一元线性回归模型的参数估计

高中统计学 · 第捌课

何濯羽 · 2024年2月23日

目录
CONTENT

第一节

一元线性回归模型

什么是线性回归模型?

根据“分解性质”，任一随机变量 Y 可以被分解为

$$Y = \underbrace{E(Y|X)}_{\text{条件期望函数}} + \underbrace{e}_{\text{误差项}}$$

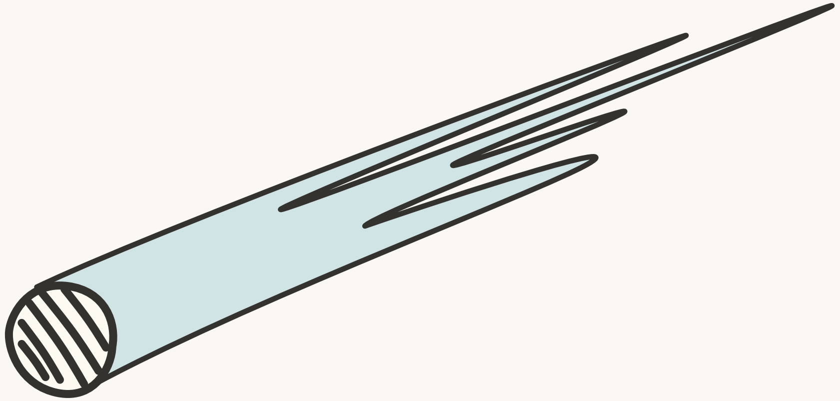
我们的目标是找到条件期望函数 $m(X) = E(Y|X)$ 的解析式。

基于理论知识，或者出于运算简便的考虑，我们假设 $m(X)$ 是一个线性函数：$m(X) = \alpha + \beta X$。

线性回归模型：

$$\overbrace{Y}^{\text{因变量}} = \underbrace{\alpha}_{\text{常数项}} + \underbrace{\beta}_{\text{自变量的系数}}\overbrace{X}^{\text{自变量}} + \overbrace{e}^{\text{误差项}}$$

什么是一元线性回归模型?



当线性回归模型只有一个自变量时，我们称其为“**一元线性回归模型**” （simple linear regression model）。

$$Y = \alpha + \beta X_1 + e$$

当线性回归模型有两个或两个以上自变量时，我们称其为“**多元线性回归模型**” （general linear regression model）。

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + e$$

在高中阶段，我们仅学习一元线性回归模型的相关知识。

第二节

最小二乘法

如何估计一元线性回归模型的参数?

一元线性回归模型一共有 2 个参数亟待估计：α 和 β。

在本节与下一节，我们将学习两种常见的构建估计量的方法——它们的结果相同，但路径不同。

考虑到数学教材的“权威性”（😔），我们先学习“**最小二乘法**”（method of least square）。

最小二乘法的思路：

我们的目标是寻找 α 和 β 的取值（记作 $\hat{\alpha}$ 和 $\hat{\beta}$），使得下方的二元函数取得最小值：

$$S(\alpha,\beta)=E[(Y-\alpha-\beta X)^2]=E(e^2)$$

由于我们无法直接观测总体（即各随机变量的联合概率分布和矩），我们可以选择先用样本数据估计 $S(\alpha,\beta)$，记作 $\hat{S}(\alpha,\beta)$，然后寻找 α 和 β 的合适取值使得 $\hat{S}(\alpha,\beta)$ 取得最小值。

根据样本类似原则，我们可以构建如下的估计量：

$$\hat{S}(\alpha,\beta)=\frac{1}{n}\sum_{i=1}^{n}(Y_i-\alpha-\beta X_i)^2$$

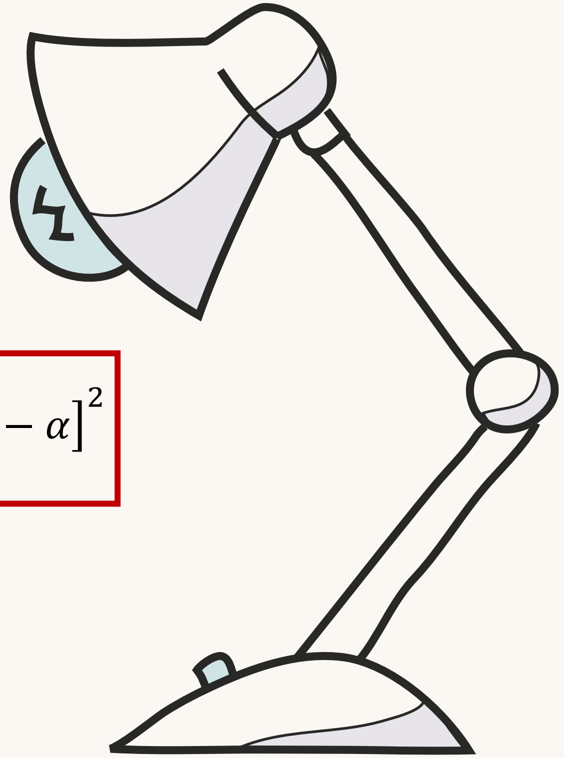
最小二乘法

如何最小化 $\hat{S}(\alpha,\beta)=\frac{1}{n}\sum_{i=1}^{n}(Y_i-\alpha-\beta X_i)^2$?

首先，$\frac{1}{n}$ 与 α,β 的取值无关，所以我们的目标转变为：寻找 α,β 使得 $\hat{Q}(\alpha,\beta)=\sum_{i=1}^{n}(Y_i-\alpha-\beta X_i)^2$ 最小化。

接着，开始枯燥的数学推导：

$$
\begin{aligned}
\hat{Q}(\alpha,\beta) &= \sum_{i=1}^{n}\left[Y_i-\beta X_i-(\bar{Y}-\beta\bar{X})+(\bar{Y}-\beta\bar{X})-\alpha\right]^2 \\
&= \sum_{i=1}^{n}\left[(Y_i-\bar{Y})-\beta(X_i-\bar{X})+(\bar{Y}-\beta\bar{X})-\alpha\right]^2 \\
&= \underbrace{\boxed{\sum_{i=1}^{n}\left[(Y_i-\bar{Y})-\beta(X_i-\bar{X})\right]^2}}_{\text{①}}+\underbrace{\boxed{2\sum_{i=1}^{n}\left[(Y_i-\bar{Y})-\beta(X_i-\bar{X})\right]\cdot\left[(\bar{Y}-\beta\bar{X})-\alpha\right]}}_{\text{②}}+\underbrace{\boxed{\sum_{i=1}^{n}\left[(\bar{Y}-\beta\bar{X})-\alpha\right]^2}}_{\text{③}}
\end{aligned}
$$



最小二乘法

第②项

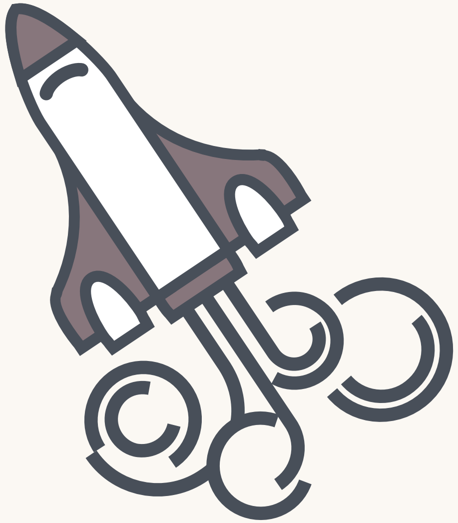
$$\sum_{i=1}^{n}\left[\left(Y_i-\bar{Y}\right)-\beta\left(X_i-\bar{X}\right)\right]\cdot\left[\left(\bar{Y}-\beta\bar{X}\right)-\alpha\right]$$

$$=\left(\bar{Y}-\beta\bar{X}-\alpha\right)\sum_{i=1}^{n}\left[\left(Y_i-\bar{Y}\right)-\beta\left(X_i-\bar{X}\right)\right]$$

$$=\left(\bar{Y}-\beta\bar{X}-\alpha\right)\left[\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)-\beta\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\right]$$

$$=\left(\bar{Y}-\beta\bar{X}-\alpha\right)\left[\left(n\bar{Y}-n\bar{Y}\right)-\beta\left(n\bar{X}-n\bar{X}\right)\right]$$

$$=0$$



最小二乘法

这样，$\hat{Q}(\alpha,\beta)$ 仅剩下第①项和第③项：

$$\hat{Q}(\alpha,\beta)=\underbrace{\sum_{i=1}^{n}\left[(Y_i-\overline{Y})-\beta(X_i-\overline{X})\right]^2}_{①}+\underbrace{\sum_{i=1}^{n}\left[(\overline{Y}-\beta\overline{X})-\alpha\right]^2}_{③}$$

其中，第③项仅与 α 有关，与 β 无关。因此，要使得 $\hat{Q}(\alpha,\beta)$ 最小，我们必须选择使得第③项为 0 的 α 值。这样，α 的取值为

$$\hat{\alpha}=\overline{Y}-\hat{\beta}\overline{X}$$

将 $\hat{\alpha}$ 和 $\hat{\beta}$ 代入 $\hat{Q}(\alpha,\beta)$ 后， 我们得到

$$\hat{Q}(\hat{\alpha},\hat{\beta})=\sum_{i=1}^{n}\left[(Y_i-\overline{Y})-\hat{\beta}(X_i-\overline{X})\right]^2=\sum_{i=1}^{n}(Y_i-\overline{Y})^2-2\hat{\beta}\sum_{i=1}^{n}(Y_i-\overline{Y})(X_i-\overline{X})+\hat{\beta}^2\sum_{i=1}^{n}(X_i-\overline{X})^2$$

仔细观察后，我们发现：$\hat{Q}(\hat{\alpha},\hat{\beta})$ 是一个关于 $\hat{\beta}$ 的开口向上的一元二次函数。因此，$\hat{Q}(\hat{\alpha},\hat{\beta})$ 的最小值在其对称轴上取得：

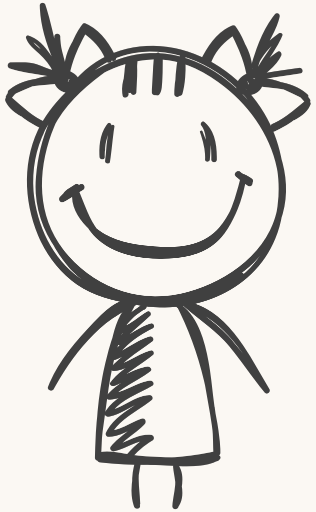
$$\hat{\beta}=\frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sum_{i=1}^{n}(X_i-\overline{X})^2}$$

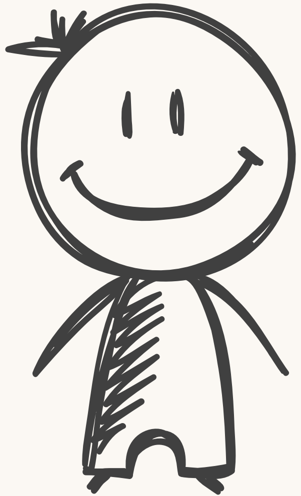
第三节

矩量估计法

与各矩量对应的样本估计量

矩量（总体）	估计量（样本）
$E(X)=\mu_X=\sum_{x\in\mathcal{X}} x\cdot \Pr(X=x)$	$\overline{X}=\frac{1}{n}\sum_{i=1}^{n} X_i$
$D(X)=\sigma_X^2=\sum_{x\in\mathcal{X}} (x-\mu_X)^2\cdot \Pr(X=x)$	$\hat{\sigma}_X^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$
$Cov(X,Y)=\sigma_{XY}=E\{[X-E(X)]\cdot[Y-E(Y)]\}$	$\hat{\sigma}_{XY}=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)\left(Y_i-\overline{Y}\right)$





矩量估计法

第二种构建一元线性回归模型中参数估计量的方法是“**矩量法**”（method of moments）。
简而言之，矩量法就是先将参数们用模型中相关的随机变量的矩表示出来（即“识别”），然后根据参数的表达式，直接构造估计量。

在第柒课中，我们已经对一元线性回归模型的参数进行了识别：

$$\alpha = E(Y) - \beta \cdot E(X) \qquad \beta = \frac{Cov(X,Y)}{D(X)}$$

于是，根据样本类似原则，我们可以构建如下的估计量：

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X} \qquad \hat{\beta} = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

这些估计量与通过最小二乘法构造的估计量一模一样！

第四节

最小二乘估计的无偏性

高斯-马尔可夫定理（简化版）

在一元线性回归模型 $Y = \alpha + \beta X + e$ 中，如果以下假设均成立，那么模型参数的最小二乘估计量就是它们的**最优无偏估计量。**“最优”指的是该估计量是所有无偏估计量中（条件）方差最小的。

1 独立同分布

样本 $(X_i, Y_i)_{i=1}^{n}$ 中每一对随机变量 (X, Y) 均相互独立，且服从同一个概率分布。

2 方差非零

自变量 X 的方差不等于 0，即 $D(X) > 0$。

3 外生性

误差项的条件期望等于 0，即 $E(e|X) = 0$。

4 同质性

误差项的条件方差为一个定值，即 $D(e|X) = \sigma^2 > 0$。

该定理经历了三个世纪的完善：由德国数学家高斯（Carl F. Gauss）于1823年提出，由俄国数学家马尔可夫（Andrey A. Markov）于1912年进行无偏性的规范化，再由美国计量经济学家汉森（Bruce E. Hansen）于2022年推广至非线性无偏估计量的领域。

第五节

线性模型的拟合效果

拟合值 & 残差

在我们将样本数据代入一元线性回归模型参数的最小二乘估计量后，我们会得到它们的估计值：$\hat{\alpha}$ 和 $\hat{\beta}$。

于是，我们可以计算

$$\hat{Y} = \hat{\alpha} + \hat{\beta} X$$

$\hat{Y}$ 被称为“**拟合值**”（fitted value）——“拟合”指的是已知某函数的若干离散函数值（来自样本数据），通过调整该函数中若干系数（如 α 和 β），使得该函数与已知样本点的差距最小。这实际上就是最小二乘法的计算原理。

我们把 Y 的观测值（observed value）与拟合值的差称为“**残差**”（residual）：

$$\hat{e} = Y - \hat{Y}$$

样本中，每一个个体都有自己独特的观测值 Y_i 和拟合值 $\hat{Y}_i$，所以每一个个体也有自己独特的残差 $\hat{e}_i$。

平方和（Sum of Squares）

✦**残差平方和** $$SS_{\text{res}}=\sum_{i=1}^{n}\hat{e}_i^2=\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)^2$$

✦**模型平方和** $$SS_{\text{reg}}=\sum_{i=1}^{n}\left(\hat{Y}_i-\overline{Y}\right)^2$$

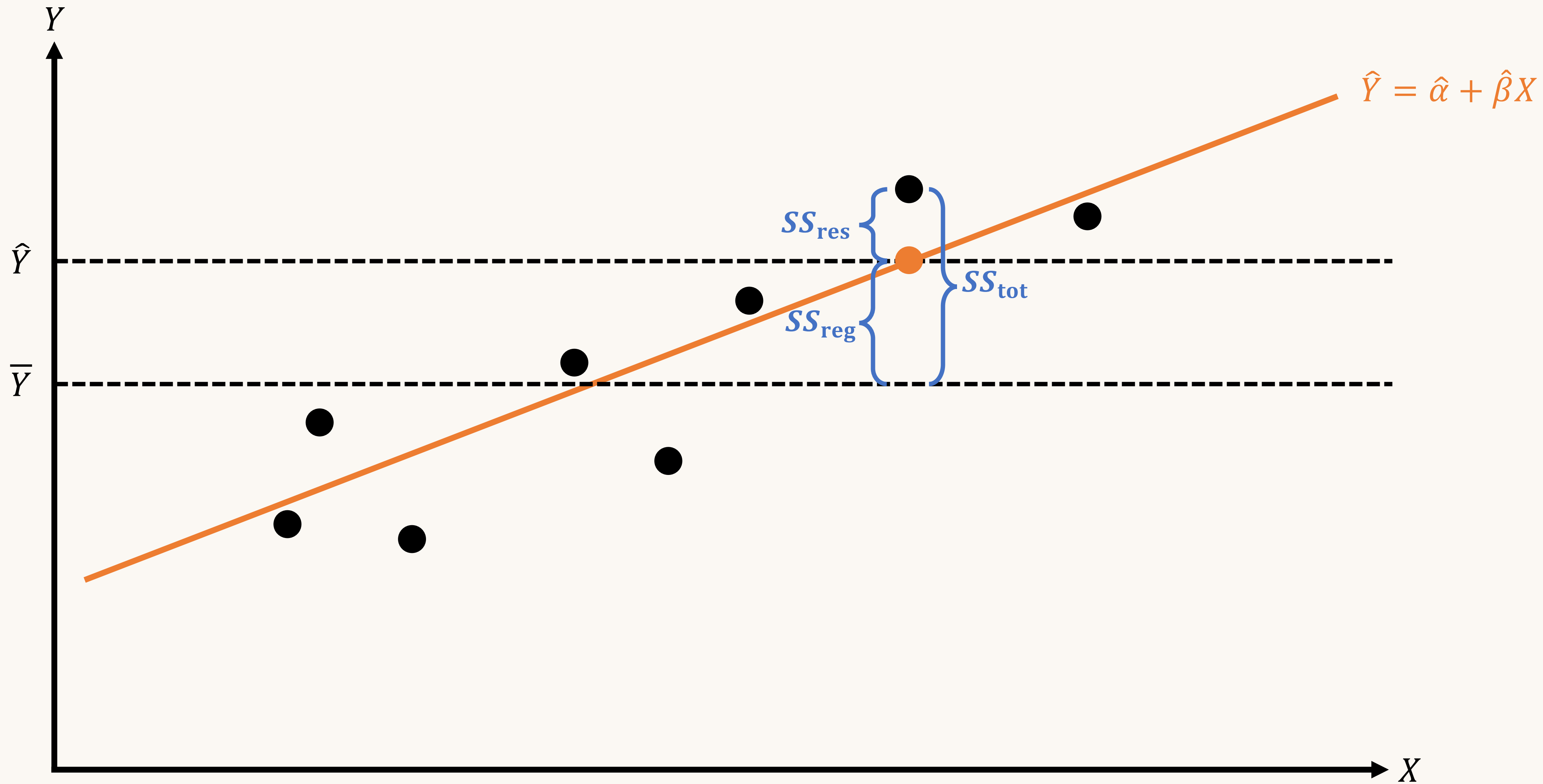
✦**总平方和** $$SS_{\text{tot}}=\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2$$

当回归模型为（一元或多元）线性模型，且包含常数项时，以下等式恒成立：

$$SS_{\text{tot}}=SS_{\text{res}}+SS_{\text{reg}}\quad\text{即}\quad\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2=\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)^2+\sum_{i=1}^{n}\left(\hat{Y}_i-\overline{Y}\right)^2$$

注：该等式的证明过程较为复杂，在此不做展示。

平方和的图示



Y
$\hat{Y} = \hat{\alpha} + \hat{\beta}X$
SS_{res}
$\hat{Y}$
SS_{tot}
SS_{reg}
$\overline{Y}$
X

决定系数

决定系数（coefficient of determination），又称“**R方**”（R squared），是衡量模型拟合效果的工具。

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}$$

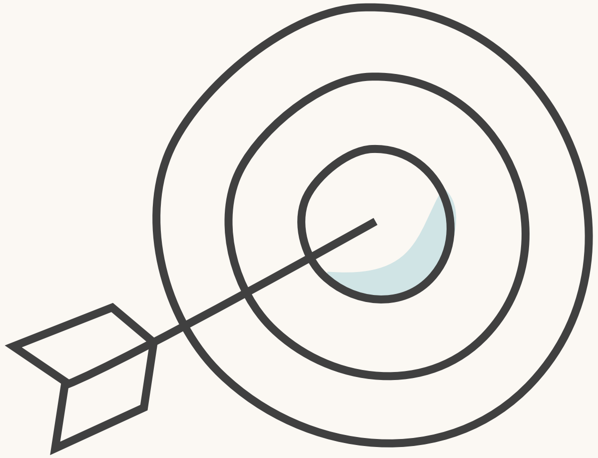
当回归模型为（一元或多元）线性模型，且包含常数项时，我们有

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} = \frac{SS_{\text{tot}} - SS_{\text{res}}}{SS_{\text{tot}}} = \frac{SS_{\text{reg}}}{SS_{\text{res}} + SS_{\text{reg}}}$$

因为 $SS_{\text{reg}} \geq 0$，$SS_{\text{res}} \geq 0$，所以 $R^2 \in [0, 1]$。

$R^2 = 1$（即 $SS_{\text{res}} = 0$）表明模型的拟合值完美契合了样本的观测值。

注意：具有完美拟合效果的模型 $\neq$ 好模型！



R方的值域

一般情况下，我们有 $SS_{\mathrm{tot}} \neq SS_{\mathrm{res}} + SS_{\mathrm{reg}}$。因此，$R^2$ 的值域不会局限在区间 $[0,1]$ 内。

$$R^2 = 1 - \frac{SS_{\mathrm{res}}}{SS_{\mathrm{tot}}} = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}$$

因为 $SS_{\mathrm{res}} \geq 0$， $SS_{\mathrm{tot}} \geq 0$，所以 $R^2 \leq 1$。

当一个模型过分糟糕时（例如：真实的 CEF 是非线性的而我们却使用了线性模型），我们会发现

$$SS_{\mathrm{res}} > SS_{\mathrm{tot}}$$

于是，$R < 0$。

综上，在一般情况下，$R^2 \in (-\infty, 1]$。

关于R方的警告

1. R^2 不能用于检测我们所使用用模型的正确性。

2. R^2 不能说明我们是否已包含正确的自变量。

3. R^2 不能说明我们是否已收集足够多的样本。

4. R^2 不能解释自变量 X 与因变量 Y 的因果关系。

因为 R^2 在实际应用方面存在极大的误导性，而且近年来 R^2 在学术界的受重视程度愈来愈低，所以我个人完全不能理解人教版 2019A 新教材为何引入 R^2 的相关知识。

感谢各位的倾听

Thanks for your listening!

何濯羽　2024年2月23日