2014.3.19

第九讲 线性回归模型（2）

参数的含义及最小二乘估计

回顾一般线性模型的两种表示：

模型表示1：

y: 响应变量，$\mathbf{x}$: 自变量(向量)。线性回归模型假设:

(i) 线性回归函数： $\mathrm{E}(y\,|\,\mathbf{x}) = a + \mathrm{b}'\mathbf{x}$,

(ii) 方差常数/齐性：$\mathrm{var}(y\,|\,\mathbf{x}) = \sigma^2$.

其中a, b, σ^2是未知参数。

模型表示2:

$$y = a + b'\mathbf{x} + \varepsilon,$$

其中

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

(2) $\mathrm{var}(\varepsilon) = \sigma^2$ (方差齐性, Homoscedasticity)

(3) ε与$\mathbf{x}$独立 (外生性, Exogeneity)

例2(单因素方差分析模型). n个个体被随机地分配于K组分别接受某种处理(*treatment*),并测量某指标。假设第k组有n_k个个体，指标测量为$y_{k1},..., y_{kn_k}$ iid $\sim \mathrm{N}(\mu_k, \sigma^2), k = 1,....K$

记 $\varepsilon_{kj} = y_{kj} - \mu_k$, $\mathrm{j} = 1,..., \mathrm{n_k}$ iid $\sim \mathrm{N}(0\;, \sigma^2), k = 1,....K$,即

$$y_{kj} = \mu_k + \varepsilon_{kj},$$

模型可表示为：$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, E\boldsymbol{\varepsilon} = 0, \mathrm{var}(\boldsymbol{\varepsilon}) = \sigma^2 I_n$， 其中

$$Y = \begin{pmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ \vdots \\ y_{K1} \\ \vdots \\ y_{Kn_K} \end{pmatrix}, \mathrm{X} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}, \beta = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_K \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n_1} \\ \vdots \\ \varepsilon_{K1} \\ \vdots \\ \varepsilon_{Kn_K} \end{pmatrix}$$

为了检验K个均值相同,通常重新参数化，比如令

$$\beta_1 = \mu_1,\ \beta_2 = \mu_2 - \mu_1,\ ...,\ \beta_K = \mu_K - \mu_1,$$

对于上述新参数$\tilde{\boldsymbol{\beta}} = (\beta_1,..., \beta_K)'$,写出模型

$\mathbf{Y} = X\tilde{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}, E\boldsymbol{\varepsilon} = 0, \mathrm{var}(\boldsymbol{\varepsilon}) = \sigma^2 I_n$

中的设计阵X

参数含义 - 以简单线性回归模型为例

简单线性模型：$y = a + bx + \varepsilon$

(1) $E(\varepsilon) = 0$, (2) $\mathrm{var}(\varepsilon) = \sigma^2$ (3) ε与x不相关/独立

■ 参数的含义

命题：记$\mu_x = E(x), \mu_y = E(y), \sigma_x = \sqrt{\mathrm{var}(x)}, \sigma_y = \sqrt{\mathrm{var}(y)}$，$\rho = \rho_{xy}$，则

(1) $b = \rho \dfrac{\sigma_y}{\sigma_x} \propto \rho$，所以$b = 0 \Leftrightarrow \rho = 0$

(2) $a = \mu_y - b\mu_x$，

(3) $\sigma^2 = (1-\rho^2)\sigma_y^2$

证明：(1) 由$\varepsilon = y - a - bx \perp x \Rightarrow 0 = \mathrm{cov}(\varepsilon, x) = \mathrm{cov}(y - a - bx, x)$
$= \mathrm{cov}(y, x) - b\,\mathrm{var}(x) \Rightarrow \quad b = \dfrac{\mathrm{cov}(y,x)}{\mathrm{var}(x)} = \rho\sqrt{\dfrac{\mathrm{var}(y)}{\mathrm{var}(x)}} = \rho\dfrac{\sigma_y}{\sigma_x}$

(2) 由$\mu_y = \mathrm{E}(y) = E(a + bx + \varepsilon) = a + b\mu_x \Rightarrow a = \mu_y - b\mu_x$

(3) 由$y = a + bx + \varepsilon$，以及ε与x独立，有如下方差分解
$\mathrm{var}(y) = \mathrm{var}(a + bx) + \mathrm{var}(\varepsilon)$，即$\sigma_y^2 = b^2\sigma_x^2 + \sigma^2$

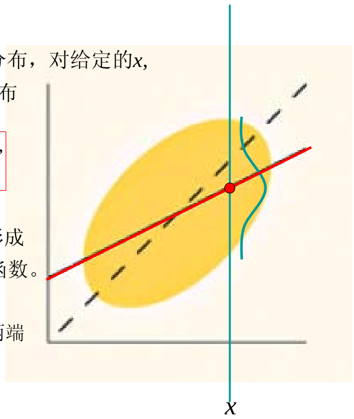
所以$\sigma^2 = \sigma_y^2 - b^2\sigma_x^2 = \sigma_y^2 - \left(\dfrac{\rho^2\sigma_y^2}{\sigma_x^2}\right)\sigma_x^2 = (1-\rho^2)\sigma_y^2$

注：由（3）我们得到 $\rho^2 = \dfrac{\sigma_y^2 - \sigma^2}{\sigma_y^2} = \dfrac{\mathrm{var}(a+bx)}{\mathrm{var}(y)}$

所以x, y相关系数的平方 =回归直线所能解释的 y的方差的百分比

■ 回归效应

假设(x, y)服从二元正态分布，对给定的x，y服从图中所示的正态分布

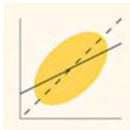


该正态分布的中心，$E(y \mid x)$，在对称轴(虚线)之下

x变化时，$f(x) = E(y \mid x)$形成回归直线(红)，称为回归函数。

相比于虚线，回归直线在两端有向中心回归的趋势。

虚线称为SD线：x偏离其中心1个SD，y也偏离其中心一个SD



假设(x, y)服从如下简单线性模型

$E(y \mid x) = a + bx, \mathrm{var}(y \mid x) = \sigma^2$

如果已知x超过其均值k个单位（$x = \mu_x + k\sigma_x$），

你是否预期y超过其均值k个单位，$E(y \mid x = \mu_x + k\sigma_x) = \mu_y + k\sigma_y$？

当$x = \mu_x + k\sigma_x$时：$E(y \mid x) = \mu_y + bk\sigma_x = \mu_y + k\rho_{xy}\sigma_y < \mu_y + k\sigma_y$

同样当$x = \mu_x - k\sigma_x$，$E(y \mid x) = \mu_y - k\rho_{xy}\sigma_y > \mu_y - k\sigma_y$.

所以回归直线不是虚线(SD线)，而是实线。

例. 课外培训教育是否能提高儿童的IQ？考察入学前IQ值低于平均水平的那些儿童，发现经过一段时间的培训，其IQ值平均提高了5分。以此现象来宣传培训效果是否可信？

实际上，学前比平均分高的儿童在入学后的IQ可能平均降低了5分！此现象可能只是回归效应，并不一定能说明真的有效。

如果入学前后平均IQ值相同，标准差也相同，那么上述现象是正常的回归效应。

又比如减肥药的宣传、高尔顿的父子身高问题等

简单线性回归模型的最小二乘法

1. 最小二乘估计

2. 最小二乘估计的性质

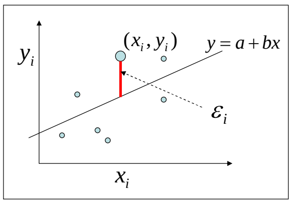
3. 正态模型下的统计推断

1. 最小二乘估计 (Least Squares, LS)

■ 方法

假设$(x_i, y_i), i=1,2,\dots,n.$ iid 满足线性模型:

$y_i = a + bx_i + \varepsilon_i$，$\varepsilon_i$ iid $\sim (0,\sigma^2)$, ε_i与x_i独立.



最小二乘法极小化误差平方和:

$$\min \sum_{i=1}^{n} \varepsilon_i^2 = \min \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

求导得到正则方程:

$$\begin{cases} \sum (y_i - a - bx_i) = 0 \\ \sum x_i (y_i - a - bx_i) = 0 \end{cases}$$

注:正则方程可理解为模型条件的"矩估计方法"实现:

$\sum (y_i - a - bx_i) = \sum \varepsilon_i = 0 \quad \longleftarrow \quad E(\varepsilon) = 0$

$\sum x_i (y_i - a - bx_i) = \sum x_i \varepsilon_i = 0 \quad \longleftarrow \quad$ x与ε独立/不相关

得到LS估计： $\begin{cases}\hat{b}=s_{xy}/s_{xx}\\ \hat{a}=\bar{y}-\hat{b}\bar{x}\end{cases}$

$s_{xx}=\sum(x_i-\bar{x})^2,$
$s_{xy}=\sum(x_i-\bar{x})(y_i-\bar{y})$

改写$\hat{b}$:

$$\hat{b}=\frac{s_{xy}}{s_{xx}}=\frac{s_{xy}}{\sqrt{s_{xx}s_{yy}}}\times\sqrt{\frac{s_{yy}}{s_{xx}}}=r_{xy}\times\frac{s_y}{s_x},$$

$$\hat{a}=\bar{y}-\hat{b}\bar{x}$$

样本标准差

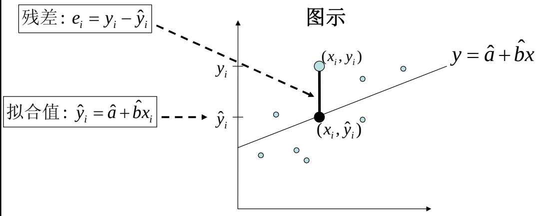
$s_x=\sqrt{\sum(x_i-\bar{x})^2/(n-1)},$
$s_y=\sqrt{\sum(y_i-\bar{y})^2/(n-1)}$

对比参数的表示：

$$(1)\ \ b=\rho\frac{\sigma_y}{\sigma_x};(2)\ \ a=\mu_y-b\mu_x,$$

■ **拟合值与残差**

拟合直线: $y=\hat{a}+\hat{b}x$ 或 $y-\bar{y}=r\frac{s_y}{s_x}(x-\bar{x})$



- 拟合值(fitted value): $\hat{y}_i=\hat{a}+\hat{b}x_i,$
- 残差(residual): $e_i=y_i-\hat{y}_i=y_i-\hat{a}-\hat{b}x_i,$
- 残差平方和RSS (Residual Sum of Squares):

$$RSS=\sum e_i^2=\sum(y_i-\hat{y}_i)^2=\sum(y_i-\hat{a}-\hat{b}x_i)^2$$

即最小的误差平方和

- $\mathrm{RSS}=s_{yy}-s_{xy}^2/s_{xx}=s_{yy}(1-r^2)$

验证： $\mathrm{RSS}=\sum\left(y_i-\hat{a}-\hat{b}x_i\right)^2=\sum\left(y_i-\bar{y}+\hat{b}\bar{x}-\hat{b}x_i\right)^2$
$=s_{yy}-2\hat{b}s_{xy}+\hat{b}^2s_{xx}=s_{yy}-s_{xy}^2/s_{xx}$

性质: $\sum e_i=0,\sum e_ix_i=0,\sum e_i\hat{y}_i=0,$

此即正则方程

$$\begin{cases}0=\sum(y_i-\hat{a}-\hat{b}x_i)=\sum e_i\\ 0=\sum x_i(y_i-\hat{a}-\hat{b}x_i)=\sum x_ie_i\end{cases}$$

■ 误差方差估计

因为$\sigma^2=\mathrm{var}(\varepsilon_i)=\mathrm{E}(\varepsilon_i^2), \varepsilon_i=y_i-a-bx_i$

而$e_i=y_i-\hat{a}-\hat{b}x_i$可看作是$\varepsilon_i$的“估计”（预测）

σ^2的“LS”估计取为：

$$\hat{\sigma}^2=\frac{1}{n-2}RSS=\frac{1}{n-2}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\frac{1}{n-2}\sum_{i=1}^{n}(y_i-\hat{a}-\hat{b}x_i)^2$$

虽然它不是由最小二乘法（极小化误差平方和）直接得到的，但通常也称它为LS估计。

■ 复相关系数平方

注意到$\sum\hat{y}_i=\sum y_i=n\bar{y}$,我们有平方和分解：

$$s_{yy}=\sum(y_i-\bar{y})^2=\sum(y_i-\hat{y}_i+\hat{y}_i-\bar{y})^2=\sum(\hat{y}_i-\bar{y})^2+\sum(y_i-\hat{y}_i)^2$$

即总平方和$s_{yy}=s_{\hat{y}\hat{y}}+RSS$ =回归平方和+残差平方和

定义：对于一般(多重)线性回归,复相关系数平方R^2定义为回归函数(或自变量)所能解释的响应变量方差的百分比：

$$\mathrm{R}^2=\frac{s_{\hat{y}\hat{y}}}{s_{yy}}=\frac{\sum(\hat{y}_i-\bar{y})^2}{\sum(y_i-\bar{y})^2}$$

对于简单回归：$s_{\hat{y}\hat{y}}=\sum(\hat{a}+\hat{b}x_i-(\hat{a}+\hat{b}\bar{x}))^2=\hat{b}^2s_{xx}=s_{xy}^2/s_{xx}$

所以$\mathrm{R}^2=s_{xy}^2/s_{xx}s_{yy}=r^2$

总结如下：

$$y=a+bx+\varepsilon,\quad \varepsilon\sim(0,\sigma^2),\varepsilon\perp x$$

参数	LS估计
(1) $b=\dfrac{\mathrm{cov}(x,y)}{\mathrm{var}(x)}=\rho\dfrac{\sigma_y}{\sigma_x}$	(1) $\hat{b}=\dfrac{s_{xy}}{s_{xx}}=r\dfrac{s_y}{s_x}$
(2) $a=\mu_y-b\mu_x$,	(2) $\hat{a}=\bar{y}-\hat{b}\bar{x}$
(3) $\sigma^2=(1-\rho^2)\sigma_y^2$	(3) $\boxed{\hat{\sigma}^2=(1-r^2)s_y^2\times\dfrac{n-1}{n-2}\approx(1-r^2)s_y^2}$

2. 最小二乘估计的性质

性质1（无偏性）. $E(\hat{b})=b, E(\hat{a})=a, E(\hat{\sigma}^2\,|\,\mathbf{x})=\sigma^2$

证明：$\hat{b}=s_{xy}/s_{xx}$

$$(1)\ E(\hat{b}\,|\,\mathbf{x})=E\left(\frac{\sum(x_i-\bar{x})y_i}{\sum(x_i-\bar{x})^2}\,\middle|\,\mathbf{x}\right)=\frac{\sum(x_i-\bar{x})E(y_i\,|\,x_i)}{\sum(x_i-\bar{x})^2}$$

$$=\frac{\sum(x_i-\bar{x})(a+bx_i+E(\varepsilon_i\,|\,x_i))}{\sum(x_i-\bar{x})^2}=\frac{\sum(x_i-\bar{x})(bx_i)}{\sum(x_i-\bar{x})^2}=b$$

$\Rightarrow E(\hat{b})=b$

由ε与x独立, $E(\varepsilon_i\,|\,x_i)=E(\varepsilon_i)=0$

(2) 另外, $\mathrm{E}(\hat{a}\,|\,\mathbf{x})=E(\bar{y}-\hat{b}\bar{x}\,|\,\mathbf{x})=a+b\bar{x}-E(\hat{b}\,|\,\mathbf{x})\bar{x}=a$

引入向量记号重述上述证明：

令$\mathbf{x}=(x_1,\ldots,x_n)',\mathbf{y}=(y_1,\ldots,y_n)',\mathbf{1}=(1,\ldots,1)'$

则$\hat{b}=s_{xy}/s_{xx}=\dfrac{1}{s_{xx}}(\mathbf{x}-\mathbf{1}\bar{x})'\mathbf{y}$

$$E(\hat{b}\mid\mathbf{x})=\frac{(\mathbf{x}-\mathbf{1}\bar{x})'}{s_{xx}}E(\mathbf{y}\mid\mathbf{x})=\frac{(\mathbf{x}-\mathbf{1}\bar{x})'}{s_{xx}}E(\mathbf{1}a+\mathbf{x}b+\boldsymbol{\varepsilon}\mid\mathbf{x})$$

$$=\frac{(\mathbf{x}-\mathbf{1}\bar{x})'\mathbf{1}a}{s_{xx}}+\frac{(\mathbf{x}-\mathbf{1}\bar{x})'\mathbf{x}b}{s_{xx}}+\frac{(\mathbf{x}-\mathbf{1}\bar{x})'}{s_{xx}}E(\boldsymbol{\varepsilon}\mid\mathbf{x})$$

$$=0+b+\frac{(\mathbf{x}-\mathbf{1}\bar{x})'}{s_{xx}}E(\boldsymbol{\varepsilon})=b$$

注：无偏性需要条件$E(\boldsymbol{\varepsilon})=0$以及$\boldsymbol{\varepsilon}$与$\mathbf{x}$的独立性，但不需要方差齐性。

(3) $\hat{y}_i=\hat{a}+\hat{b}x_i$，向量形式：$\hat{\mathbf{y}}=\mathbf{1}\hat{a}+\mathbf{x}\hat{b}$

而$\hat{b}=s_{xy}/s_{xx}=\dfrac{(\mathbf{x}-\mathbf{1}\bar{x})'\mathbf{y}}{(\mathbf{x}-\mathbf{1}\bar{x})'(\mathbf{x}-\mathbf{1}\bar{x})}$，

将拟合值向量写成y的线性变换形式：

$$\hat{\mathbf{y}}=\mathbf{1}\hat{a}+\mathbf{x}\hat{b}=\mathbf{1}(\bar{y}-\hat{b}\bar{x})+\mathbf{x}\hat{b}=\mathbf{1}\bar{y}+(\mathbf{x}-\bar{x}\mathbf{1})\hat{b}$$

$$=\frac{\mathbf{1}\mathbf{1}'}{n}\mathbf{y}+(\mathbf{x}-\bar{x}\mathbf{1})\frac{(\mathbf{x}-\mathbf{1}\bar{x})'\mathbf{y}}{(\mathbf{x}-\mathbf{1}\bar{x})'(\mathbf{x}-\mathbf{1}\bar{x})}$$

$$=\left(\frac{\mathbf{1}\mathbf{1}'}{n}+\frac{(\mathbf{x}-\bar{x}\mathbf{1})(\mathbf{x}-\mathbf{1}\bar{x})'}{(\mathbf{x}-\mathbf{1}\bar{x})'(\mathbf{x}-\mathbf{1}\bar{x})}\right)\mathbf{y}$$

残差向量 $\mathbf{e}=\mathbf{y}-\hat{\mathbf{y}}=\left(I_n-\dfrac{\mathbf{1}\mathbf{1}'}{n}-\dfrac{(\mathbf{x}-\bar{x}\mathbf{1})(\mathbf{x}-\mathbf{1}\bar{x})'}{(\mathbf{x}-\mathbf{1}\bar{x})'(\mathbf{x}-\mathbf{1}\bar{x})}\right)\mathbf{y}\overset{\text{记为}}{=}C\mathbf{y}$

容易验证C是对称幂等矩阵，所以$(I-C)$也是，而且：

$\mathrm{tr}(C)=n-2$，$C\mathbf{1}=0$，$C\mathbf{x}=0$

则给定$\mathbf{x}$的条件下，

$$\mathrm{E}(\mathrm{RSS}\mid\mathbf{x})=E(\mathbf{e}'\mathbf{e}\mid\mathbf{x})=E(\mathbf{y}'C\mathbf{y}\mid\mathbf{x})$$

$$=(\mathbf{1}a+\mathbf{x}b)'C(\mathbf{1}a+\mathbf{x}b)+tr(C)\sigma^2$$

$$=(n-2)\sigma^2$$

所以$E(\hat{\sigma}^2\mid\mathbf{x})=\mathrm{E}(\mathrm{RSS}/(n-2)\mid\mathbf{x})=\sigma^2$。

注：事实上，$\mathrm{P}=I-C$是$\mathrm{L}(\mathbf{1},\mathbf{x})$对应的投影矩阵。

$\hat{\mathbf{y}}=\mathrm{P}\mathbf{y}$是$\mathbf{y}$在$\mathrm{L}(\mathbf{1},\mathbf{x})$上的投影。