引言

马学俊 (主讲) 晁越 (助教)

苏州大学

数学科学学院

https://xuejunma.github.io/

1. 多元线性回归模型
2. 回归参数的估计
3. 参数估计量的性质
4. 约束的最小二乘
5. 回归方程的显著性检验
6. 中心化和标准化
7. 相关阵与偏相关系数
8. 本章小结与评注

多元线性回归模型的一般形式

设随机变量 y 与一般变量 $x_1, x_2, \ldots, x_p$ 的线性回归模型为：

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

对于随机误差项假定：

$$\begin{cases} E(\epsilon) = 0 \\ var(\epsilon) = \sigma^2 \end{cases}$$

对于 n 组观测数据 $(x_{i1}, x_{i2}, \cdots, x_{ip}; y_i), i = 1, 2, \ldots, n$，线性回归模型表示为：

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_p x_{1p} + \epsilon_1 \\ y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_p x_{2p} + \epsilon_2 \\ \quad \vdots \\ y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_p x_{np} + \epsilon_n \end{cases}$$

多元线性回归模型的一般形式

写成矩阵的形式为

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

其中

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \qquad \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \text{(设计矩阵)}$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \qquad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

多元线性回归模型的基本假定

1. 解释变量 $x_1, x_2, \ldots, x_p$ 是确定性变量，不是随机变量，且要求 $rank(\boldsymbol{X}) = p + 1 < n$。表明设计矩阵 $\boldsymbol{X}$ 中的自变量列之间不相关，$\boldsymbol{X}$ 是一满秩矩阵。
2. 随机误差项具有 0 均值和等方差，即

$$\begin{cases} E(\epsilon_i) = 0, \quad i = 1, 2, \ldots, n \\ cov(\epsilon_i, \epsilon_j) = \begin{cases} \sigma^2, i = j \\ 0, i \neq j \end{cases} \quad i, j = 1, 2, \ldots, n \end{cases}$$

这个假定称为 Gauss-Markov 条件

3. 正态分布的假定条件为:

$$\begin{cases} \epsilon_i \sim N(0, \sigma^2) \\ \epsilon_1, \epsilon_2, \ldots, \epsilon_n \quad \text{相互独立} \end{cases}$$

矩阵形式表示为

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

多元线性回归模型的基本假定

在正态假定下：

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

$$E(\boldsymbol{y}) = \boldsymbol{X}\beta$$

$$var(\boldsymbol{y}) = \sigma^2 \boldsymbol{I}_n$$

多元线性回归方程的解释

- y 表示空调机的销售量，
- x_1 表示空调机的价格，
- x_2 表示消费者可用于支配的收入。

建立二元线性回归模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

在 x_2 保持不变时，$\frac{\partial E(y)}{\partial x_1} = \beta_1$

- β_1 可解释为在消费者收入 x_2 保持不变时，空调机价格 x_1 每增加一个单位，空调机销售量 y 的平均增加幅度。

在 x_1 保持不变时，$\frac{\partial E(y)}{\partial x_2} = \beta_2$

回归参数的普通最小二乘估计

最小二乘估计要寻找 $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_p$，使得

$$Q(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_p) = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_p x_{ip})^2$$

$$= \min_{\beta_0, \beta_1, \beta_2, \ldots, \beta_p} \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_p x_{ip})^2$$

回归参数的普通最小二乘估计

$$
\begin{cases}
\dfrac{\partial Q}{\partial \beta_0}\Big|_{\beta_0=\hat{\beta}_0} = -2\sum\limits_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_p x_{ip}) \\
\dfrac{\partial Q}{\partial \beta_1}\Big|_{\beta_1=\hat{\beta}_1} = -2\sum\limits_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_p x_{ip})x_{i1} \\
\dfrac{\partial Q}{\partial \beta_2}\Big|_{\beta_2=\hat{\beta}_2} = -2\sum\limits_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_p x_{ip})x_{i2} \\
\qquad\vdots \\
\dfrac{\partial Q}{\partial \beta_p}\Big|_{\beta_p=\hat{\beta}_p} = -2\sum\limits_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_p x_{ip})x_{ip}
\end{cases}
$$

回归参数的普通最小二乘估计

整理后得到矩阵形式表示的正规方程组

$$\boldsymbol{X}'(\boldsymbol{y}-\boldsymbol{X}\hat{\beta})=\boldsymbol{0}$$

移项得

$$\boldsymbol{X}'\boldsymbol{X}\hat{\beta}=\boldsymbol{X}'\boldsymbol{y}$$

当 $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ 存在时，即得到回归参数的最小二乘估计为

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

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} \quad\Rightarrow\quad |\boldsymbol{X}'\boldsymbol{X}|\neq 0 \quad\Rightarrow\quad rank(\boldsymbol{X}'\boldsymbol{X})=p+1$$
$$\Rightarrow\quad rank(\boldsymbol{X})\geq p+1 \quad\Rightarrow\quad \boldsymbol{X}_{n\times(p+1)} \quad\Rightarrow\quad n\geq p+1$$

回归值与残差

称

$$\hat{y}_i = y_i - \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_p x_{ip}$$

为观测值 y_i 得回归拟合值。

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

称

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

为帽子矩阵。其主对角线元素记为 h_{ii}，帽子矩阵得迹为 $tr(\boldsymbol{H}) = \sum_{i=1}^{n} h_{ii} = p+1$。
证明需要用到迹得性质 $tr(\boldsymbol{AB}) = tr(\boldsymbol{BA})$:

$$tr(\boldsymbol{H}) = tr(\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}') = tr((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}) = tr(\boldsymbol{I}_{p+1}) = p+1$$

回归值与残差

回归残差向量

$$\boldsymbol{e} = (e_1, e_2, \cdots, r_n)' = \boldsymbol{y} - \hat{\boldsymbol{y}} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y}$$

$$\begin{aligned} D(\boldsymbol{e}) &= cov(\boldsymbol{e}, \boldsymbol{e}) \\ &= cov((\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y}, (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y}) \\ &= (\boldsymbol{I} - \boldsymbol{H})cov(\boldsymbol{y}, \boldsymbol{y})(\boldsymbol{I} - \boldsymbol{H})' \\ &= \sigma^2(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{I}_n(\boldsymbol{I} - \boldsymbol{H})' \\ &= \sigma^2(\boldsymbol{I} - \boldsymbol{H}) \end{aligned}$$

得到

$$D(e_i) = (1 - h_{ii})\sigma^2$$

回归值与残差

又因为

$$E(\sum_{i=1}^{n} e_i^2) = \sum_{i=1}^{n} D(e_i) = (n-p-1)\sigma^2$$

可得

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n-p-1} SSE = \frac{1}{n-p-1}(\boldsymbol{e}'\boldsymbol{e}) \\ &= \frac{1}{n-p-1}\sum_{i=1}^{n} e_i^2\end{aligned}$$

回归参数得最大似然估计

$\boldsymbol{y}$ 得概率分布为

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

可以得到似然函数为

$$L = (2\pi)^{-n/2}(\sigma^2)^{-n/2} \exp(-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\beta)'(\boldsymbol{y} - \boldsymbol{X}\beta))$$

两边同时取对数似然

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

等价于使 $(\boldsymbol{y} - \boldsymbol{X}\beta)'(\boldsymbol{y} - \boldsymbol{X}\beta)$ 达到最小，这又与普通最小二乘相同。

参数估计量的性质

- 性质 1：$\hat{\boldsymbol{\beta}}$ 是随机向量 $\boldsymbol{y}$ 的一个线性变换。

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

- 性质 2：$\hat{\boldsymbol{\beta}}$ 是 $\boldsymbol{\beta}$ 的无偏估计。

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

参数估计量的性质

- 性质 3： $D(\hat{\beta}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$

$$
\begin{aligned}
D(\hat{\beta}) &= cov(\hat{\beta}, \hat{\beta}) \\
&= cov((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}, (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}) \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' cov(\boldsymbol{y}, \boldsymbol{y})((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')' \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\sigma^2\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}
\end{aligned}
$$

参数估计量的性质

当 $p = 1$ 时即一元线性回归的情况，此时

$$\boldsymbol{X}'\boldsymbol{X} = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}$$

$$\begin{aligned}
(\boldsymbol{X}'\boldsymbol{X})^{-1} &= \frac{1}{|\boldsymbol{X}'\boldsymbol{X}|} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix} \\
&= \frac{1}{nL_{xx}} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{nL_{xx}} \sum_{i=1}^n x_i^2 & -\frac{\bar{x}}{L_{xx}} \\ -\frac{\bar{x}}{L_{xx}} & \frac{1}{L_{xx}} \end{pmatrix}
\end{aligned}$$

$$D(\hat{\boldsymbol{\beta}}) = \begin{pmatrix} var(\hat{\beta}_0) & cov(\hat{\beta}_0, \hat{\beta}_1) \\ cov(\hat{\beta}_0, \hat{\beta}_1) & var(\hat{\beta}_1) \end{pmatrix}$$

参数估计量的性质

- 性质 4：Gauss-Markov 定理
  预测函数 $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_{10} + \hat{\beta}_2 x_{20} + \cdots + \hat{\beta}_p x_{p0}$ 是 $\hat{\boldsymbol{\beta}}$ 的线性函数。
  Gauss-Markov 定理：
  在假定 $E(\boldsymbol{y}) = \boldsymbol{X\beta}, D(\boldsymbol{y}) = \sigma^2 \boldsymbol{I}_n$ 时，$\boldsymbol{\beta}$ 的任一线性函数 $\boldsymbol{c}'\boldsymbol{\beta}$ 的最小方差线性无偏估计（BLUE）为 $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$，其中，$\boldsymbol{c}$ 是任一维常数向量，$\hat{\boldsymbol{\beta}}$ 是 $\boldsymbol{\beta}$ 的最小二乘估计。

1. 取常数向量 $\boldsymbol{c}$ 的第 $j(j = 0, 1, \ldots, p)$ 个分量为 1，其余分量为 0，这时 G-M 定理表明最小二乘估计 $\hat{\beta}_j$ 是 β_j 的最小方差线性无偏估计。
2. 可能存在 $y_1, y_2, \ldots, y_n$ 的非线性函数，作为 $\boldsymbol{c}'\boldsymbol{\beta}$ 的无偏估计，比最小二乘估计 $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ 的方差更小。
3. 可能存在 $\boldsymbol{c}'\boldsymbol{\beta}$ 的有偏估计量，在某种意义（例如均方误差最小）下比最小二乘估计 $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ 更好。
4. 在正态假定下，$\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ 是 $\boldsymbol{c}'\boldsymbol{\beta}$ 的最小方差无偏估计。也就是说，既不可能存在 $y_1, y_2, \ldots, y_n$ 的非线性函数，也不可能存在 $y_1, y_2, \ldots, y_n$ 的其它线性函数，作为 $\boldsymbol{c}'\boldsymbol{\beta}$ 的无偏估计，比最小二乘估计 $\boldsymbol{c}'\hat{\boldsymbol{\beta}}$ 方差更小。

- 在线性回归中，在给定 $p+1$ 维的常向量 $\boldsymbol{c}$，则在 $\boldsymbol{c}^\top\boldsymbol{\beta}$ 的所以线性无偏估计中，最小二乘 $\boldsymbol{c}^\top\hat{\boldsymbol{\beta}}$ 唯一具有最小方差估计。[1]

证明：设 $\boldsymbol{\alpha}^\top\boldsymbol{Y}$ 为 $\boldsymbol{c}^\top\boldsymbol{\beta}$ 的任意一个线性无偏估计，于是对于一切的 $\boldsymbol{\beta}$，有

$$\boldsymbol{c}^\top\boldsymbol{\beta} = \mathbf{E}(\boldsymbol{\alpha}^\top\boldsymbol{Y}) = \boldsymbol{\alpha}^\top\boldsymbol{\alpha}^\top\boldsymbol{X}\boldsymbol{\beta}$$

由于 $\boldsymbol{\beta}$ 的任意性可得 $\boldsymbol{c}^\top = \boldsymbol{\alpha}^\top\boldsymbol{Y}$。任意一个向量 $||\boldsymbol{Z}||^2 = \boldsymbol{Z}^\top\boldsymbol{Z}$，则 $||\boldsymbol{Z}||^2 = 0$ 当且仅当 $|\boldsymbol{Z} = 0$。

$$\begin{aligned}
Var(\boldsymbol{\alpha}^\top\boldsymbol{Y}) &= \boldsymbol{\alpha}^\top Var(\boldsymbol{Y})\boldsymbol{\alpha} = \sigma^2||\boldsymbol{\alpha}||^2 \\
&= \sigma^2||\boldsymbol{\alpha} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{c} + \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{c}||^2 \\
&= \sigma^2||\boldsymbol{\alpha} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{c}||^2 + \sigma^2\boldsymbol{c}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{c} \\
&+ 2\sigma\boldsymbol{c}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{\alpha} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{c})
\end{aligned}$$

将 $\boldsymbol{c}^\top = \boldsymbol{\alpha}^\top\boldsymbol{Y}$ 代入第三项得此项为零。$Var(\boldsymbol{c}^\top\hat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{c}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{c}$，故 $Var(\boldsymbol{\alpha}^\top\boldsymbol{Y}) \geq Var(\boldsymbol{c}^\top\hat{\boldsymbol{\beta}})$ 等号当且仅当 $\boldsymbol{\alpha} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{c} = 0$ 成立，即 $\boldsymbol{\alpha}^\top\boldsymbol{Y} = \boldsymbol{\alpha}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y}$

[1] 梅长林，王宁。近代回归分析方法，2012，中国科学出版社.p.7.

- **性质 5** $cov(\hat{\boldsymbol{\beta}}, \boldsymbol{e}) = 0$
  此性质说明 $\hat{\boldsymbol{\beta}}$ 与 $\boldsymbol{e}$ 不相关，在正态假定下 $\hat{\boldsymbol{\beta}}$ 与 $\boldsymbol{e}$ 等价于 $\hat{\boldsymbol{\beta}}$ 与 $\boldsymbol{e}$ 独立，从而 $\hat{\boldsymbol{\beta}}$ 与 $SSE = \boldsymbol{e}'\boldsymbol{e}$ 独立。

- **性质 6**：当 $\boldsymbol{y} \sim N(\boldsymbol{X}\beta, \sigma^2\boldsymbol{I}_n)$ （等价$\boldsymbol{\varepsilon} \sim N(0, \sigma^2\boldsymbol{I})$）时，则
  1. $\hat{\beta} \sim N(\beta, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})$
  2. $SSE/\sigma^2 \sim \chi^2(n-p-1) \Rightarrow E(SSE) = \sigma^2(n-p-1)$
  3. $\hat{\beta}$ 与 SSE/σ^2 相互独立。

证明 (2)

由于 $e = (\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Y} = (\boldsymbol{I}-\boldsymbol{H})(\boldsymbol{X}\beta+\varepsilon) = (\boldsymbol{I}-\boldsymbol{H})\varepsilon$ ，从而 $SSE = \varepsilon^\top(\boldsymbol{I}-\boldsymbol{H})\varepsilon$ 再由于 $\boldsymbol{I}-\boldsymbol{H}$ 是对称幂等矩阵，故 $rank(\boldsymbol{I}-\boldsymbol{H}) = n-p-1$，从而存在 n 阶正交矩阵 $\boldsymbol{P}$，使得

$$\boldsymbol{P}^\top(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{P} = \begin{pmatrix} \boldsymbol{I}_{n-p-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}$$

令

$$\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^\top = \boldsymbol{P}^\top\varepsilon \Rightarrow \boldsymbol{\varepsilon} = \boldsymbol{P}\boldsymbol{\eta}$$

则 $E(\boldsymbol{\eta}) = 0, Var(\boldsymbol{\eta}) = \sigma^2\boldsymbol{P}^\top\boldsymbol{P} = \sigma^2\boldsymbol{I}$，从而 $\boldsymbol{\eta} \sim N(0, \sigma^2\boldsymbol{I})$，进而

$$\frac{1}{\sigma^2}SSE = \frac{1}{\sigma^2}\boldsymbol{\eta}^\top\boldsymbol{P}^\top(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{P}\boldsymbol{\eta} = \frac{1}{\sigma^2}(\eta_1^2 + \ldots + \eta_{n-p-1}^2)$$

由于 $\eta_i \sim N(0, \sigma^2)$ 相互独立，并且 $\eta_i/\sigma \sim N(0,1)$，从而 $\eta_i^2/\sigma^2 \sim \chi^2(1)$，再由卡方分布得可加性得 $SSE/\sigma^2 \sim \chi^2(n-p-1)$

证明 (3)

将 n 阶正交矩阵 $\boldsymbol{P}$ 分成两块 $\boldsymbol{P}=(\boldsymbol{P}_1,\boldsymbol{P}_2)$，其中 $\boldsymbol{P}_1$ 由 $\boldsymbol{P}$ 的前 $n-p-1$ 列组成，从而

$$\boldsymbol{\eta}=\begin{pmatrix}\boldsymbol{\eta}_1\\ \boldsymbol{\eta}_2\end{pmatrix}=\begin{pmatrix}\boldsymbol{P}_1^\top\\ \boldsymbol{P}_2^\top\end{pmatrix}\varepsilon=\begin{pmatrix}\boldsymbol{P}_1^\top\varepsilon\\ \boldsymbol{P}_2^\top\varepsilon\end{pmatrix}$$

由于 $\boldsymbol{\eta}\sim N(0,\sigma^2\boldsymbol{I})$，所以 $\boldsymbol{\eta}_1\sim N(0,\sigma^2\boldsymbol{I}_{n-p-1})$，$\boldsymbol{\eta}_2\sim N(0,\sigma^2\boldsymbol{I}_{p+1})$，且 $\boldsymbol{\eta}_1$ 和 $\boldsymbol{\eta}_2$ 相互独立

$$\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}=\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\varepsilon=\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{P}\eta\overset{\triangle}{=\!=}\boldsymbol{D}\eta$$

其中 $\boldsymbol{D}=\boldsymbol{X}(\boldsymbol{X}^\top)^{-1}\boldsymbol{P}=(\boldsymbol{D}_1,\boldsymbol{D}_2)$, $\boldsymbol{D}_1$ 由 $\boldsymbol{D}$ 的前 $n-p-1$ 列组成。

$$(\boldsymbol{D}_1,\boldsymbol{D}_2)\begin{pmatrix}\boldsymbol{I}_{n-p-1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}\boldsymbol{D}\boldsymbol{P}^\top(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{P}=\boldsymbol{X}(\boldsymbol{X}^\top)^{-1}(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{P}=0$$

从而 $\boldsymbol{D}_1=0$，所以

$$\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}=(\boldsymbol{D}_1,\boldsymbol{D}_2)\begin{pmatrix}\boldsymbol{\eta}_1\\ \boldsymbol{\eta}_2\end{pmatrix}=\boldsymbol{D}_2\boldsymbol{\eta}_2$$

- **性质 7**：当 $\boldsymbol{y} \sim N(\boldsymbol{X}\beta, \sigma^2 \boldsymbol{I}_n)$（等价$\varepsilon \sim N(0, \sigma^2 \boldsymbol{I})$）时，则
  1. $E(\boldsymbol{e}) = 0$
  2. $Var(\boldsymbol{e}) = \sigma^2(\boldsymbol{I} - \boldsymbol{H})$
  3. $\boldsymbol{e} \sim N(\mathbf{0}, \sigma^2(\boldsymbol{I} - \boldsymbol{H}))$

约束的最小二乘

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon, \varepsilon \sim N(0, \sigma^2 \boldsymbol{I})$$

约束条件

$$\boldsymbol{A}\beta = \boldsymbol{b}$$

其中 $\boldsymbol{A}$ 是 $m \times (p+1)$ 阶行满秩常值矩阵。

- 如何估计？

利用 Lagrange 方法求解：设 $\boldsymbol{\lambda}$ 为 Lagrange 乘子，构造辅助函数

$$
\begin{aligned}
F(\boldsymbol{\beta},\boldsymbol{\lambda}) &= (\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})+2\boldsymbol{\lambda}^\top(\boldsymbol{A}\boldsymbol{\beta}-\boldsymbol{b}) \\
&= \boldsymbol{Y}^\top\boldsymbol{Y}-2\boldsymbol{\beta}^\top\boldsymbol{X}\boldsymbol{Y}+\boldsymbol{\beta}^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta}+2\boldsymbol{\beta}^\top\boldsymbol{A}^\top\boldsymbol{\lambda}-2\boldsymbol{b}^\top\boldsymbol{\lambda}
\end{aligned}
$$

求偏导

$$
\frac{\partial F(\boldsymbol{\beta},\boldsymbol{\lambda})}{\partial\boldsymbol{\beta}} = -2\boldsymbol{X}^\top\boldsymbol{Y}+2\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta}+2\boldsymbol{A}^\top\boldsymbol{\lambda}=0
$$

$$
\frac{\partial F(\boldsymbol{\beta},\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}} = 2(\boldsymbol{A}\boldsymbol{\beta}-\boldsymbol{b})=0
$$

进而得到

$$
\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{A}^\top\boldsymbol{\lambda}=\boldsymbol{X}^\top\boldsymbol{Y} \tag{1}
$$

$$
\boldsymbol{A}\boldsymbol{\beta}=\boldsymbol{b} \tag{2}
$$

令 $\hat{\boldsymbol{\beta}}_c$ 和 $\hat{\boldsymbol{\lambda}}_c$ 为上述方程组的解，

$$
\hat{\boldsymbol{\beta}}_c=(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y}-(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{A}^\top\hat{\boldsymbol{\lambda}}_c
$$

$$\hat{\beta}_c = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y} - (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{A}^\top\hat{\lambda}_c$$

将代入1的第二等式 $\boldsymbol{A}\beta = \boldsymbol{b}$，从而

$$\hat{\lambda}_c = [\boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{A}^\top]^{-1}(\boldsymbol{A}\hat{\beta} - \boldsymbol{b})$$

可以得到

$$\hat{\beta}_c = \hat{\beta} - (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{A}^\top[\boldsymbol{A}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{A}^\top]^{-1}(\boldsymbol{A}\hat{\beta} - \boldsymbol{b})$$

为了确保 $\hat{\beta}_c$ 的确为约束条件下的 β 的最小二乘估计，我们还需证明下面两点（**作业**）

- (1)$\boldsymbol{A}\hat{\beta}_c = \boldsymbol{b}$
- (2) 对一切满足条件的 $\boldsymbol{A}\beta - \boldsymbol{b}$ 的 β，都有 $||\boldsymbol{Y} - \boldsymbol{X}\beta||^2 \geq ||\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}_c||^2$

F 检验

原假设

$$H_0: \quad \beta_1 = \beta_2 = \cdots = \beta_p = 0$$

利用总离差平方和的分解式

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

$$SST = SSR + SSE$$

构造 F 检验统计量如下

$$F = \frac{SSR/p}{SSE/(n-p-1)}$$

当原假设成立时，F 服从自由度为 $(p, n-p-1)$ 的 F 分布。

方差来源	自由度	平方和	均方	F 值	p 值
回归	p	SSR	$\frac{SSR}{p}$	$\frac{SSR/p}{SSE/(n-p-1)}$	$P(F > F)$
残差	$n-p-1$	SSE	$\frac{SSE}{n-p-1}$		
总和	$n-1$	SST			

一般约束的检验

$$\boldsymbol{y} = \boldsymbol{X}\beta + \varepsilon, \varepsilon \sim N(0, \sigma \boldsymbol{I})$$

下面我们假设

$$H_0 : \boldsymbol{A}\beta = \boldsymbol{b}$$

其中 $\boldsymbol{A}$ 是 $m \times (p+1)$ 阶行满秩常值矩阵。令

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

$$SSE(H_0) = (\boldsymbol{y} - \boldsymbol{X}\hat{\beta}(H_0))^\top (\boldsymbol{y} - \boldsymbol{X}\hat{\beta}(H_0))$$

小组作业

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

- $SSE/\sigma^2 \sim \chi^2_{n-p-1}$
- 若 H_0 成立，则 $(SSE(H_0) - SSE)/\sigma^2 \sim \chi^2_m$
- SSE 和 $SSE(H_0) - SSE$ 相互独立
- 若 H_0 成立

$$F = \frac{(SSE(H_0) - SSE)/m}{\cdots} \sim F(m, n-p-1)$$

t 检验

原假设

$$H_0: \quad \beta_j = 0, \quad j = 1, 2, \ldots, p$$

已知

$$\hat{\beta} \sim N(\beta, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})$$

记

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = (c_{ij}), \quad i, j = 0, 1, 2, \ldots, p$$

由此可以构造 t 统计量

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

其中

$$\hat{\sigma} = \sqrt{\frac{1}{n-p-1}\sum_{i=1}^{n} e_i^2} = \sqrt{\frac{1}{n-p-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

回归方程的显著性检验: 偏 F 统计量

从另外一个角度考虑自变量 x_j 的显著性。

- y 对自变量 $x_1, x_2, \ldots, x_p$ 线性回归的残差平方和为 SSE，回归平方和为 SSR
- 在剔除掉 x_j 后，用 y 对其余的 $p-1$ 个自变量做回归，记所得的残差平方和为 $SSE_{(j)}$，回归平方和为 $SSR_{(j)}$，则自变量 x_j 对回归的贡献为 $\Delta SSR_{(j)} = SSR - SSR_{(j)}$，称为 x_j 的偏回归平方和。由此构造偏 F 统计量

$$F_j = \frac{\Delta SSR_{(j)}/1}{SSE/(n-p-1)}$$

- 当原假设 $H_{0j}: \beta_j = 0$ 成立时，偏 F 统计量 F_j 服从自由度为 $(1, n-p-1)$ 的 F 分布，此 F 检验 t 检验是一致的，可以证明 $F_j = t_j^2$。

回归系数的置信区间

因为

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

可得 β_j 的置信度为 $1-\alpha$ 的置信区间为

$$(\hat{\beta}_j - t_{\alpha/2}\sqrt{c_{jj}}\hat{\sigma}, \hat{\beta}_j + t_{\alpha/2}\sqrt{c_{jj}}\hat{\sigma})$$

拟合优度

- 定义样本的决定系数

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

- y 关于 $x_1, x_2, \ldots, x_p$ 的样本复相关系数

$$R = \sqrt{R^2} = \sqrt{\frac{SSR}{SST}}$$

- 调整 R^2

$$R^2 = 1 - \frac{n-}{n-p-1}(1 - R^2)$$

通常用来比较自变量个数不同的模型拟合效果，不能理解为变量 Y 的总方差中能由自变量解释的比例。

中心化

多元线性回归模型的经验回归方程为

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_p x_p$$

经过样本中心 $(\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_p; \bar{y})$，将坐标原点移到样本中心，即

$$x'_{ij} = x_{ij} - \bar{x}_j, \quad i = 1, 2, \ldots, n \quad j = 1, 2, \ldots, p$$

$$y'_i = y_i - \bar{y}, \quad i = 1, 2, \ldots, n$$

上述经验方程转换为

$$\hat{y}' = \hat{\beta}_1 x'_1 + \hat{\beta}_2 x'_2 + \cdots + \hat{\beta}_p x'_p$$

回归常数项为

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2 - \cdots - \hat{\beta}_p \bar{x}_p$$

标准化回归系数

样本数据的标准化公式为

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{\sqrt{L_{jj}}}, \quad i = 1, 2, \cdots, n; \quad j = 1, 2, \cdots, p$$

$$y_i^* = \frac{y_i - \bar{y}}{\sqrt{L_{yy}}}, \quad i = 1, 2, \cdots, n$$

其中

$$L_{jj} = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$$

标准化样本的经验回归方程为

$$\hat{y}^* = \hat{\beta}_1^* x_1^* + \hat{\beta}_2^* x_2^* + \cdots + \hat{\beta}_p^* x_p^*$$

$$\hat{\beta}_j^* = \frac{\sqrt{L_{jj}}}{\sqrt{L_{yy}}} \hat{\beta}_j, \quad j = 1, 2, \cdots, p$$

样本相关阵

r_{ij} 为 x_i 与 x_j 之间简单的相关系数，自变量样本相关阵

$$\boldsymbol{r} = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{pmatrix}$$

同时，相关阵可表示为：

$$\boldsymbol{r} = (\boldsymbol{X}^*)'\boldsymbol{X}^*$$

$\boldsymbol{X}^* = (x_{ij}^*)_{n\times p}$ 表示中心标准化的设计阵。
增广的样本相关阵为：

$$\boldsymbol{r} = \begin{pmatrix} 1 & r_{y1} & r_{y2} & \cdots & r_{yp} \\ r_{1y} & 1 & r_{12} & \cdots & r_{1p} \\ r_{2y} & r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{py} & r_{p1} & r_{p2} & \cdots & 1 \end{pmatrix}$$

偏相关系数

当其他变量被固定后，给定的任两个变量之间的相关系数，叫偏相关系数。

偏相关系数可以度量 $p+1$ 个变量 $y, x_1, x_2, \cdots x_p$ 之中任意两个变量的线性相关程度，而这种相关程度是在固定其余 $p-1$ 个变量的影响下的线性相关。

偏判定系数测量在回归方程中已包含若干个自变量时，再引入某一个新的自变量后 y 的剩余变差的相对减少量，它衡量 y 的变差减少的边际贡献。

两个自变量的偏判定系数

二元线性回归模型为：

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i, \quad i = 1, 2, \cdots, n$$

记 $SSE(x_2)$ 是模型中只含有自变量 x_2 时 y 的残差平方和，$SSE(x_1\ x_2)$ 是模型中同时含有自变量 x_1 和 x_2 时 y 的残差平方和。因此模型中已含有 x_2 时再加入 x_1 使 y 的剩余变差的相对减小量为

$$r^2_{y1;2} = \frac{SSE(x_2) - SSE(x_1, x_2)}{SSE(x_2)}$$

此即模型中已含有 x_2 时，y 与 x_1 的偏判定系数。
同样，模型中已含有 x_1 时，y 与 x_2 的偏判定系数为：

$$r^2_{y2;2} = \frac{SSE(x_1) - SSE(x_1, x_2)}{SSE(x_1)}$$

一般情况

$$r^2_{y1;2,\cdots,p} = \frac{SSE(x_2,\cdots,x_p) - SSE(x_1,x_2,\cdots,x_p)}{SSE(x_2,\cdots,x_p)}$$

偏决定系数与回归系数显著性检验的偏 F 值是等价的。

偏相关系数

偏判定系数的平方根称为偏相关系数，其符号与相应的回归系数的符号相同。偏相关系数与回归系数显著性检验的 t 值是等价的。

固定 $x_3, \cdots, x_p$ 保持不变时，x_1 与 x_2 之间的偏相关系数为：

$$r_{12;3,\cdots,p} = \frac{-\Delta_{12}}{\sqrt{\Delta_{11} \cdot \Delta_{22}}}$$

其中符号 Δ_{ij} 表示相关矩阵 $(r_{ij})_{p\times p}$ 第 i 行第 j 列元素的代数余子式。验证以下关系

$$r_{12;3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}}$$

本章小结与评注

- 样本决定系数 R^2 越大，说明回归方程拟合原始数据 y 的观测值的效果越好。
- 但由于 R^2 的大小与样本容量 n 以及自变量个数 p 有关，当 n 与 p 的数目接近时，R^2 容易接近于 1，这说明 R^2 中隐含着一些虚假成分
- 仅由 R^2 的值很大，去推断模型优劣一定要慎重。

对于回归方程的显著性检验，我们用 F 统计量去判断假设 $H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0$ 是否成立。

- 当给定显著性水平 α 时，$F > F_\alpha(p, n-p-1)$，则拒绝假设 H_0，否则不拒绝 H_0。
- 接受假设 H_0 和拒绝假设 H_0 对于回归方程来说意味着什么，这仍需慎重对待。
    - 一般来说，当接受假设 H_0 时，认为在给定的显著性水平 α 之下，自变量 $x_1, x_2, \cdots, x_p$ 对因变量 y 无显著性影响，于是通过 $x_1, x_2, \cdots, x_p$ 去推断 y 也就无多大意义。
    - 在这种情况下，一方面可能这个问题本来应该用非线性模型去描述，而我们误用线性模型描述了，使得自变量对因变量无显著影响
    - 另一方面可能是在考虑自变量时由于我们认识上的局限性把一些影响因变量 y 的自变量漏掉了。这就从两个方面提醒我们去重新考虑建模问题。

本章小结与评注

当我们拒绝了假设 H_0 时，我们也不能过于相信这个检验，认为这个回归模型已经很完美了。

- 当拒绝 H_0 时，我们只能认为这个回归模型在一定程度上说明了自变量 $x_1, x_2, \cdots, x_p$ 与因变量 y 的线性关系。
- 因为这时仍不能排除我们漏掉了一些重要的自变量。
- 研究者在事前根据专业知识及经验，认为已把较重要的自变量选入了，且在一定误差限度内认为模型为线性是合理的。经过样本数据计算后，可以用来验证一下，原先的考虑是否周全。
- 若拒绝 H_0，可认为至少并不与他原来的设想矛盾。如果接受 H_0，可以认为模型是不能反映因变量 y 与自变量 $x_1, x_2, \cdots, x_p$ 的线性关系，这个模型就不能应用于实际预测和分析

。

补充

矩阵 QR 分解

设 $\boldsymbol{A} \in \mathbb{R}^{n\times m}$，则存在 $\boldsymbol{Q} \in \mathbb{R}^{n\times p}$，且 $\boldsymbol{Q}^\top \boldsymbol{Q} = \boldsymbol{I}_p$，以及一个上三角矩阵 $\boldsymbol{R} \in \mathbb{R}^{p\times p}$ 使得

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$$

奇异值分解代码

```
> rm(list=ls())
> A <- matrix(1:6, 2, 3)
> A.qr <- qr(A)
> Q <- qr.Q(A.qr)
> crossprod(Q)
                [,1]         [,2]
[1,] 1.000000e+00 5.551115e-17
[2,] 5.551115e-17 1.000000e+00
> R <- qr.R(A.qr)
> Q %*% R
     [,1] [,2] [,3]
[1,]    1    3    5
[2,]    2    4    6
```

线性回归求解是一个技术问题，参数估计表达式 $\hat{\beta} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X} \boldsymbol{Y}$ 逆的解法比较复杂，假设

$$\boldsymbol{X} = \boldsymbol{Q}\boldsymbol{R}$$

则

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

第三步使用逆的性质 $(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$。这样大大降低了计算成本。上三角矩阵的逆求解比普通矩阵的逆好很多。下面我们模拟数据看看这个计算效果。我们比较了 lm 和 QR 分解，发现结果一致。原因是前者的求解是基于 QR 分解得到的（是否？查看？rm 文档）

QR 代码

```
> rm(list=ls())
> library(MASS)
> ###data begin
> p <- 5
> n <- 100
> mu <- rep(0, p)
> sigm <- diag(p)
> x <- MASS::mvrnorm(n=n, mu=mu, Sigma=sigm)
> beta <- 1 + runif(p, min = 0, max = 1)
> y <- x %*% beta + rnorm(n)
> ###data end
>
> ###X QR begin
> x.qr <- qr(x)
> Q <- qr.Q(x.qr)
> R <- qr.R(x.qr)
> ###X QR end
>
> beta_hat <- solve(R) %*% t(Q) %*% y
> beta_ols <- coef(lm(y~-1))
> cbind(beta_ols, beta_hat)
```

作业

- 作业:
  - 证明:
  - 实例分析: p.84 3.11 备注: 自己编程
- 小组作业: 证明