高级计量经济学 2022-3-11

Thursday, February 24, 2022 20:37

The linear regression model.

$$E[y \mid x_1, x_2, \cdots, x_k]$$

$$y = \underbrace{f(x_1, x_2, \cdots, x_k)} + \varepsilon$$
$$= \underbrace{x_1\beta_1 + x_2\beta_2 + \cdots + x_k\beta_k} + \varepsilon$$

依赖 依靠 依靠

y : dependent variable. 从属变量 因

$x_1, \cdots, x_k$: independent variable. 独立变量. 自

f : population regression function

ε : random disturbance.

Assume.

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{ik}\beta_k + \varepsilon_i.$$

- 估计. $\beta_1, \beta_2, \cdots, \beta_k$.
- 解释理论.
- 预测. y

Matrix. matrices.

- scalar variables: $y_i, x_{i1}, \cdots, x_{ik}, \varepsilon_i$
- column vectors:

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$

d a b, y, x

2, a, b, y, x.

$$y=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

- matrices.

$$X=\begin{array}{c|c} & \xrightarrow{k} \\ \downarrow n & \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & & & \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \end{array} \qquad (n\times k)$$

- transposes: 转置. T

$$\beta=\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \qquad \beta'=\beta^T=[\beta_1 \cdots \beta_k].$$

$$X=[x_1\ x_2\ \cdots\ x_k], \quad X'=\begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_k' \end{bmatrix}=\begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ & & \vdots & \\ x_{1k} & x_{2k} & \cdots & x_{nk} \end{bmatrix}$$

- inner/outer product.

$$a\cdot b=a'b=[a_1,\cdots,a_n]\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}=\sum_{i=1}^{n} a_i b_i .$$

$$ab'=\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}[b_1 \cdots b_k]=\begin{bmatrix} a_1b_1 & a_1b_2 & \cdots & a_1b_k \\ a_2b_1 & a_2b_2 & \cdots & a_2b_k \\ & \vdots & & \\ a_nb_1 & a_nb_2 & \cdots & a_nb_k \end{bmatrix}$$

$(n\times k)$

$$X=[x_1, x_2, \cdots, x_k] \qquad x_i'=[x_{1i}\ x_{2i} \cdots x_{ni}]$$

$$\begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{bmatrix} \quad i\text{th column} \quad \begin{bmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ni} \end{bmatrix} \quad x_{\cdot i}$$

$$X = \begin{bmatrix} & & & \\ x_{i1}, & x_{i2}, & \cdots, & x_{ik} \\ & & & \end{bmatrix} \leftarrow i\text{th row}$$

x_i'

x_j'

$i, j, t, s \rightarrow$ rows.

k, $\rightarrow$ column. x_k

Linear regression model

n observation, k variables. x_{ij} : the ith obse. of the jth variable

$$y_1 = \underbrace{x_{11}\beta_1 + x_{12}\beta_2 + \cdots + x_{1k}\beta_k} + \varepsilon_1,$$

$$y_2 = x_{21}\beta_1 + x_{22}\beta_2 + \cdots + x_{2k}\beta_k + \varepsilon_2,$$

$$\vdots$$

$$y_n = x_{n1}\beta_1 + x_{n2}\beta_2 + \cdots + x_{nk}\beta_k + \varepsilon_n.$$

Let $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, $X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ \vdots & & & \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$, $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$

$(n\times 1)$ $\quad \underset{=}{X} \quad (n\times k) \quad (k\times 1) \quad (n\times 1)$

$$\boxed{y = X\beta + \varepsilon}$$

$$[x_{11}\ x_{12} \cdots x_{1k}]\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_2 \end{bmatrix} = \sum_{j=1}^{k} x_{1j}\beta_j$$

$$n\begin{bmatrix} \\ \end{bmatrix} = n\begin{bmatrix} \\ \end{bmatrix}\begin{bmatrix} \\ \end{bmatrix}k + \begin{bmatrix} \\ \end{bmatrix}n$$

k

$n[\;]$

The ith obs. : $y_i = x_{i1}\beta_1 + \cdots + x_{ik}\beta_k + \varepsilon_i$.

$y_i = x_i'\beta + \varepsilon_i$.

the ith row

Probability

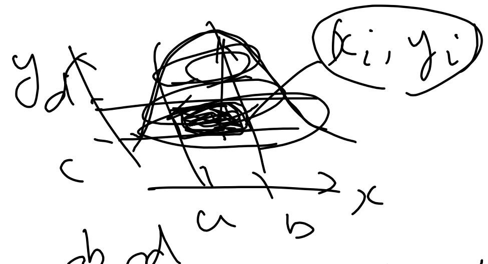
· The expected value : $E[x] = \begin{cases} \sum_i x_i f(x_i) \\ \int x f(x) dx \end{cases}$

· The variance : $Var[x] = E[(x-\mu_x)^2]$ $\quad \mu_x = E[x]$

VAR $\quad = \int (x-\mu_x)^2 f(x) dx$

· Joint density.

$f(x,y)$.

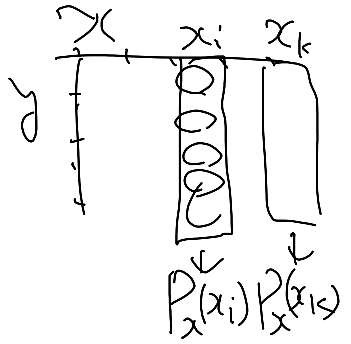


$\text{Prob}(a \le x \le b,\ c \le y \le d) = \int_a^b \int_c^d f(x,y) dy\, dx$

· marginal densities ;

$f_x(x) = \int_y f(x,y) dy$,

$f_y(y) = \int_x f(x,y) dx$.



· Independence. $P(X,Y)=P(X)\cdot P(Y)$

$$f(x,y)=f_x(x)\cdot f_y(y)$$, x and y are independent.

· marginal mean and variance.

$$E[x]=\int_x x\, f_x(x)\,dx$$
$$=\int_x x\left[\int_y f(x,y)\,dy\right]dx$$
$$=\int_x\int_y x\, f(x,y)\,dy\,dx.$$

$$Var[x]=\int_x (x-\mu_x)^2 f_x(x)\,dx$$
$$=\int_x\int_y (x-\mu_x)^2 f(x,y)\,dy\,dx.$$

· Covariance.

$$Cov[x,y]=E[(x-\mu_x)(y-\mu_y)]$$
$$=\int_x\int_y (x-\mu_x)(y-\mu_y)f(x,y)\,dy\,dx$$
$$=E[xy]-\mu_x\mu_y \qquad \underbrace{f(x,y)}_{=\,f_x(x)\,f_y(y)}$$
$$=\sigma_{xy}. \qquad \int_x \; dx \quad \underline{\int_y (y-\mu_y)\, f_y(y)\,dy}.$$

If x and y are independent, $E[y-\mu_y]$ $= E[y]-\mu_y=0.$

$Cov[x,y]=0$.

Correlation. ... $V[x]=\sigma^2$

Correlation $\quad\quad\quad\quad$ $Var[x] = \sigma_x$

$$r[x, y] = \rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$