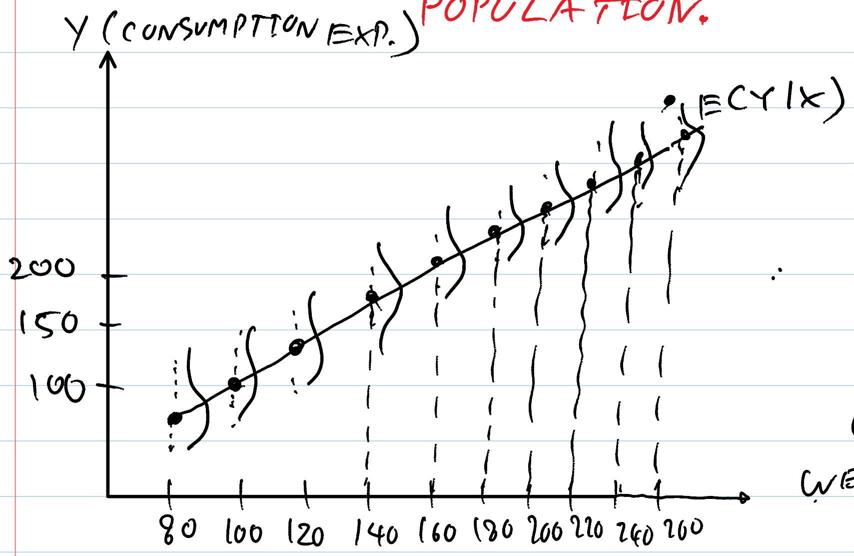
BASIC IDEAS ABOUT REGRESSION ANALYSIS

STORY: • WEEKLY FAMILY INCOME & CONSUMPTION EXPENDITURE.
• WE TAKE 60 FAMILIES AS "POPULATION."



E(Y|X) = CONDITIONAL MEANS OF Y

E(Y) = UNCONDITIONAL MEAN OF Y

EX: E(Y) = 121.20

$E(Y|X=80) = 65$

$E(Y|X=260) = 173$

$$\underbrace{E(Y|X_i)}_{\text{DEPENDENT VARIABLE}} = f(X_i)$$

X_i: EXPLANATORY VARIABLE

⇒ IT IS CALLED POPULATION REGRESSION FUNCTION. (PRF)

LET'S ASSUME THAT THE PRF OF $E(Y|X_i)$ IS A LINEAR FUNCTION OF X_i.

$$\boxed{E(Y|X_i) = a + bX_i}$$

" REGRESSION EQUATION OR REGRESSION MODEL

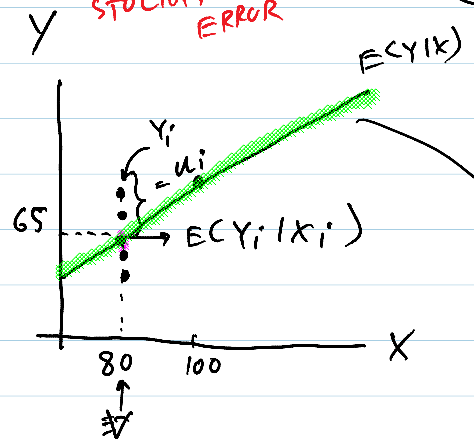
SHOWS THAT IF INCOME ↑, ON AVERAGE, CONSUMPTION EXPENDITURE ↑ [NOTE THAT THIS MIGHT NOT TRUE FOR ALL POPULATION,
EX: SOME FAMILIES HAVE HIGHER

[INCOME BUT LOWER CONSUMPTION]

SO THIS EQUATION SHOWS AVERAGE RELATIONSHIP OF CONSUMPTION & WEEKLY INCOME.

∴ THERE WILL BE ERRORS.

STOCHASTIC DISTURBANCE OR STOCHASTIC ERROR ← $u_i = \underbrace{Y_i - E(Y_i \mid X_i)}$



DEVIATION OF Y_i FROM ITS CONDITIONAL MEAN.

$$E(Y_i \mid X_i) = f(X_i)$$

MEANING OF "LINEAR MODEL"

LINEARITY

- IN VARIABLES
  - EX: $a + bx^2$ → NON-LINEAR IN VARIABLE
  - $Y = \alpha + \beta X_i + \gamma X_i^2 + u_i$
- IN PARAMETERS
  - EX $Y = \alpha + \beta^2 X_i$ → NON LINEAR IN PARAMETER

EX: $Y = \alpha_0 + \alpha_1 \ln X_i + u_i$ → LINEAR IN VARIABLES, LINEAR IN PARAMETERS

YOU COULD EVEN DEFINE $\ln X_i = Z_i$, SO

$$Y = \alpha_0 + \alpha_1 Z_i + u_i$$

IN THIS CLASS (EE325), WHEN WE TALK ABOUT LINEAR MODEL, WE MEAN "LINEAR IN PARAMETER"

EX: COBB-DOUGLAS PRODUCTION FUNCTION:

$Y_i = A K_i^{\alpha} L_i^{\beta} e^{u_i}$ IS A NON-LINEAR

MODEL, BUT IT IS INTRINSICALLY LINEAR.

$$\ln Y_i = \ln A + \alpha \ln K_i + \beta \ln L_i + u_i$$

$$Y' = A' + \alpha K' + \beta L' + u_i \rightarrow \text{LINEARIZED}$$

<u>SUMMARY</u>

LINEAR REGRESSION MODEL REFERS TO A MODEL THAT IS LINEAR IN PARAMETER.

∴ IT MAY OR MAY NOT BE LINEAR IN VARIABLES.

EX:

$$Y = \beta_1 + \beta_2 X + \beta_3 X^2 \quad \text{(QUADRATIC)} \rightarrow \text{LINEAR.}$$

$$Y = \boxed{e^{\beta_1 + \beta_2 X}} \quad \text{(EXPONENTIAL)} \rightarrow \text{LINEAR}$$

$$Y = \beta_1 + \beta_2 X + \beta_3 X^2 + \beta_4 X^3 \quad \text{(CUBIC)}$$

↓ LINEAR

$$\ln Y = \beta_1 + \beta_2 X$$

$$Y' = \beta_1 + \beta_2 X$$

LINEAR MODEL

CONSIDER $u_i = Y_i - E(Y_i | X_i)$.

OR $\underline{\underline{Y_i}} = \underbrace{E(Y_i | X_i)}_{\text{SYSTEMATIC PART OR DETERMINISTIC PART}} + \underbrace{u_i}_{\text{NON-SYSTEMATIC PART OR STOCHASTIC PART OR RANDOM PART}}$.

SO IF $E(Y_i | X_i)$ IS ASSUMED TO BE LINEAR IN X,

$$\boxed{Y_i = a + bX_i + u_i}$$

$$\boxed{Y_i = \underline{a + bX_i} + \underline{u_i}}$$

CONSUMPTION EXP. DEPENDS ON (1) WEEKLY INCOME IN A LINEAR MANNER, AND (2) THE DISTURBANCE TERM.

SUPPOSE $X = 200$

$$\begin{aligned}
Y_1 &= 120 = a + b(200) + u_1 \\
Y_2 &= 136 = a + b(200) + u_2 \\
Y_3 &= 140 = a + b(200) + u_3 \\
Y_4 &= 144 = a + b(200) + u_4 \\
Y_5 &= 145 = \underbrace{a + b(200)}_{\text{DETERMINISTIC COMPONENT}} + \underbrace{u_5}_{\text{STOCHASTIC COMPONENT}}.
\end{aligned}$$

FROM: $$Y_i = E(Y|X_i) + u_i$$

TAKE CONDITIONAL EXPECTED VALUE TO THE ABOVE EQUATION:

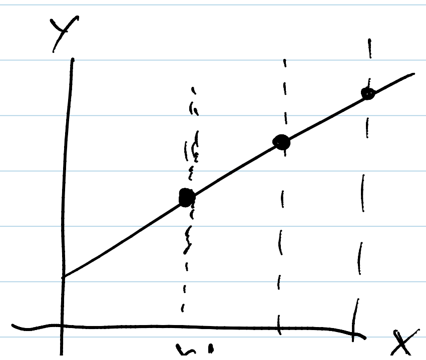
$$E(Y_i|X_i) = E\left[E(Y|X_i)\right] + E(u_i|X_i)$$

$$= E(Y|X_i) + E(u_i|X_i)$$

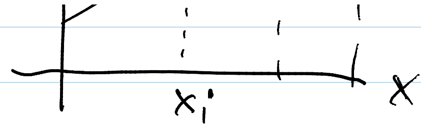
SINCE $E(Y_i|X_i) = E(Y|X_i)$,

SO $\boxed{E(u_i|X_i) = 0.}$

IMPLICATION?



● REFERS $E(Y|X_i)$



WHEN WE ASSUME THAT THE REGRESSION LINE PASSES THROUGH CONDITIONAL MEANS OF Y, IT MUST BE THE CASE THAT

THE CONDITIONAL MEAN VALUE OF u_i (CONDITIONAL UPON THE GIVEN X'S) ARE ZERO.