回归分析 (01714601)
课程主页：http://staff.ustc.edu.cn/~ynyang/lm2020

第五讲 线性回归模型

2020.3.4

勾股定理：$\mathrm{var}(y)=\mathrm{var}\big(E(y\,|\,\mathbf{x})\big)+E\big(\mathrm{var}(y\,|\,\mathbf{x})\big)$

回归效应：$E\big(E(y\,|\,\mathbf{x})\big)=E(y)$

内容

- 响应变量与自变量
- 回归函数
- 线性回归模型
- 回归效应

响应变量与自变量

y： 响应变量(response, 1维, 感兴趣的变量);

$\mathbf{x}$： 自变量/因变量/协变量/解释变量(covariate, 1维或多维);

$E(y\,|\,\mathbf{x})$：回归函数,回归分析研究响应变量与自变量的关系。

其它名称：

y：dependent variable（相依变量）; ouput（输出）; predicted variable(被预测变量);

$\mathbf{x}$：indenpendent variable（独立变量）; input（输入）; predictor(预测变量);
feature(特征); controlled variable（控制变量）; regressor.

回归函数

条件期望$E(y|\mathbf{x})$称为回归函数,它是所有$\mathbf{x}$函数中，对y的最佳逼近(平方误差意义下，参见后面命题2)。

令$\varepsilon = y - E(y|\mathbf{x})$称为误差，则$y$可表示为:

$$y = E(y|\mathbf{x}) + \varepsilon$$

这是一般的回归方程，其中的误差项ε与回归函数$E(y|\mathbf{x})$不相关，所以它是一个"正交"分解(命题1).

假定回归函数具有特殊的形式就得到各种特殊的回归模型。
比如假设回归函数是$\mathbf{x}$的线性函数，我们得到线性回归模型.

命题1. 假设 y 是任一随机变量，$\mathbf{x}$是随机向量，令$\varepsilon = y - E(y\,|\,\mathbf{x})$，

则一般回归方程$y = E(y\,|\,\mathbf{x}) + \varepsilon$有如下性质：

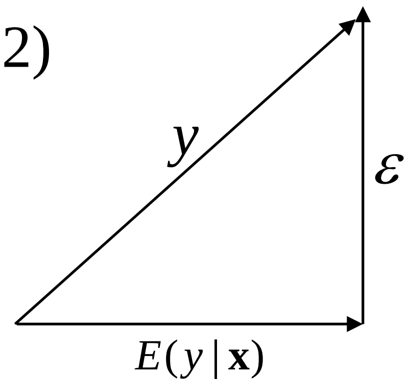
(1) $E(\varepsilon) = 0$

(2) "正交"：ε与$E(y\,|\,\mathbf{x})$不相关 (ε与$\mathbf{x}$不相关)

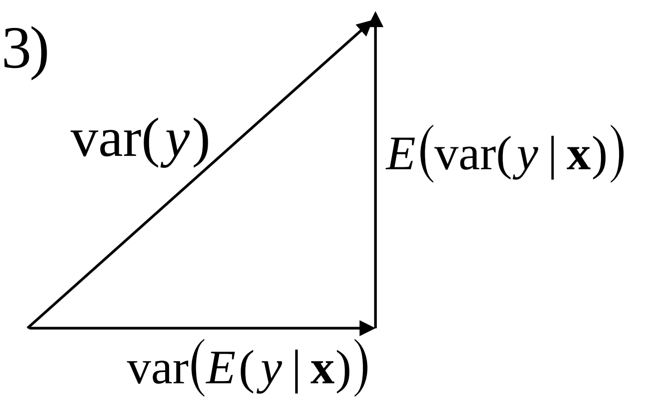
(3) 方差分解/勾股定理：

$$\mathrm{var}(y) = \mathrm{var}\big(E(y\,|\,\mathbf{x})\big) + \mathrm{var}(\varepsilon) = \mathrm{var}\big(E(y\,|\,\mathbf{x})\big) + E\big(\mathrm{var}(y\,|\,\mathbf{x})\big)$$

(2)



(3)



证明：(1)　由Tower property,　$E\left(E(y\mid\mathbf{x})\right)=E(y),$

$$\Rightarrow E(\varepsilon)=E(y)-E\left(E(y\mid\mathbf{x})\right)=0$$

(2) 记$\mathrm{g}(\mathbf{x})=E(y\mid\mathbf{x}),\quad \mu=E(g(\mathbf{x}))=E(y)$

$$\operatorname{cov}(g(\mathbf{x}),\varepsilon)=E(g(\mathbf{x})-\mu)\varepsilon=E(g(\mathbf{x})-\mu)(y-\mathrm{g}(\mathbf{x}))$$

$$=E\left[E(g(\mathbf{x})-\mu)(y-\mathrm{g}(\mathbf{x}))\mid\mathbf{x})\right]$$

Tower property

$$=E\left[\underline{(g(\mathbf{x})-\mu)}E\left((y-\mathrm{g}(\mathbf{x}))\mid\mathbf{x}\right)\right]$$

<u>常数项移到期望之外</u>

而$E\left((y-\mathrm{g}(\mathbf{x}))\mid\mathbf{x}\right)=E(y\mid\mathbf{x})-E\left(\mathrm{g}(\mathbf{x})\mid\mathbf{x}\right)=g(\mathbf{x})-g(\mathbf{x})=0$

所以　$\operatorname{cov}(g(\mathbf{x}),\varepsilon)=0$

(3) $y = g(\mathbf{x}) + \varepsilon$,因为$\operatorname{cov}(g(\mathbf{x}), \varepsilon) = 0 \Rightarrow \operatorname{var}(y) = \operatorname{var}\left(g(\mathbf{x})\right) + \operatorname{var}(\varepsilon)$,

只需验证: $\operatorname{var}(\varepsilon) = E\left(\operatorname{var}(y \mid \mathbf{x})\right)$.

$$\operatorname{var}(\varepsilon) = E\left(y^2 - 2yg(\mathbf{x}) + g(\mathbf{x})^2\right)$$
$$= E\left\{E\left(y^2 - 2yg(\mathbf{x}) + g(\mathbf{x})^2\right)\middle|\mathbf{x}\right\}$$

Tower property

$$= E\left\{E(y^2 \mid \mathbf{x}) - 2E(y \mid \mathbf{x})g(\mathbf{x}) + g(\mathbf{x})^2\right\}$$
$$= E\left\{E(y^2 - g(\mathbf{x})^2) \mid \mathbf{x})\right\}$$

$$\operatorname{var}(y \mid \mathbf{x}) = E\left[\left(y - g(\mathbf{x})\right)^2 \middle| \mathbf{x}\right]$$
$$= E\left(y^2 - g(\mathbf{x})^2)\middle|\mathbf{x}\right)$$

$$= E\left\{\operatorname{var}(y \mid \mathbf{x})\right\}$$

命题2. 以$\mathbf{x}$的函数$f(\mathbf{x})$逼近y，定义误差为$E(y-f(\mathbf{x}))^2$,则

$$E(y-f(\mathbf{x}))^2 \geq E(y-E(y|\mathbf{x}))^2,$$

即在$f(\mathbf{x})=E(y|\mathbf{x})$时误差最小。

证1：记$\mu=E(y)$. 我们知道$E(y-a)^2$的极小值在$a=\mu$处达到

(因为对任何常数$a, E(y-a)^2=E(y-\mu)^2+a^2 \geq E(y-\mu)^2$)。

$\Rightarrow$ 给定$\mathbf{x}$时，$E[(y-f(\mathbf{x}))^2|\mathbf{x}]$的极小值在$E(y|\mathbf{x})$处达到，

$\Rightarrow$ $E[(y-f(\mathbf{x}))^2]=E\{E[(y-f(\mathbf{x}))^2|\mathbf{x}]\}$极小值在$E(y|\mathbf{x})$处达到。

证2：记$g(\mathbf{x})=E(y\,|\,\mathbf{x})$. 首先$E(\mathrm{g}(\mathbf{x}))=E(y)$, $E(y-\mathrm{g}(\mathbf{x}))=0$.

$$E(y-f(\mathbf{x}))^2=E(y-g(\mathbf{x})+g(\mathbf{x})-f(\mathbf{x}))^2$$

$$=E(y-g(\mathbf{x}))^2+E(g(\mathbf{x})-f(\mathbf{x}))^2+\underline{2E(y-g(\mathbf{x}))(g(\mathbf{x})-f(\mathbf{x}))}$$

$$=E(y-g(\mathbf{x}))^2+E(g(\mathbf{x})-f(\mathbf{x}))^2\geq E(y-g(\mathbf{x}))^2$$

其中$E\underline{(y-g(\mathbf{x}))(g(\mathbf{x})-f(\mathbf{x}))}=0$,

这是因为$y-g(\mathbf{x})=y-E(y\,|\,\mathbf{x})$与任何$\mathbf{x}$的函数不相关(参见命题1(2))：

$$E\left([y-g(\mathbf{x})]\varphi(\mathbf{x})^{\top}\right)=E\left(E\left[(y-g(\mathbf{x}))\varphi(\mathbf{x})^{\top}\,|\,\mathbf{x}\right]\right)$$

$$=E\left(E\left[(y-g(\mathbf{x}))\,|\,\mathbf{x}\right]\varphi(\mathbf{x})^{\top}\right)=E\left(\left[E(y\,|\,\mathbf{x})-g(\mathbf{x})\right]\varphi(\mathbf{x})^{\top}\right)=0$$

线性回归模型

线性回归模型假设回归函数是线性的

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: $E(y\,|\,\mathbf{x}) = a + \mathbf{b}^{\top}\mathbf{x}$,

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

参数: 回归系数$a, \mathbf{b}$, 误差方差σ^2. 它们都需要估计。

- 简单线性回归模型: 自变量$\mathbf{x}$一维
  (simple linear regression model)
- 多重(不是多元)线性回归模型: 自变量$\mathbf{x}$多维
  (multiple linear regression model)

"多元线性回归"指的是响应为多维变量的情形,属于多元分析。

但通常，线性模型如下定义：

线性回归模型假设响应变量y和自变量(向量)$\mathbf{x}$满足方程：

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

其中误差项ε是不可观测的随机变量，满足Gauss-Markov假设

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

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

以及

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

注：

- Gauss-Markov假设(1),(2)简单记为$\varepsilon \sim (0, \sigma^2)$。
- 教科书上通常不提条件(3),是因为假定自变量$\mathbf{x}$给定。

两种定义方式基本等价(差别在于(3)和(3*))：

$(1),(2),(3) \Rightarrow (i),(ii)$

验证：$E(y\,|\,\mathbf{x}) = E(a + \mathbf{b}^\top \mathbf{x} + \varepsilon\,|\,\mathbf{x}) = a + \mathbf{b}^\top \mathbf{x} + E(\varepsilon\,|\,\mathbf{x}) = a + \mathbf{b}^\top \mathbf{x}$

$\mathrm{var}(y\,|\,\mathbf{x}) = \mathrm{var}(a + \mathbf{b}^\top \mathbf{x} + \varepsilon\,|\,\mathbf{x}) = \mathrm{var}(\varepsilon\,|\,\mathbf{x}) = \mathrm{var}(\varepsilon) = \sigma^2$

反之，令$\varepsilon = y - E(y\,|\,\mathbf{x}) = y - (a + \mathbf{b}^\top \mathbf{x})$，则

$(i),(ii) \Rightarrow (1),(2),(3^*)$：$\varepsilon$与$\mathbf{x}$不相关.

验证：$E(\varepsilon) = E(y - E(y\,|\,\mathbf{x})) = 0,$

$\mathrm{var}(\varepsilon) = \mathrm{var}(y - a - \mathbf{b}^\top \mathbf{x}) = \mathrm{var}(E(y - a - \mathbf{b}^\top \mathbf{x}\,|\,\mathbf{x})) + E(\mathrm{var}(y - a - \mathbf{b}^\top \mathbf{x}\,|\,\mathbf{x}))$

$= 0 + E(\mathrm{var}(y\,|\,\mathbf{x})) = \sigma^2.$

为什么线性？

没有一个确定的答案，或许最好的理由是:线性函数最简单

理由2： 联合正态情形下回归函数是线性的：

假设$\mathbf{x}, y$服从联合正态分布: $\begin{pmatrix} y \\ \mathbf{x} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mu_y \\ \mu_{\mathrm{x}} \end{pmatrix}, \begin{pmatrix} \Sigma_{yy} & \Sigma_{y\mathrm{x}} \\ \Sigma_{\mathrm{x}y} & \Sigma_{\mathrm{xx}} \end{pmatrix}\right)$,则

$$y \mid \mathbf{x} \sim N(\mu_y + \Sigma_{y\mathrm{x}}\Sigma_{\mathrm{xx}}^{-1}(\mathbf{x} - \mu_{\mathrm{x}}), \Sigma_{yy} - \Sigma_{y\mathrm{x}}\Sigma_{\mathrm{xx}}^{-1}\Sigma_{\mathrm{x}y}),$$

(i) $\mathrm{E}(y \mid \mathbf{x}) = \mu_y + \Sigma_{y\mathrm{x}}\Sigma_{\mathrm{xx}}^{-1}(\mathbf{x} - \mu_{\mathrm{x}}) \mathrel{\hat{=}} a + \mathbf{b}^{\top}\mathbf{x}$,

(ii) $\mathrm{var}(y \mid \mathbf{x}) = \Sigma_{yy} - \Sigma_{y\mathrm{x}}\Sigma_{\mathrm{xx}}^{-1}\Sigma_{\mathrm{x}y} \mathrel{\hat{=}} \sigma^2$.

(3) 令$\varepsilon = y - (a + \mathbf{b}^{\top}\mathbf{x})$,由$y \mid \mathbf{x} \sim N(a + \mathbf{b}^{\top}\mathbf{x}, \Sigma_{yy} - \Sigma_{y\mathrm{x}}\Sigma_{\mathrm{xx}}^{-1}\Sigma_{\mathrm{x}y}) \Rightarrow$

$\varepsilon \mid \mathbf{x} \sim N(0, \Sigma_{yy} - \Sigma_{y\mathrm{x}}\Sigma_{\mathrm{xx}}^{-1}\Sigma_{\mathrm{x}y})$与$\mathbf{x}$无关，故$\varepsilon$和$\mathbf{x}$独立。

理由3:分组数据可以用线性模型表示(例如两组正态)

如果z是因子变量,比如$z=0,1$代表两组,而每组内y服从正态:

$$y|_{\mathrm{z}=1} \sim N(\mu_1, \sigma^2), \quad y|_{\mathbf{x},\mathrm{z}=0} \sim N(\mu_0, \sigma^2)$$

记$a=\mu_0, b=\mu_1-\mu_0$,则上述两个正态分布可以表示为

$$y|_{\mathrm{z}} \sim N(\mu_0+(\mu_1-\mu_0)\mathrm{z}, \sigma^2) \triangleq N(a+b\mathrm{z}, \sigma^2),$$

(i) $E(y|z)=a+bz$

(ii) $\mathrm{var}(E(y|z)=\sigma^2$

均值中存在自变量时类似: $y|_{\mathbf{x},\mathrm{z}=1} \sim N(\alpha_1+\boldsymbol{\beta}^\top\mathbf{x}, \sigma^2), \quad y|_{\mathbf{x},\mathrm{z}=0} \sim N(\alpha_0+\boldsymbol{\beta}^\top\mathbf{x}, \sigma^2)$

$\Leftrightarrow y|_{\mathbf{x},\mathrm{z}} \sim N(a+b\mathrm{z}+\boldsymbol{\beta}^\top\mathbf{x}, \sigma^2),$

参数含义：不相关化

记号：$\mathbf{x}=(x_1,\ldots,x_{p\text{-}1})^{\top}$，$\mathbf{x}_{-k}=(x_1,\ldots,x_{k-1},x_{k+1},\ldots,x_{p\text{-}1})^{\top}$.

命题3. 假设线性回归模型：

$$y=\alpha+\boldsymbol{\beta}^{\top}\mathbf{x}+\varepsilon,\quad \varepsilon\sim(0,\sigma^2)\text{与}\mathbf{x}\text{独立}$$

其中$\boldsymbol{\beta}=(\beta_1,\ldots,\beta_{p\text{-}1})^{\top}$. 记$\mu_y=E(y)$，$\boldsymbol{\mu}_x=E(\mathbf{x})$，则

(1) $\boldsymbol{\beta}=\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y}$， $\alpha=\mu_y-\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\boldsymbol{\mu}_x$

从而$\varepsilon=(y-\mu_y)-\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}(\mathbf{x}-\boldsymbol{\mu}_x)$是$y$关于$\mathbf{x}$的不相关化。

(2) $\beta_k=\Sigma_{x_kx_k\bullet\mathbf{x}_{-k}}^{-1}\Sigma_{x_ky\bullet\mathbf{x}_{-k}}$

(3) $\sigma^2=\Sigma_{yy\bullet\mathbf{x}}=\Sigma_{yy}-\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y}$

$\Sigma_{\mathbf{uv}}=\operatorname{cov}(\mathbf{u},\mathbf{v})$

$y^{\perp}=y-\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\mathbf{x}$

$\Sigma_{\mathbf{uv}\bullet\mathbf{w}}=$

$\Sigma_{\mathbf{uv}}-\Sigma_{\mathbf{uw}}\Sigma_{\mathbf{ww}}^{-1}\Sigma_{\mathbf{wv}}$

注. 线性模型与第四讲中的不相关化本质上几乎完全相同，不同的是在这里多出了均值项，而在不相关化过程中我们仅考虑协方差，没必要考虑均值.

注1. 表达式$\boldsymbol{\beta}=\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y}$, $\alpha=\mu_y-\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\boldsymbol{\mu}_x, \sigma^2=\Sigma_{yy\bullet\mathbf{x}}$ 提示了应该如何估计参数$\boldsymbol{\beta},\alpha,\sigma^2$:

如果$\Sigma_{\mathbf{xx}},\Sigma_{\mathbf{x}y},\mu_y,\boldsymbol{\mu}_x$的样本估计为$S_{\mathbf{xx}},S_{\mathbf{x}y},\bar{y},\bar{\mathbf{x}}$,则参数可如下估计$(LS)$:

$$\hat{\boldsymbol{\beta}}=S_{\mathbf{xx}}^{-1}S_{\mathbf{x}y},\quad \hat{\alpha}=\bar{y}-S_{y\mathbf{x}}S_{\mathbf{xx}}^{-1}\bar{\mathbf{x}},\quad \hat{\sigma}^2=S_{yy\bullet\mathbf{x}}。$$

注2. β_k正比于偏相关系数。

$$\beta_k=\Sigma_{x_kx_k\bullet\mathbf{x}_{-k}}^{-1}\Sigma_{x_ky\bullet\mathbf{x}_{-k}}=\frac{\Sigma_{x_ky\bullet\mathbf{x}_{-k}}}{\sqrt{\Sigma_{x_kx_k\bullet\mathbf{x}_{-k}}}\sqrt{\Sigma_{yy\bullet\mathbf{x}_{-k}}}}\times\sqrt{\frac{\Sigma_{yy\bullet\mathbf{x}_{-k}}}{\Sigma_{x_kx_k\bullet\mathbf{x}_{-k}}}}$$

$$\propto\rho_{x_ky\bullet\mathbf{x}_{-k}}=y\text{与}x_k\text{的偏相关系数}$$

证：(1) 由 $0=\mathrm{cov}(\mathbf{x},\varepsilon)=\mathrm{cov}(\mathbf{x},y-\alpha-\boldsymbol{\beta}^{\top}\mathbf{x})=\mathrm{cov}(\mathbf{x},y)-\mathrm{cov}(\mathbf{x},\mathbf{x})\boldsymbol{\beta}$

$=\Sigma_{\mathbf{x}y}-\Sigma_{\mathbf{xx}}\boldsymbol{\beta}\Rightarrow\boldsymbol{\beta}=\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y}$。

回归方程两边取均值， 由$0=E(\varepsilon)$

$\mu_y=E(y)=\alpha+\boldsymbol{\beta}^{\top}E(\mathbf{x})=\alpha+\boldsymbol{\beta}^{\top}\boldsymbol{\mu}_x\Rightarrow\alpha=\mu_y-\boldsymbol{\beta}^{\top}\boldsymbol{\mu}_x$,

（3）回归方程两边取方差：

因为$\mathbf{x}$与ε独立

$$\Sigma_{yy}=\mathrm{var}(y)=\mathrm{var}(\alpha+\boldsymbol{\beta}^{\top}\mathbf{x}+\varepsilon)=\mathrm{var}(\boldsymbol{\beta}^{\top}\mathbf{x})+\mathrm{var}(\varepsilon)=\boldsymbol{\beta}^{\top}\Sigma_{\mathbf{xx}}\boldsymbol{\beta}+\sigma^2$$

$$\Rightarrow\sigma^2=\Sigma_{yy}-\boldsymbol{\beta}^{\top}\Sigma_{\mathbf{xx}}\boldsymbol{\beta}=\Sigma_{yy}-\Sigma_{y\mathbf{x}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y}=\Sigma_{yy\bullet\mathbf{x}}$$。

(2) 考察$k=1$情形（一般情形，只需把$\boldsymbol{\beta}$和$\mathbf{x}$同时置换一下），

为了简单，记$\boldsymbol{\beta}_2=\boldsymbol{\beta}_{-1}$，$\mathbf{x}_2=\mathbf{x}_{-1}$，即$\boldsymbol{\beta}=\begin{pmatrix}\beta_1\\ \boldsymbol{\beta}_{-1}\end{pmatrix}=\begin{pmatrix}\beta_1\\ \boldsymbol{\beta}_2\end{pmatrix}$, $\mathbf{x}=\begin{pmatrix}x_1\\ \mathbf{x}_{-1}\end{pmatrix}=\begin{pmatrix}x_1\\ \mathbf{x}_2\end{pmatrix}$,

模型为$y=\alpha+\boldsymbol{\beta}\mathbf{x}+\varepsilon=\alpha+\beta_1x_1+\boldsymbol{\beta}_2\mathbf{x}_2+\varepsilon$

首先将x_1关于$\mathbf{x}_2$不相关化，令

$$x_1^{\perp}=x_1-\Sigma_{x_1\mathbf{x}_2}\Sigma_{\mathbf{x}_2\mathbf{x}_2}^{-1}\mathbf{x}_2,\quad \mathrm{cov}(x_1^{\perp},\mathbf{x}_2)=0,$$

类似于(1)我们利用$\mathrm{cov}(\mathbf{x},\varepsilon)=0$求$\beta_1$的表达，这里我们不需要$\boldsymbol{\beta}_2$，故将$x_1$关于$\mathbf{x}_2$不相关化.

因为ε与$\mathbf{x}$不相关，而$x_1^{\perp}$是$\mathbf{x}=\begin{pmatrix}x_1\\ \mathbf{x}_2\end{pmatrix}$的线性组合，故$\varepsilon$与$x_1^{\perp}$不相关

细节：由$x_1^{\perp}=x_1-\Sigma_{x_1\mathbf{x}_2}\Sigma_{\mathbf{x}_2\mathbf{x}_2}^{-1}\mathbf{x}_2=(1,-\Sigma_{x_1\mathbf{x}_2}\Sigma_{\mathbf{x}_2\mathbf{x}_2}^{-1})\begin{pmatrix}x_1\\ \mathbf{x}_2\end{pmatrix}\hat{=}A\mathbf{x}$,

以及$\mathrm{cov}(\mathbf{x},\varepsilon)=0\Rightarrow\mathrm{cov}(x_1^{\perp},\varepsilon)=\mathrm{cov}(A\mathbf{x},\varepsilon)=0$

另一方面，

$$0=\mathrm{cov}(x_1^{\perp},\varepsilon)=\mathrm{cov}(x_1^{\perp},y-\alpha-\beta_1x_1-\boldsymbol{\beta}_2\mathbf{x}_2)$$

$$=\mathrm{cov}(x_1^{\perp},y)-\mathrm{cov}(x_1^{\perp},x_1)\beta_1-\cancel{\mathrm{cov}(x_1^{\perp},\mathbf{x}_2)\boldsymbol{\beta}_2^{\mathsf{T}}}$$

$$=\mathrm{cov}(x_1^{\perp},y^{\perp})-\mathrm{cov}(x_1^{\perp},x_1^{\perp})\beta_1$$

其中$y^{\perp}=y-\Sigma_{y\mathbf{x}_2}\Sigma_{\mathbf{x}_2\mathbf{x}_2}^{-1}\mathbf{x}_2$

因为

$\mathrm{cov}(x_1^{\perp},x_1-x_1^{\perp})=0$

$\mathrm{cov}(x_1^{\perp},y-y^{\perp})=0$

所以$\beta_1=\mathrm{var}(x_1^{\perp})^{-1}\mathrm{cov}(y^{\perp},x_1^{\perp})=\mathrm{cov}(y^{\perp},x_1^{\perp})/\mathrm{var}(x_1^{\perp})$

而 $\mathrm{var}(x_1^{\perp})=\Sigma_{x_kx_k\bullet\mathbf{x}_{-k}}$， $\mathrm{cov}(y^{\perp},x_1^{\perp})=\Sigma_{x_1y\bullet\mathbf{x}_{-1}}$

$$\Rightarrow\beta_1=\Sigma_{x_1x_1\bullet\mathbf{x}_{-1}}^{-1}\Sigma_{x_1y\bullet\mathbf{x}_{-1}}$$

证明2(利用分块矩阵的逆矩阵公式化):

$\mathbf{x}=\begin{pmatrix} x_1 \\ \mathbf{x}_2 \end{pmatrix}$,为了叙述简单，其协方差阵记为$\Sigma=\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, $\Sigma_{\mathbf{x}y}=\begin{pmatrix} \Sigma_{1y} \\ \Sigma_{2y} \end{pmatrix}$。

由(1), $\boldsymbol{\beta}=\begin{pmatrix} \beta_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}=\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{x}y}=\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}^{-1}\begin{pmatrix} \Sigma_{1y} \\ \Sigma_{2y} \end{pmatrix}$

$$=\begin{pmatrix} \Sigma_{11\bullet 2}^{-1} & -\Sigma_{11\bullet 2}^{-1}\Sigma_{12}\Sigma_{22}^{-1} \\ * & * \end{pmatrix}\begin{pmatrix} \Sigma_{1y} \\ \Sigma_{2y} \end{pmatrix}=\begin{pmatrix} \Sigma_{11\bullet 2}^{-1}\Sigma_{1y}-\Sigma_{11\bullet 2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{2y} \\ * \end{pmatrix}$$

故$\beta_1=\Sigma_{11\bullet 2}^{-1}\Sigma_{1y}-\Sigma_{11\bullet 2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{2y}=\Sigma_{11\bullet 2}^{-1}(\Sigma_{1y}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{2y})=\Sigma_{11\bullet 2}^{-1}\Sigma_{1y\bullet 2}$

回到原来的记号，此即$\beta_1=\Sigma_{x_1x_1\bullet\mathbf{x}_{-1}}^{-1}\Sigma_{x_1y\bullet\mathbf{x}_{-1}}$

命题3(特例):假设简单线性模型 $y = a + bx + \varepsilon, \varepsilon \sim (0, \sigma^2)$, ε与x独立.
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) $\sigma^2 = (1 - \rho^2)\sigma_y^2$

回归直线$y = a + bx$即为: $\dfrac{y - \mu_y}{\sigma_y} = \rho \times \dfrac{x - \mu_x}{\sigma_x}$

回归效应（regression towards the mean）

弗朗西斯.高尔顿研究儿子身高与父亲身高关系时发现，
儿子身高（y轴）与父亲相比有趋中的趋势：

如果父亲很高，则儿子很可能会相对矮一些；
如果父亲很矮，则儿子可能会相对高一些。

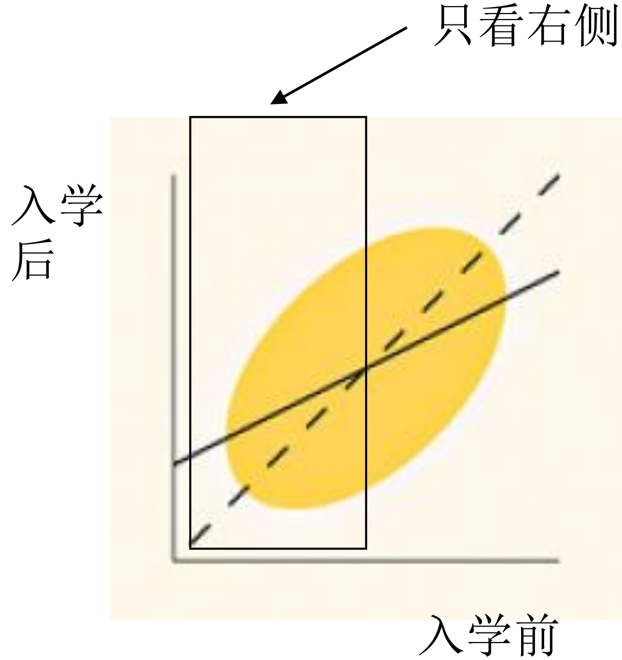
他称这种现象为回归（实线）。

如果x和y地位对等，那么上图数据的最佳直线逼近应是虚线。

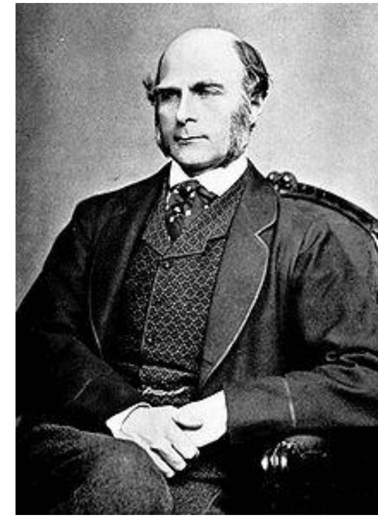
但上述研究中关心的是儿子身高，父亲的身高x作为解释变量（以变量x的线性函数逼近响应y），两者的地位是不对等的。最佳逼近直线-回归直线-是实线而不是虚线。

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。



弗朗西斯. 高尔顿 (Francis Galton)



弗朗西斯高尔顿爵士 **(1822 –1911)**是达尔文**(Charles Darwin)**的表弟，英国维多利亚时期的博物学家**(polymath)**：

Anthropologist, eugenicist, tropical explorer, geographer, inventor, meteorologist, proto-geneticist, psychometrician, and statistician.

他自称为**private gentleman,** 建立了众多学科或概念**(**包括统计**):**

Regression toward the mean (回归**), Correlation (**相关系数**), Standard deviation (**标准差**), Galton board (**高尔顿板**), eugenics (**优生学**), Weather map (**气象图**), Fingerprint (**指纹**), nature vs nurture (**先天与后天）**…**

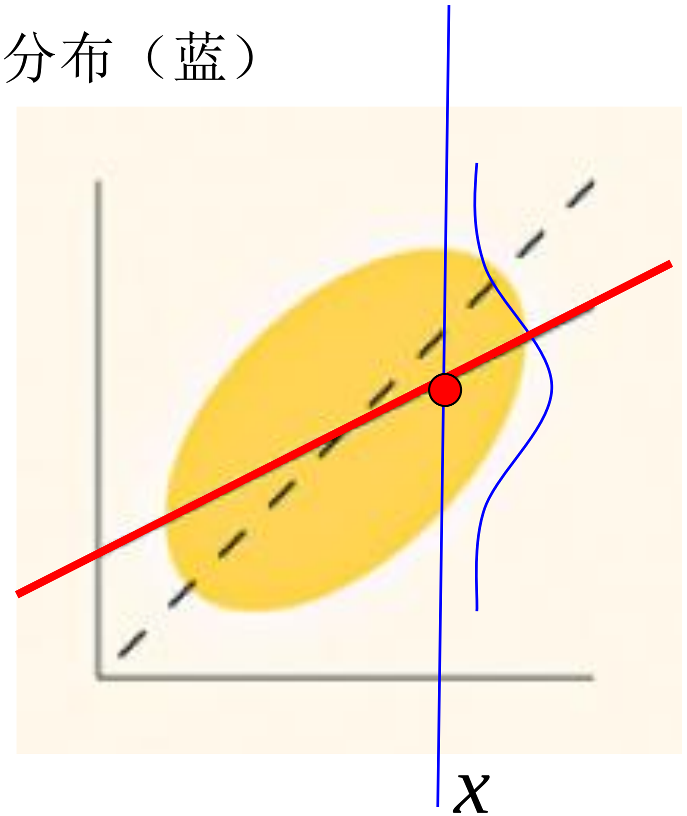
SD线

假设(x,y)服从二元正态分布，其分布形状如左图

虚线称为SD线：$y=\mu_y+\dfrac{\sigma_y}{\sigma_x}(x-\mu_x)$ 或 $\dfrac{y-\mu_y}{\sigma_y}=\dfrac{x-\mu_x}{\sigma_x}$

它表示x偏离中心一个单位σ_x时，y也偏离其中心一个单位σ_y

对给定的x,假设变量y服从图中所示的正态分布（蓝）



该正态分布的中心(红点):

$$E(y\,|\,x) = \mu_y + \rho\frac{\sigma_y}{\sigma_x}\left(x - \mu_x\right)$$

x变化时，形成回归直线(红):

$$\frac{y-\mu_y}{\sigma_y} = \rho \times \frac{x-\mu_x}{\sigma_x}$$

相比于虚线(虚线没有ρ: $\frac{y-\mu_y}{\sigma_y} = \frac{x-\mu_x}{\sigma_x}$),

回归直线在两端有向中心回归的趋势。

上页给出了直观，现在严格证实一下：

虚线（SD线）: $\dfrac{y-\mu_y}{\sigma_y}=\dfrac{x-\mu_x}{\sigma_x}$

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

$$y=a+bx+\varepsilon, \varepsilon\sim(0,\sigma^2)\text{与}x\text{独立}$$

如果已知x超过其均值$k>0$个标准差（$x=\mu_x+k\sigma_x$),

$$\text{则}E(y\,|\,x=\mu_x+k\sigma_x)=\mu_y+bk\sigma_x=\mu_y+k\rho_{xy}\sigma_y<\mu_y+k\sigma_y$$

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

Tower property 可能是回归效应的根本原因： $E\left(E(y \mid x)\right)=E(y)$

对于一般的(未必线性的)回归函数 $E(y \mid x)$，是否也有回归现象？
比如何时会有如下回归现象？

$$\frac{E(y \mid x)-E(y)}{\sigma_y}=k \times \frac{x-E(x)}{\sigma_x}, |k|<1$$

或更一般地，假设x, y同分布，对于$c>\mu$，何时会有下述回归现象

$$\mu<E(y \mid x=c)<c ?$$