23 January 2024 08:06 PM

Parameters ← from population } population mean = μ, population variance = σ^2

Statistics ← from a sample } sample mean = $\bar{x}$, Sample variance = s^2

(100) N: $X_1, X_2 \ldots\ldots X_N$ (population) $\Rightarrow \mu$, $\Rightarrow \sigma^2$ } Unknown (we have to estimate)

1, 2, 3, 4, ..., 100

↓

(10) n: $x_1, x_2 \ldots\ldots x_n$ (sample) →

Statistic $\left\{ \begin{array}{l} \bar{x} = \frac{1}{n}\sum x_i \\ s^2 = \frac{1}{n}\sum (x_i - \bar{x})^2 \end{array} \right.$

(1): 1, 2, 3, ... 10 (10) → Statistic

(2): 3, 7, 19, ... 20 (10) → Static

Techniques of estimation
(1) Point estimation
(2) Interval estimation

How to select the best estimate of a parameter?

Properties of a best estimator of parameter

1) Unbiasedness
2) Min variance
3) Consistency
4) Efficiency
5) Sufficiency

from sample, two statistics t_1 and t_2

$\boxed{E(t_1) = \theta}$ $\Rightarrow E(t) = \theta$ or $E(t) - \theta = 0$

$E(\bar{x}) = \mu$ ← parameter of a population

statistic from sample

Hypothesis Testing ⇒ (Testing of the values of parameters)

$H_0: \mu = 0$? which is

a) $H_0 : \mu = 0$
$H_1 : \mu \neq 0$ } Which is true }

b) $H_0 : \sigma^2 = 1$
$H_1 : \sigma^2 \neq 1$ } Which is true }

one independent variable (x_i).

(Simple Linear Regression Model)

Topic : OLS (Ordinary Least Square Method)

$$Y_i = \alpha + \beta \overset{\text{indep}}{x_i} + U_i$$

Y_i: dependent

α: constant (intercept)

β: regression coefficient (slope of regression line)

U_i: (random disturbance) or stochastic error

Assumptions of U_i (random error / disturbance) in a CLRM model.

$\left[V(x) = E(x - E(x))^2 \right.$

① $E(u_i) = 0$ for $i = 1, 2, \dots n$

② $V(u_i) = E(u_i - E(u))^2$
$= E(U_i^2) = \sigma_u^2$ (constant)

This is called homoscedasticity.

Note: violation of ② point i.e. $V(u) = \sigma_{u_i}^2$ = non constant

it is problem of heteroscedasticity
⇒ OLS not applicable.

③ $\text{Cov}(U_i, U_j) = E(U_i U_j) = 0$

... correlated.

③ $Cov(u_i, u_j) = E(u_i u_j) = 0$

Two random errors are uncorrelated.

Note: If u_i and u_j are correlated, then violation of pt ③ ⇒ called problem of autocorrelation

④ Independent variable X_i is non-stochastic.

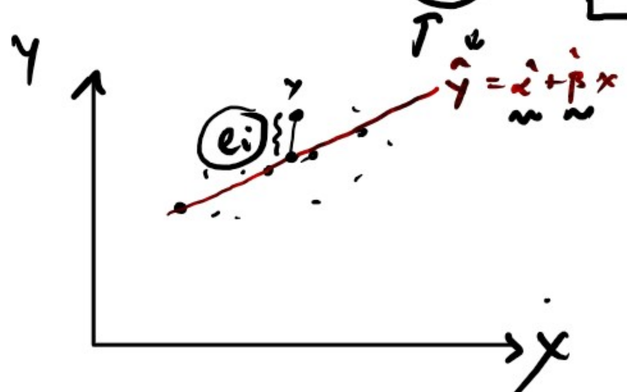
⑤ a) $cov(x, u) = 0$

b) $cov(x_i, x_j) = 0$ ⇒ Note: if $cov(x_i, x_j) \neq 0$ then problem of multicollinearity.

The estimated regression equation is written as,

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$$

→ this gives a straight-line equation.



What is estimation error?

$$\boxed{e_i = y_i - \hat{Y}_i}$$

objective is to estimate the value of $\hat{\alpha}$ and $\hat{\beta}$ such that the error of estimation is minimised.

$$e_i = Y_i - \hat{Y}_i$$

$$e_i^2 = (Y_i - \hat{Y}_i)^2$$

$$e_i^2 = (y_i - \hat{y}_i)^2$$

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

or, $$\sum e_i^2 = \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta} x_i)^2$$ ✓

f.o.c for minimisation requires,

$$\frac{\partial \sum e_i^2}{\partial \hat{\alpha}} = 0$$

or, $$2 \sum (y_i - \hat{\alpha} - \hat{\beta} x_i)(-1) = 0$$

or, $$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \hat{\alpha} - \hat{\beta} \sum_{i=1}^{n} x_i = 0$$

or, $$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{\alpha} + \hat{\beta} \sum_{i=1}^{n} x_i$$

or, $$\sum_{i=1}^{n} y_i = n\hat{\alpha} + \hat{\beta} \sum_{i=1}^{n} x_i \quad \text{——} \ (1)$$

Again, $$\frac{\partial \sum e_i^2}{\partial \hat{\beta}} = 0$$

or, $$2 \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta} x_i)(-x_i) = 0$$

or, $$\sum x_i y_i - \hat{\alpha} \sum x_i - \hat{\beta} \sum x_i^2 = 0$$

or, $$\sum x_i y_i = \hat{\alpha} \sum x_i + \hat{\beta} \sum x_i^2 \quad \text{——} \ (2)$$

or, $\sum x_i y_i = \hat{\alpha} \sum x_i + \hat{\beta} \sum x_i^2$ — (2)

eq (1) and eq (2) are called normal equations.

$$\sum_{i=1}^{n} y_i = n\hat{\alpha} + \hat{\beta} \sum x_i$$

$$\sum x_i y_i = \hat{\alpha} \sum x_i + \hat{\beta} \sum x_i^2$$

Let us write the system of equation in matrix form

$$\begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}$$

$$C = \underbrace{\qquad}_{A} \quad B$$

Let $|A| = \begin{vmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{vmatrix} = n \sum x_i^2 - \left(\sum x_i\right)^2$

Now $\hat{\beta} = \dfrac{\begin{vmatrix} n & \sum y_i \\ \sum x_i & \sum x_i y_i \end{vmatrix}}{n \sum x_i^2 - (\sum x_i)^2} = \dfrac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$ (dividing by n^2)

$$= \frac{\frac{1}{n} \sum x_i y_i - \frac{1}{n^2} \sum x \sum y}{\frac{1}{n} \sum x_i^2 - \left(\frac{\sum x_i}{n}\right)^2}$$

$$\hat{\beta} = \frac{\mathrm{cov}(x, y)}{\mathrm{var}(x)} = \frac{\frac{1}{n}\sum (x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n}\sum (x_i - \bar{x})^2}$$

Let $X_i - \bar{X} = x_i$, $Y_i - \bar{Y} = y_i$ $\Big\}$ then $\hat{\beta}$ can also be are $X_i - \bar{X} = x_i$

$Y_i - \bar{Y} = y_i$ } then $\hat{\beta}$ can also be written as

$$\boxed{\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}}$$

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

or, $\boxed{\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}}$