

简单线性回归

包含一个自变量 x 和一个因变量 y

以上两个变量关系可用一条直线来拟合。

若自变量有两个及以上，则称为多元回归分析。

模型:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

因变量
自变量
偏差
(服从正态分布)

方程:

$$E(y) = \beta_0 + \beta_1 x$$

因变量的均值
(期望值)

此方程对应一条直线，称为回归线， β_0 为回归线截距， β_1 为其斜率。

$E(y)$ 为在一个给定 x 值下 y 的均值。

关系:

正向线性关系 负向线性关系 无关系。

是要得出估计的简单线性回归方程。

$$\hat{y} = b_0 + b_1 x$$

估计值
估计的截距
估计的斜率

线性回归分析的流程

Estimation Process (估计流程)

Sample Data

x	y
x_1	y_1
x_2	y_2
\vdots	\vdots
x_n	y_n

\Rightarrow

Estimated Regression Equation.

$$\hat{y} = b_0 + b_1 x$$

\Downarrow

Regression Equation.

$$E(y) = \beta_0 + \beta_1 x.$$

关于偏差 ε

- ① ε 是一个随机变量, 均值为 0
- ② ε 的方差对于所有的自变量 x 是一样的.
- ③ ε 的值是独立的.
- ④ ε 满足正态分布.

如何估计 b_0, b_1 ($\hat{y} = b_0 + b_1 x$)

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

例: Number of TV Ads and Number of Cars sold problem.

TV Ads (x)

1

3

2

$$\bar{x} = 2$$

$$\bar{y} = 20$$

Cars sold (y)

14

24

28

5	$x = 2$	27
2		28
1	$\bar{y} = 20.$	17
3		27

$$b_1 = \frac{(1-2)(14-20) + (3-2)(24-20) + (2-2)(18-20) + (1-2)(17-20) + (3-2)(27-20)}{(1-2)^2 + (3-2)^2 + (2-2)^2 + (1-2)^2 + (3-2)^2}$$

$$= 5$$

$$b_0 = \bar{y} - b_1 \bar{x} = 10.$$

$\therefore \hat{y} = 10 + 5x$ 模型建立好了之后，就可以开始预测了。

$$\text{如: } x_{\text{given}} = 6 \quad \therefore y_{\text{hat}} = 10 + 5 \times 6 = 40.$$