WAGE1

Thursday, October 15, 2020 11:15 AM

Regression w/ binary variable

$y = \beta_0 + \beta_1 x$

x is binary

$E(y|x) = \beta_0 + \beta_1 x + \varepsilon$
$E(y|x=0) = \beta_0$
$E(y|x=1) = \beta_0 + \beta_1$

$\beta_1 = E(y|x=1) - \beta_0$
$= E(y|x=1) - E(y|x=0)$

β_1 = difference in average value of Y over subpopulations w/ $x=1$ and $x=0$

Relation between score in QMB and gender

$score = \beta_0 + \beta_1(gender)$ male = 1, female = 0

$E(score|gender = female) = \beta_0$

given

$E(score|gender = male) = \beta_0 + \beta_1$

$\beta_1 = E(score|male) - \beta_0$
$= E(score|male) - E(score|female)$

β_1 = diff in male and female scores

Population model

$Y = \beta_0 + \beta_1 X + \varepsilon$

slope

intercept

X and Y are random

Population model is not known. Use sample x and y data to estimate β_0 and β_1

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \varepsilon_i$

ε_i = error term that contains unobservable factor other than x that affect Y

ex:

$y = \{5, 10, 15, 2, 1\}$ $x = \{2, 2, 2, 2, 2\}$ ← bad

2: zero conditional mean
$E(\varepsilon|X) = 0$

3: X and Y are randomly distributed

Estimate $\hat{\beta}_0$ and $\hat{\beta}_1$ from the sampled data using OLS (ordinary least squares) estimation

$\hat{\beta}_1$ and $\hat{\beta}_0$ are unbiased estimators of β_0 and β_1

If the sample is typical and representative of population, then $\hat{\beta}_0$ and $\hat{\beta}_1$ should be near (or) close to population values β_0 and β_1

Homoskedasticity → constant variance

$Var(\varepsilon|x) = constant$
$= \sigma^2$

variance of error term (or) unobservable factors conditional on x is constant

$Var(y|x) = constant$

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X$

Residual = observed − fitted
$y - \hat{y}$