。

- $\hat{y}_0$ 是线性预测。

定理2.9 设 $\widehat{\boldsymbol{\beta}}$ 是 $\boldsymbol{\beta}$ 的最小二乘估计。对于任一 $(p+1)$ 维常数向量 $\boldsymbol{c}$，$\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}}$ 是 $\boldsymbol{c}^\top \boldsymbol{\beta}$ 的最小方差无偏估计。

推论 在 y_0 的一切线性无偏预测中，$\hat{y}_0$ 的方差最小。

点预测

- **定理2.9的证明**
- 设 $\boldsymbol{d}^\top\boldsymbol{y}$ 是 $\boldsymbol{c}^\top\boldsymbol{\beta}$ 的一个线性无偏估计，即

$$\boldsymbol{c}^\top\boldsymbol{\beta}=E(\boldsymbol{d}^\top\boldsymbol{y})=E(\boldsymbol{d}^\top(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}))=\boldsymbol{d}^\top\boldsymbol{X}\boldsymbol{\beta}$$

于是，

$$\boldsymbol{c}^\top=\boldsymbol{d}^\top\boldsymbol{X}$$

又因为

$$\mathrm{Var}(\boldsymbol{d}^\top\boldsymbol{y})=\sigma^2\boldsymbol{d}^\top\boldsymbol{d}$$

和

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

点预测

- **定理2.9的证明**

于是

$$\begin{aligned}\mathrm{Var}(\boldsymbol{d}^\top \boldsymbol{y}) - \mathrm{Var}\left(\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}}\right) &= \sigma^2 \boldsymbol{d}^\top \boldsymbol{d} - \sigma^2 \boldsymbol{d}^\top \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{d} \\ &= \sigma^2 \boldsymbol{d}^\top \boldsymbol{d} - \sigma^2 \boldsymbol{d}^\top \boldsymbol{H} \boldsymbol{d} \\ &= \sigma^2 \boldsymbol{d}^\top (\boldsymbol{I} - \boldsymbol{H}) \boldsymbol{d} \geq 0\end{aligned}$$

所以，$\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}}$ 是 $\boldsymbol{c}^\top \boldsymbol{\beta}$ 的最小方差无偏估计。

点预测

- 从理论分析中，不难看出 $\hat{y}_0$ 是 y_0 一个不错的预测值。
- 在实际中， $\hat{y}_0$ 仍与 y_0 存在偏差。分解一下偏差的构成方式。

$$
\begin{aligned}
&(\hat{y}_0 - y_0)^2 \\
&= (\hat{y}_0 - E(\hat{y}_0) + E(\hat{y}_0) - E(y_0) + E(y_0) - y_0)^2 \\
&= (\hat{y}_0 - E(\hat{y}_0))^2 + 2(\hat{y}_0 - E(\hat{y}_0))(E(\hat{y}_0) - E(y_0)) \\
&\quad + (E(\hat{y}_0) - E(y_0))^2 + 2(E(\hat{y}_0) - E(y_0))(E(y_0) - y_0) \\
&\quad + (E(y_0) - y_0)^2 + 2(\hat{y}_0 - E(\hat{y}_0))(E(y_0) - y_0) \\
&= (\hat{y}_0 - E(\hat{y}_0))^2 + 2(\hat{y}_0 - E(\hat{y}_0))(E(\hat{y}_0) - E(y_0)) \\
&\quad + (E(\hat{y}_0) - E(y_0))^2 + 2(E(\hat{y}_0) - E(y_0))(-\varepsilon_0) \\
&\quad + \varepsilon_0^2 + 2(\hat{y}_0 - E(\hat{y}_0))(-\varepsilon_0)
\end{aligned}
$$

点预测

- 从理论分析中，不难看出 $\hat{y}_0$ 是 y_0 一个不错的预测值。
- 在实际中， $\hat{y}_0$ 仍与 y_0 存在偏差。分解一下偏差的构成方式。
- 于是有，

$$E(\hat{y}_0 - y_0)^2 = \mathrm{Var}(\hat{y}_0) + \mathrm{Bias}^2(\hat{y}_0) + \mathrm{Var}(\varepsilon_0)$$

区间估计

- 点估计为

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \cdots + \hat{\beta}_p x_{0p} = \boldsymbol{x}_0^\top \widehat{\boldsymbol{\beta}}$$

- 其分布为

$$\hat{y}_0 \sim N(\boldsymbol{x}_0^\top \boldsymbol{\beta}, \sigma^2 \boldsymbol{x}_0^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_0)$$

- 因为 σ^2 是冗余参数，其常见的估计为 $\hat{\sigma}^2 = SS_E/(n-p-1)$。
- 又因为

$$\frac{SS_E}{\sigma^2} \sim \chi^2(n-p-1)$$

区间估计

- 枢轴量为

$$\frac{\hat{y}_0 - \boldsymbol{x}_0^\top \boldsymbol{\beta}}{\hat{\sigma}\sqrt{\boldsymbol{x}_0^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_0}} \sim t(n-p-1)$$

- 取置信水平为 $1-\alpha$。
- 于是，$\mu_0 = E(y_0)$ 的 $1-\alpha$ 置信区间为

$$\hat{y}_0 \pm t_{1-\frac{\alpha}{2}}(n-p-1)\hat{\sigma}\sqrt{\boldsymbol{x}_0^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_0}$$

区间预测

- 点预测为

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \cdots + \hat{\beta}_p x_{0p} = \boldsymbol{x}_0^\top \widehat{\boldsymbol{\beta}}$$

- 其分布为

$$\hat{y}_0 \sim N(\boldsymbol{x}_0^\top \boldsymbol{\beta}, \sigma^2 \boldsymbol{x}_0^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_0)$$

- 而真实值 y_0 的分布为

$$y_0 \sim N(\boldsymbol{x}_0^\top \boldsymbol{\beta}, \sigma^2)$$

$\hat{y}_0$ 和 y_0 相互独立

- $\hat{y}_0 - y_0$ 的分布为

$$\hat{y}_0 - y_0 \sim N(0, \sigma^2(1 + \boldsymbol{x}_0^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_0))$$

区间预测

- $\hat{y}_0 - y_0$ 的分布为

$$\hat{y}_0 - y_0 \sim N(0, \sigma^2(1 + \boldsymbol{x}_0^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_0))$$

- 因为 σ^2 是冗余参数，其常见的估计为 $\hat{\sigma}^2 = SS_E/(n-p-1)$。
- 又因为

$$\frac{SS_E}{\sigma^2} \sim \chi^2(n-p-1)$$

- 于是有

$$\frac{\hat{y}_0 - y_0}{\hat{\sigma}\sqrt{1 + \boldsymbol{x}_0^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_0}} \sim t(n-p-1)$$

区间预测

- 令

$$P\left(c_1 < \frac{\hat{y}_0 - y_0}{\hat{\sigma}\sqrt{1 + \boldsymbol{x}_0^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_0}} < c_2\right) \geq 1 - \alpha$$

- y_0 的$1 - \alpha$ 预测区间为

$$\hat{y}_0 \pm t_{1-\frac{\alpha}{2}}(n - p - 1)\hat{\sigma}\sqrt{1 + \boldsymbol{x}_0^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{x}_0}$$

总结与思考

- y_0 的点预测

$$\hat{y}_0 = \boldsymbol{x}_0^\top \widehat{\boldsymbol{\beta}}$$

- $\mu_0 = E(y_0)$ 的 $1-\alpha$ 置信区间为

$$\hat{y}_0 \pm t_{1-\frac{\alpha}{2}}(n-p-1)\hat{\sigma}\sqrt{\boldsymbol{x}_0^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{x}_0}$$

- y_0 的$1-\alpha$ 预测区间为

$$\hat{y}_0 \pm t_{1-\frac{\alpha}{2}}(n-p-1)\hat{\sigma}\sqrt{1+\boldsymbol{x}_0^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{x}_0}$$

谢谢

EAST CHINA NORMAL UNIVERSITY
華東師範大學

DaSE
Data Science
& Engineering
SCHOOL OF DATA
SCIENCE & ENGINEERING
数据科学与工程学院