对称回归

回归与逆回归

被预测的是响应。在身高-体重数据中，可以从身高x预测体重y，也可从体重y预测身高x（如果前者称为回归，后者称为逆回归）.
如果x, y地位对等，则以直线描述/逼近数据的问题称为对称回归，尽管对称问题中已经没有回归现象了。

$y \sim x$

回归: 假设y_i是响应（为主）， x_i是自变量。假设模型

$$y_i = a + bx_i + \varepsilon_i,$$

以x_i预测y_i的误差（ y -轴方向）为 $\varepsilon_i = y_i - a - bx_i$

极小化 $\Sigma\varepsilon_i^2 = \Sigma(y_i - a - bx_i)^2 \Rightarrow \hat{b} = s_{xy}/s_{xx}$

$x \sim y$

逆回归: 假设x_i是响应（为主）， y_i是自变量。假设模型

$$x_i = c + dy_i + \delta_i,$$

以y_i预测x_i的误差（ x -轴方向）： $\delta_i = x_i - c - dy_i$

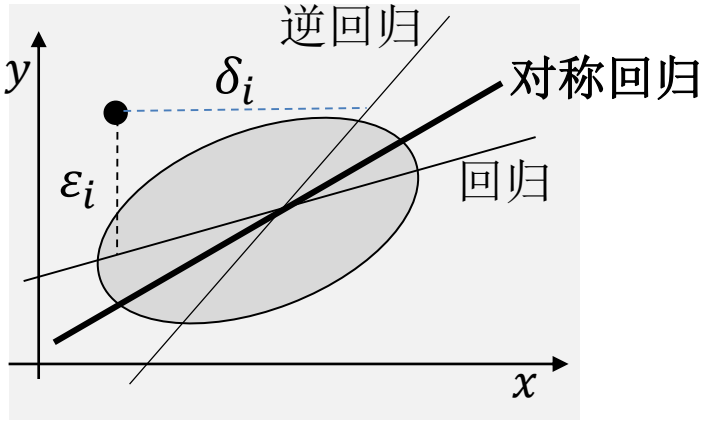
极小化 $\Sigma\delta_i^2 = \Sigma(x_i - c - dy_i)^2 \Rightarrow \hat{d} = s_{xy}/s_{yy}$

$\Rightarrow x$系数$\hat{b}_{\text{inv}} = 1/\hat{d} = s_{yy}/s_{xy}$

由回归和逆回归得到估计$\hat{b}, \hat{b}_{\text{inv}}$满足

$$\hat{b}\hat{b}_{\text{inv}} = s_{yy}/s_{xx} \geq 0, \qquad \hat{b}/\hat{b}_{\text{inv}} = r^2 \leq 1$$

所以$\hat{b}, \hat{b}_{\text{inv}}$符号相同，但$|\hat{b}| \leq |\hat{b}_{\text{inv}}|$，对称回归在回归与逆回归之间。



对称回归得到的斜率估计在回归估计与逆回归估计之间

至少以下两种情形，我们可以考虑对称回归

x, y 对称

有些时候两个变量对称、平等，我们并不希望用一个变量去预测/描述另外一个变量, 而是单纯描述两者之间的关系，可用相关系数进行描述，也可使用**对称回归**。

自变量带误差

通常的回归分析是在给定自变量条件下进行的，可认为自变量是常数。如果需要考虑自变量的随机性，或自变量不可直接测量而只能测量其替代指标，称为自变量误差模型（error-in-variable model), 此时可使用对称回归。

对称回归的估计方法不再用通常的LS方法，一般采用Total least squares，常用的有

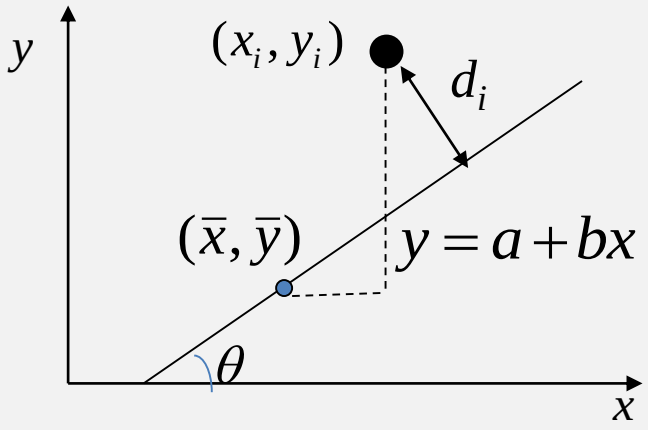
- 主轴回归、
- 约简的主轴回归、
- double回归

主轴回归/ Deming回归

数据$(x_i, y_i), i = 1,..., n$

目标：求对称回归直线$y = a + bx$

主轴回归： $\min \sum d_i^2$，其中d_i为(x_i, y_i)到$y = a + bx$的垂直距离



$b = \tan(\theta)$

$d_i = (y_i - \bar{y})\cos(\theta) - (x_i - \bar{x})\sin(\theta),$

$\sum d_i^2$对θ求导得

$$\frac{2s_{xy}}{s_{xx} - s_{yy}} = \tan(2\theta) = \frac{2b}{1 - b^2}$$

$$\Rightarrow \hat{b}_{ma} = \frac{s_{xx} - s_{yy} + \sqrt{(s_{xx} - s_{yy})^2 + 4s_{xy}^2}}{2s_{xy}}$$

$\hat{\theta} = \arctan(\hat{b}_{ma})$为二元正态分布等概率椭圆

$$\frac{x^2}{s_{xx}} - 2r\frac{xy}{\sqrt{s_{xx}s_{yy}}} + \frac{y^2}{s_{yy}} = c$$

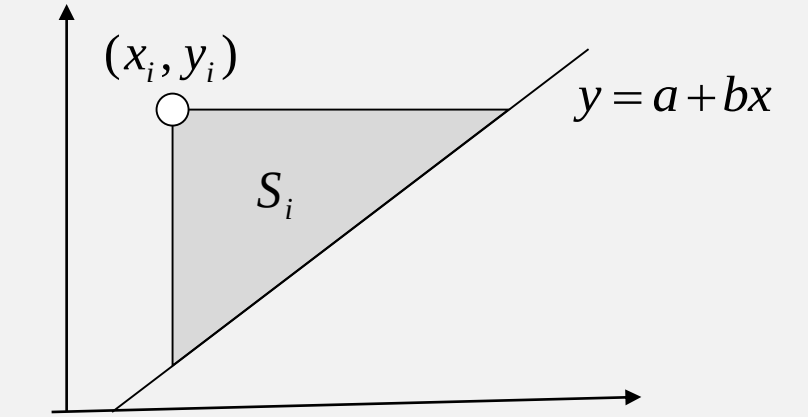
的主轴方向。

约简的主轴回归：SD线

Reduced major - axis regression

目标： $\min\limits_{a,b} \sum S_i$,

S_i为数据点(x_i, y_i)与直线$y = a + bx$之间的三角形面积

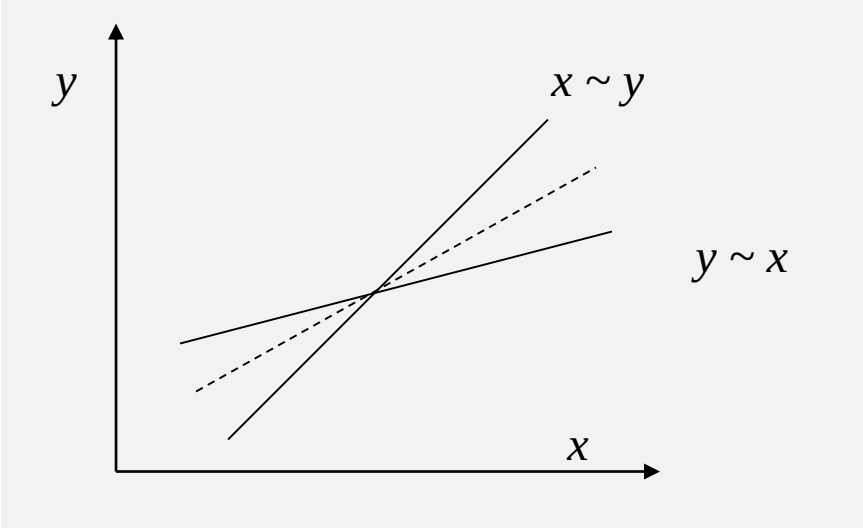


$\hat{b}_{RMA} = \text{sgn}(r)\sqrt{\frac{s_{yy}}{s_{xx}}}$，直线方程为SD Line：$y - \bar{y} = \text{sgn}(r)\sqrt{\frac{s_{yy}}{s_{xx}}}(x - \bar{x})$

$$\frac{y - \bar{y}}{\sqrt{s_{yy}}} = \text{sgn}(r)\left(\frac{x - \bar{x}}{\sqrt{s_{xx}}}\right)$$

Bisector regression
(double regression)

平分回归和逆回归之间的夹角



$$\hat{b}_{\text{bisect}} = \frac{\hat{b}_1\hat{b}_2 - 1 + \sqrt{(1+\hat{b}_1^2)(1+\hat{b}_2^2)}}{\hat{b}_1 + \hat{b}_2}$$

$\hat{b}_1, \hat{b}_2$分别是回归和逆回归得到的x的系数的估计。