天善智能
TIANSHAN SOFT

专注于商业智能BI和大数据的垂直社区平台

一元线性回归的参数估计

Allen

www.hellobi.com

课程目录

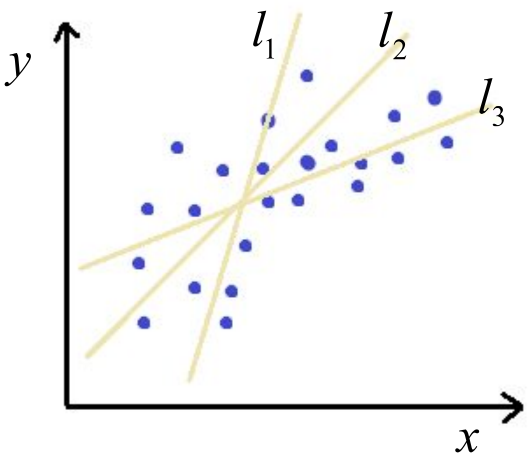
- 最小二乘法
- 回归参数的最小二乘估计
- 最小二乘估计量的性质
- 小结

天善智能
TIANSHAN SOFT

最小二乘法

- 一元线性回归模型可表示为:$y = \beta_0 + \beta_1 x + \varepsilon$
- 一元线性回归中估计的回归方程为: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

如果不加限制，通过样本点，可以求出多个回归方程



怎么选取估计的回归方程呢？

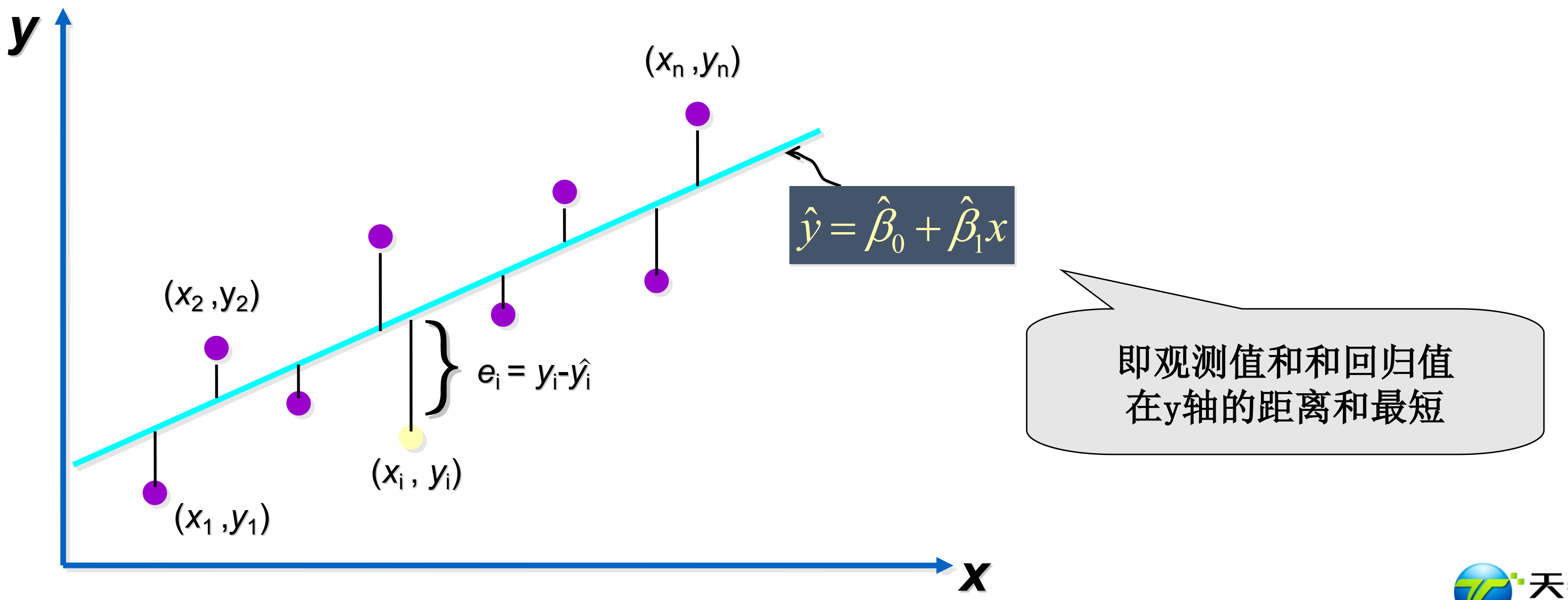
最小二乘法

- 现假设已知一元线性回归中估计的回归方程为: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$
- 若有一个样本点 (x_i, y_i) ，样本点带入估计的回归方程中求得的估计值为 $\hat{y}_i$
- 很自然的想法就是使估计值 $\hat{y}_i$ 与真实观测值 y_i 的差 $e_i = y_i - \hat{y}_i$ 尽量小，把 $e_i = y_i - \hat{y}_i$ 称为**残差**，最小二乘法就是选择一条直线使得残差平方和最小

问题转化 → $\min\left(\sum e_i\right)$ → $\min\left(\sum |e_i|\right)$ → $\min\left(\sum e_i^2\right)$

最小二乘法

- 想要使残差和最小，即 $\min\left(\sum e_i\right)$或者 $\min\left(\sum |e_i|\right)$或者 $\min\left(\sum e_i^2\right)$ ，如图



最小二乘法

- 想要使残差平方和可表示为 $Q(\hat{\beta}_0,\hat{\beta}_1)=\sum e_i^2=\sum(y_i-\hat{y}_i)^2=\sum(y_i-\hat{\beta}_0-\hat{\beta}_1x_i)^2$
- 要使上式达到最小，根据求极值原理可得如下方程组：

$$\begin{cases}\dfrac{\partial Q(\hat{\beta}_0,\hat{\beta}_1)}{\partial\hat{\beta}_0}==\dfrac{\partial\sum(y_i-\hat{\beta}_0-\hat{\beta}_1x_i)^2}{\partial\hat{\beta}_0}=0\\ \dfrac{\partial Q(\hat{\beta}_0,\hat{\beta}_1)}{\partial\hat{\beta}_1}==\dfrac{\partial\sum(y_i-\hat{\beta}_0-\hat{\beta}_1x_i)^2}{\partial\hat{\beta}_1}=0\end{cases}$$

天善智能 TIANSHAN SOFT

最小二乘法

- 将以上方程组整理得：$\begin{cases} \sum\left(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i\right)=0 \\ \sum\left(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i\right)x_i=0 \end{cases}$

- 解以上方程组得：$\begin{cases} \hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x} \\ \hat{\beta}_1=\dfrac{n\sum x_i y_i-\sum x_i\sum y_i}{n\sum x_i^2-\left(\sum x_i\right)^2}=\dfrac{\sum\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sum\left(x_i-\bar{x}\right)^2} \end{cases}$

- 通常称$\hat{\beta}_0$和$\hat{\beta}_1$为β_0和β_1的最小二乘估计

最小二乘估计量的性质——期望

- 1.求 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的均值：
$$\begin{cases} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 = \dfrac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \end{cases}$$
- 首先将 $\hat{\beta}_1$ 化简得：$\hat{\beta}_1 = \dfrac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \dfrac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2} - \dfrac{\sum (x_i - \bar{x}) \bar{y}}{\sum (x_i - \bar{x})^2}$
- 由于：$\sum (x_i - \bar{x}) = \sum x_i - \sum \bar{x} = n\bar{x} - n\bar{x} = 0$
- 化简：$\hat{\beta}_1 = \dfrac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2} = \dfrac{\sum (x_i - \bar{x})(\beta_0 + \beta_1 x_i + \varepsilon_i)}{\sum (x_i - \bar{x})^2} = \dfrac{\beta_1 \sum (x_i - \bar{x}) x_i + \sum (x_i - \bar{x}) \varepsilon_i}{\sum (x_i - \bar{x})^2}$

最小二乘估计量的性质——期望

$$\hat{\beta}_1 = \frac{\sum\left(x_i - \bar{x}\right)y_i}{\sum\left(x_i - \bar{x}\right)^2} = \frac{\sum\left(x_i - \bar{x}\right)\left(\beta_0 + \beta_1 x_i + \varepsilon_i\right)}{\sum\left(x_i - \bar{x}\right)^2} = \frac{\beta_1 \sum\left(x_i - \bar{x}\right)x_i + \sum\left(x_i - \bar{x}\right)\varepsilon_i}{\sum\left(x_i - \bar{x}\right)^2}$$

$$= \frac{\beta_1 \sum\left(x_i - \bar{x}\right)x_i - \beta_1 \sum\left(x_i - \bar{x}\right)\bar{x} + \beta_1 \sum\left(x_i - \bar{x}\right)\bar{x} + \sum\left(x_i - \bar{x}\right)\varepsilon_i}{\sum\left(x_i - \bar{x}\right)^2} = \frac{\beta_1 \sum\left(x_i - \bar{x}\right)^2 + \beta_1 \sum\left(x_i - \bar{x}\right)\bar{x} + \sum\left(x_i - \bar{x}\right)\varepsilon_i}{\sum\left(x_i - \bar{x}\right)^2}$$

$$= \beta_1 + \frac{\sum\left(x_i - \bar{x}\right)\varepsilon_i}{\sum\left(x_i - \bar{x}\right)^2}$$

- 两边去期望值：$E\left(\hat{\beta}_1\right) = \beta_1 + \dfrac{\sum\left(x_i - \bar{x}\right)E\left(\varepsilon_i\right)}{\sum\left(x_i - \bar{x}\right)^2} = \beta_1$

表明$\hat{\beta}_1$是β_1的无偏估计量

最小二乘估计量的性质——期望

- 由于：$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$

- 所以：$E(\hat{\beta}_0) = E(\overline{y} - \hat{\beta}_1 \overline{x})$ → $= E(\beta_0 + \beta_1 \overline{x} + \overline{\varepsilon} - \hat{\beta}_1 \overline{x})$ → $= \beta_0 + \beta_1 \overline{x} + E(\overline{\varepsilon}) - \overline{x} E(\hat{\beta}_1)$

  → $= \beta_0 + \beta_1 \overline{x} + 0 - \overline{x} \beta_1$ → $= \beta_0$

表明$\hat{\beta}_0$是β_0的无偏估计量

最小二乘估计量的性质——方差

- $\hat{\beta}_1$ 的方差：$D\left(\hat{\beta}_1\right)=E\left(\hat{\beta}_1-E\left(\hat{\beta}_1\right)\right)^2=E\left(\hat{\beta}_1-\beta_1\right)^2$

- 因为：$\hat{\beta}_1=\beta_1+\dfrac{\sum\left(x_i-\bar{x}\right)\varepsilon_i}{\sum\left(x_i-\bar{x}\right)^2}$ 即：$\hat{\beta}_1-\beta_1=\dfrac{\sum\left(x_i-\bar{x}\right)\varepsilon_i}{\sum\left(x_i-\bar{x}\right)^2}$

- 所以：$\left(\hat{\beta}_1-\beta_1\right)^2=\left(\dfrac{\sum\left(x_i-\bar{x}\right)\varepsilon_i}{\sum\left(x_i-\bar{x}\right)^2}\right)^2=\dfrac{\left(\left(x_1-\bar{x}\right)\varepsilon_1+\cdots+\left(x_n-\bar{x}\right)\varepsilon_n\right)^2}{\left(\sum\left(x_i-\bar{x}\right)^2\right)^2}$

$$=\frac{\left(\sum\left(x_i-\bar{x}\right)^2\varepsilon_i{}^2+\sum_{i\neq j}\left(x_i-\bar{x}\right)\left(x_j-\bar{x}\right)\varepsilon_i\varepsilon_j\right)}{\left(\sum\left(x_i-\bar{x}\right)^2\right)^2}$$

天善智能 TIANSHAN SOFT

最小二乘估计量的性质——方差

- 两边取期望值：$$E\left(\hat{\beta}_1-\beta_1\right)^2=\frac{\left(\sum\left(x_i-\overline{x}\right)^2 E\left(\varepsilon_i^{\ 2}\right)+\sum_{i\neq j}\left(x_i-\overline{x}\right)\left(x_j-\overline{x}\right)E\left(\varepsilon_i\varepsilon_j\right)\right)}{\left(\sum\left(x_i-\overline{x}\right)^2\right)^2}$$
- 由假设可知：$E\left(\varepsilon_i^{\ 2}\right)=\sigma^2 \quad E\left(\varepsilon_i\varepsilon_j\right)=0$
- 所以：$E\left(\hat{\beta}_1-\beta_1\right)^2=\frac{\sum\left(x_i-\overline{x}\right)^2\sigma^2}{\left(\sum\left(x_i-\overline{x}\right)^2\right)^2}=\frac{\sigma^2}{\sum\left(x_i-\overline{x}\right)^2}$ 即：$D\left(\hat{\beta}_1\right)=\frac{\sigma^2}{\sum\left(x_i-\overline{x}\right)^2}$
- 同理：$D\left(\hat{\beta}_0\right)=\frac{\sigma^2\sum x_i^{\ 2}}{n\sum\left(x_i-\overline{x}\right)^2}$

天善智能 TIANSHAN SOFT

小结

- 最小二乘法
- 回归参数的最小二乘估计
- 最小二乘估计量的性质
- 小结