26 August 2023 07:37 PM

# Regression Estimator.

$x \quad y \quad (x-\bar{x}) \quad y-\bar{y} \quad (\ )(\ )$

$\Sigma(\ ) \qquad \Sigma(\ )$

$$\boxed{\hat{y} = a + bx}$$

$$b_{yx} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2} = \text{ans}.$$

$$\boxed{a = \bar{y} - b_{yx}\bar{x}}$$

Mean

$x = \mu_x \qquad a = \bar{y} - b\bar{x} \Rightarrow \hat{\mu}_L = (\bar{y} - b\bar{x}) + b\mu_x$

$$\boxed{\hat{\mu}_L = a + b\mu_x}$$

$y \quad x \quad \mu_L \rightarrow$ by.

$$\hat{Var}(\hat{\mu}_L) = \frac{N-n}{N\times n} \cdot \frac{\Sigma(y_i - a - bx_i)^2}{n-2}$$

$$= \frac{N-n}{N\times n} \cdot MSE$$

at $(1-\alpha)100\%$ CI :

$$\hat{\mu}_L \pm t_{n-2,\alpha/2}\sqrt{\hat{Var}(\hat{\mu}_L)}$$

$$\hat{\tau}_L = N\cdot\hat{\mu}_L = N\bar{y} + b(\tau_x - N\bar{x})$$

$$\hat{Var}(\hat{\tau}_L) = N^2\,\hat{Var}(\hat{\mu}_L)$$

$$= \frac{N\times(N-n)}{n} \cdot MSE$$

$(1-\alpha)100\%$ CI for $\tau$ is $\quad \hat{\tau}_L \pm t_{n-2,\alpha/2}\sqrt{\hat{V}(\hat{\tau}_L)}$

$n=10$ ~~Score~~ Achivement ~~est~~ Score.

$x \qquad y$

$n=10$

| ~~score~~ $X$ | ~~at~~ score $Y$ |
|---|---|
| 39 | 65 |
| 43 | 78 |
| 21 | 52 |
| ⋮ | ⋮ |
| 52 | 75 |

$$\hat{\mu}_L = \bar{y} + b(\mu_x - \bar{x})$$
$$= a + b\mu_x$$

# Stratified Sampling

$L \Rightarrow$ no. of strata.

Population total estimate $\hat{\tau}_{st} = \sum_{h=1}^{L} \hat{\tau}_h$

$$\text{Variance}(\hat{\tau}_{st}) = \sum_{h=1}^{L} \text{Var}(\hat{\tau}_h)$$

For stratified random sampling ie, take a random sample within each stratum

$$\hat{\tau}_h = N_h \bar{y}_h$$

$$\hat{\text{Var}}(\hat{\tau}_{st}) = \sum_{h=1}^{L} N_h \cdot (N_h - n_h) \cdot \frac{s_h^2}{n_h}$$

$$s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2$$

$$\hat{\mu}_{st} = \frac{\hat{\tau}_{st}}{N} \quad \text{and} \quad \hat{\text{Var}}(\hat{\mu}_{st}) = \frac{1}{N^2} \hat{\text{Var}}(\hat{\tau}_{st})$$

For stratified random sampling

$$\bar{y}_{st} = \frac{1}{N} \sum_{h=1}^{L} N_h \bar{y}_h$$

$$\bar{y}_{st} = \frac{1}{N} \sum_{h=1}^{L} N_h y_h$$

$$\hat{Var}(\bar{y}_{st}) = \sum_{h=1}^{L} \left(\frac{N_h}{N}\right)^2 \left(\frac{N_h - n_h}{N_h}\right) \frac{s_h^2}{n_h}$$

| | | |
|---|---|---|
| Town A | 35, 43, ... 34 (20) | $N_1 = 155$ |
| Town B | 27 ... 9 ... (8) | $N_2 = 62$ |
| Rural C | 8, 14 ... (12) | $N_3 = 93$ |

$E(\bar{x})$

$\frac{s}{\sqrt{n}}$

| variables | $n_h$ | Mean ($\bar{y}_h$) | std dev | SE (mean) |
|---|---|---|---|---|
| A | 20 | 33.9 | 5.95 | 1.33 |
| B | 8 | 25.12 | 15.25 | 5.39 |
| C | 12 | 19 | 9.36 | 2.70 |

$L = 3$ $\quad N_1 = 155 \quad N_2 = 62 \quad N_3 = 93$

$$N = 155 + 62 + 93 = 310$$

$$\bar{y}_{st} = \frac{1}{N}(N_1 \bar{y}_1 + N_2 \bar{y}_2 + N_3 \bar{y}_3)$$

$$= \frac{1}{310}(155 \times 33.9 + 62 \times 25.12 + 93 \times 19)$$

$$\bar{y}_{st} = 27.7$$

$$Var(\bar{y}_{st}) = \sum_{h=1}^{3} \left(\frac{N_h}{N}\right)^2 \left(\frac{N_h - n_h}{N_h}\right) \frac{s_h^2}{n_h}$$

$$= \frac{1}{(310)^2} \left\{ \left( (155)^2 \frac{(155-20)}{155} \cdot \frac{(5.95)^2}{20} \right) \right\}$$

$$= \frac{1}{(310)^2}\left\{\left((155) \frac{(155-20)}{155} \cdot \frac{(5.95)}{20}\right) + \left( \quad \right) + \left( \quad \right)\right\}$$

$$= 1.97$$

$$\hat{\tau}_{st} = N \cdot \bar{y}_{st} = 310 \times 27.7 = 8587$$

$$\hat{Var}(\hat{\tau}_{st}) = N^2 \hat{V}(\bar{y}_{st})$$

$$= (310)^2 \times 1.97 = 189317.$$

CI : $\hat{\tau}_{st} \pm t_{df, \alpha/2} \sqrt{\hat{V}(\hat{\tau}_{st})}$

$P(A) = 3/10$

samp. A $\rightarrow$ 1 2 3 5 ... 6 10

B $\rightarrow$ 1 2, 11, 12, ... 6 20

$P(B) = \cancel{2}\, 2/20$

$P(A \cap B) =$ 1, 2, 6 = ?

$P(A/B) = \dfrac{P(A \cap B)}{P(B)}$