METHODOLOGY OF ECONOMETRICS (8 STEPS)

1. STATEMENT OF THEORY OR HYPOTHESIS
2. SPECIFICATION OF MATHEMATICAL MODEL OF THE THEORY
3. SPECIFICATION OF THE ECONOMETRIC MODEL
4. DATA COLLECTION
5. ESTIMATION OF THE PARAMETERS OF THE ECONOMETRIC MODEL
6. HYPOTHESES TESTING
7. FORCASTING OR PREDICTION
8. USING THE MODEL FOR POLICY PURPOSES

AS INCOME RISES, PEOPLE TEND TO CONSUME MORE, VICE VERSA.

$Y_i = \beta_1 + \beta_2 X_i \quad , \quad 0 < \beta_2 < 1$

WHERE Y_i = CONSUMPTION EXP.

$\Rightarrow Y_i = \beta_1 + \beta_2 X_i + u_i$; X_i = DISPOSABLE INCOME ; $u_i \rightarrow$ ERROR TERM OR DISTURBANCE TERM.

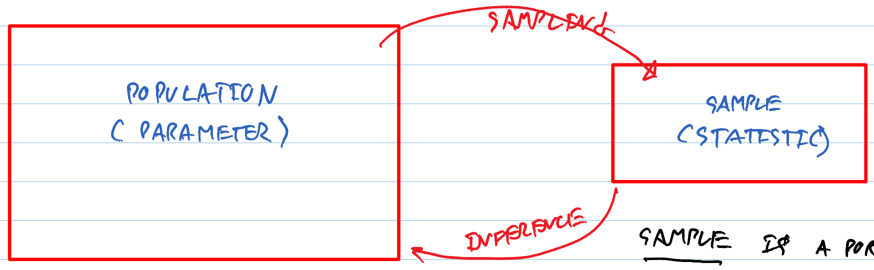
N = 30 OBSERVATIONS, LET'S SAY $\hat{\beta}_1$, $\hat{\beta}_2 \Rightarrow MPC = \frac{\Delta Y}{\Delta X}$

$\hat{Y}_i = -2000 + 0.7 X_i$

??? MPC = 0.7
TO TEST IF MPC IS STATISTICALLY SIGNIFICANT.

SUPPOSE, X_i = 5000 BAHT / WK
THEN $\hat{Y}_i = -2000 + 0.7 \cdot 5000$ = ____

~~REVIEW OF INFERENTIAL STATISTICS~~



POPULATION IS A COLLECTION OF ALL POSSIBLE INDIVIDUALS, OBJECTS, OR MEASUREMENTS OF INTEREST.

SAMPLE IS A PORTION OR PART OF THE POPULATION WE ARE INTERESTED IN

~~PARAMETER~~ IS A NUMERICAL MEASUREMENT DESCRIBING " CHARACTERISTIC OF A POPULATION "

~~STATISTIC~~ IS A NUMERICAL MEASUREMENT DESCRIBING CHARACTERISTIC OF A SAMPLE "

$\hat{GPA}_i = \beta_1 + \beta_2 X_i$

$\hat{GPA}_i = 2.25 + 0.15 X_i \qquad N = 150$ OBS.

0.35

$\hat{\beta}_2$ = ESTIMATED VALUE OF β_2

$H_0 : \beta_2 = 0$

$H_1 : \beta_2 \neq 0$

0.15

0

$H_1: \beta_2 \neq 0$

$$\hat{\beta}_2 \longrightarrow t_{\hat{\beta}_2} = \frac{\overset{0.5}{\hat{\beta}_2} - \overset{0}{\beta_2}}{s.e.(\hat{\beta}_2)} = ?$$

17.01.2013

REGRESSION: AN ATTEMPT TO EXPLAIN THE VARIATION IN A DEPENDENT VARIABLE USING THE VARIATION IN INDEPENDENT VARIABLE(S).

REGRESSION IS THUS AN EXPLANATION OF CAUSATION.

IF THE INDEPENDENT VARIABLE(S) SUFFICIENTLY EXPLAIN THE VARIATION IN THE DEPENDENT VARIABLE, THE MODEL CAN BE USED FOR PREDICTION.

$$\boxed{Y_i = \beta_1 + \beta_2 X_i + u_i}$$

2-VARIABLE REGRESSION MODEL

✓ DEPENDENT VARIABLE	⟷	INDEPENDENT VARIABLE
✓ EXPLAINED VARIABLE	⟷	EXPLANATORY VARIABLE
✓ REGRESSAND	⟷	REGRESSOR
OUTCOME	⟷	STIMULUS
PREDICTED VARIABLE	⟷	PREDICTOR
ETC.		ETC.

REMARKS: (1) A STATISTICAL RELATIONSHIP IN ITSELF CANNOT LOGICALLY IMPLY ~~CAUSATION~~ N. (2 WAYS OF RELATIONSHIP)

Ex: CROP YIELD ⟵ RAIN FALL

CROP YIELD $\not\longrightarrow$ RAIN FALL

(2) REGRESSION Vs. CORRELATION

$Y = f(X)$

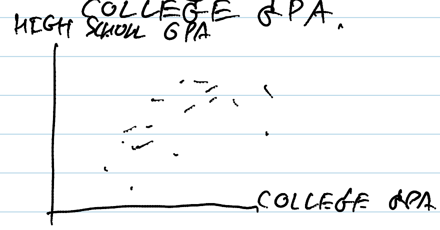
X CAUSES Y

X: EXPLANATORY VARIABLE

Y: DEPENDENT VARIABLE

CORRELATION: MEASURES THE DEGREE OF ASSOCIATION BETWEEN TWO VARIABLES

Ex: HIGH SCHOOL GPA & COLLEGE GPA.



TWO-VARIABLE REGRESSION ANALYSIS

TWO-VARIABLE REGRESSION ANALYSIS

READ: GUJARATI, CH. 2

CONSIDER

$$\boxed{Y_i = \beta_1 + \beta_2 X_i + u_i}$$

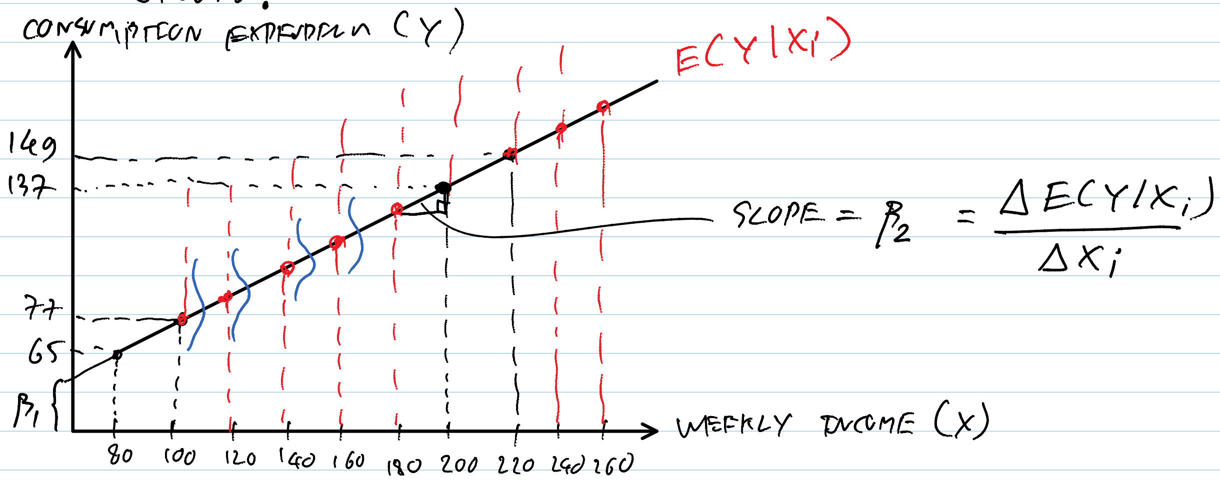
EXPLANATORY VARIABLE, INDEPENDENT VARIABLE

DEPENDENT VARIABLE

OBJECTIVE: HOW Y VARIES WITH THE CHANGE IN X

LET'S TAKE A HYPOTHETICAL EXAMPLE...

- A TOTAL POPULATION OF 60 FAMILIES IN A VILLAGE
- THEIR WEEKLY INCOME (X)
  THEIR WEEKLY CONSUMPTION EXPENDITURE (Y)
- THE 60 FAMILIES ARE DIVIDED INTO 10 INCOME GROUPS.



$E(Y|X_i)$ = CONDITIONAL MEAN OF Y
(AVERAGE VALUE OF Y GIVEN X)
(MEAN VALUE OF Y GIVEN X)
(CENTER OF DISTRIBUTION OF Y GIVEN X)

$E(Y)$ = UNCONDITIONAL MEAN OF Y

$E(Y) = \frac{7272}{60} = \$121.20$

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

$E(Y|X=220) = \$149$

FROM THE PICTURE ABOVE, WE CAN WRITE DOWN "A FUNCTION"

$$\boxed{E(Y|X_i) = f(X_i)}$$ ⇒ CALLED "POPULATION REGRESSION FUNCTION"

READ: AVERAGE VALUE OF Y GIVEN WEEKLY INCOME LEVEL (X) DEPENDS ON X_i.

LET'S ASSUME THAT $E(Y|X_i)$ IS "LINEARLY RELATED" WITH X_i:

$$\boxed{E(Y|X_i) = \beta_1 + \beta_2 X_i}$$. ⇒ CALLED "~~POPULATION~~" REGRESSION

WHEN X_i.

$$\boxed{E(Y|X_i) = \beta_1 + \beta_2 X_i}$$. ⇒ CALLED "~~POPULATION~~" REGRESSION

WHERE β_1 = INTERCEPT TERM
β_2 = SLOPE COEFFICIENT

AS INCOME (X) RISES, ON AVERAGE, CONSUMPTION EXPENDITURE ALSO RISES, OTHER THING BEING EQUAL,
(CETERIS PARIBUS)

* THIS EQUATION SHOWS "AVERAGE RELATIONSHIP" OF CONSUMPTION EXPENDITURE AND FAMILY INCOME.

WHEN WE SAY THIS, IT FURTHER IMPLIES THAT

THERE WILL BE ERRORS !!!

OBSERVE THAT IT IS NOT ALWAYS THE CASE THAT

FOR SOME FAMILIES, THEIR EXPENDITURE ARE LOWER

THAN THOSE FAMILIES W/ LOWER INCOME LEVEL

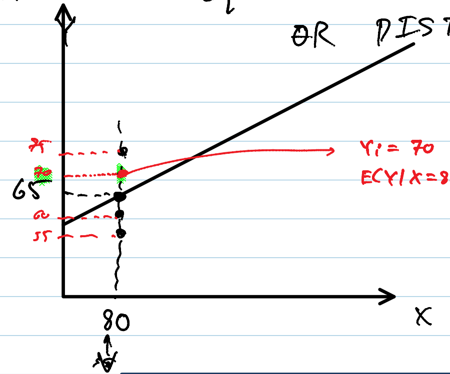
$X_i = 100$, $Y_i = 65$ → $X_i = 80$, $Y_{1i} = 70$, $Y_{2i} = 75$

A FAMILY

TAKE A LOOK AT THIS :

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

WHERE u_i = STOCHASTIC ERROR
OR DISTURBANCE TERM



$Y_i = 70$
$E(Y|X=80) = 65$ } $u_i = 70 - 65$
$= +5$

$$\boxed{Y_i = \underbrace{E(Y|X_i)}_{\text{systematic part}} + \underbrace{u_i}_{\text{random or stochastic error term}}} \qquad (1)$$

IF $E(Y|X_i)$ ABOVE IS LINEARLY RELATED IN X_i,

WE CAN WRITE EQ(1) AS:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad (2)$$

EX: LOOK AT $X = 80$.

$$\left.\begin{aligned}
Y_1 &= 55 = \beta_1 + \beta_2(80) + u_1 \\
Y_2 &= 60 = \beta_1 + \beta_2(80) + u_2 \\
Y_3 &= 65 = \beta_1 + \beta_2(80) + u_3 \\
Y_4 &= 70 = \beta_1 + \beta_2(80) + u_4 \\
Y_5 &= 75 = \beta_1 + \beta_2(80) + u_5 \\
& \beta_1 + \beta_2(80) + u_5
\end{aligned}\right\} \qquad (3)$$

DETERMINISTIC OR SYSTEMATIC

STOCHASTIC NON SYSTEMATIC

EQ(3) TELLS US THAT CONSUMPTION EXP. OF A FAMILY IS LINEARLY RELATED TO ~~ITS~~ INCOME PLUS THE DISTURBANCE TERM.

FROM EQ(1): $Y_i = E(Y|X_i) + u_i$, IF WE TAKE THE EXPECTED VALUE OF EQ(1) ON BOTH SIDES, WE GET

$$E(Y|X_i) = E[E(Y|X_i)] + E[u_i|X_i] \qquad (4)$$

EXPECTED VALUE OF A CONSTANT NUMBER IS A CONSTANT ITSELF

SO, $E(Y|X_i) = E[E(Y|X_i)]$!

AS A RESULT, EQ(4) IMPLIES THAT

$$E(u_i|X_i) = 0$$

THAT IS, THE CONDITIONAL MEAN VALUES OF u_i (CONDITIONAL UPON THE GIVEN X_i) ARE ZERO !

MESSAGE IS THIS: THE ASSUMPTION THAT THE PRF LINE PASSES THROUGH THE CONDITIONAL MEAN OF Y ~~IMPLIES THAT~~ $E(u_i|X_i) = 0$.

$$E(Y|X_i) = \beta_1 + \beta_2 X_i \quad \text{vs.} \quad Y_i = \beta_1 + \beta_2 X_i + u_i$$

$E(Y \mid X_i) = \beta_1 + \beta_2 X_i$ vs. $Y_i = \beta_1 + \beta_2 X_i + u_i$

(A) THIS IS PRF LINE (B)

EQ (A) AND EQ (B) ARE EQUIVALENT IF $E(u_i / X_i) = 0$

- $Y_i = \beta_1 + \beta_2 X_i + u_i$ IS THE STOCHASTIC SPECIFICATION OF $E(Y \mid X_i) = \beta_1 + \beta_2 X_i$.
- EQ (B) SAYS THAT THERE ARE OTHER VARIABLES BESIDE INCOME THAT AFFECT CONSUMPTION EXPENDITURE AND THAT AN INDIVIDUAL FAMILY'S EXPENDITURE CANNOT BE FULLY EXPLAINED ONLY BY THE VARIABLE(S) INCLUDED IN THE REGRESSION MODEL.

THE SIGNIFICANCE OF THE STOCHASTIC DISTURBANCE TERM

(ERROR TERM)

- AS NOTED, THE DISTURBANCE TERM u_i IS "A PROXY" FOR ALL THOSE VARIABLES THAT ARE OMITTED FROM THE MODEL BUT COLLECTIVELY AFFECT Y.
- A QUESTION ARISES: WHY DON'T WE INCLUDE THESE VARIABLES INTO THE MODEL EXPLICITLY?
- ANSWER: THE REASONS ARE MANY.

1. UNAVAILABILITY OF DATA.
2. CORE VARIABLES Vs. PERIPHERAL VARIABLES
   - "MAKE THE MODEL AS SIMPLE AS POSSIBLE"
   - THE COMBINED EFFECT OF PERIPHERAL VARIABLES CAN BE INCORPORATED IN u_i.
3. INTRINSIC RANDOMNESS IN HUMAN BEHAVIOR: EVEN YOU TRY VERY HARD BY PUTTING ALL VARIABLE YOU THINK THEY ARE IMPORTANT, STILL, YOU CANNOT FULLY EXPLAIN Y. SO, u_i REFLECTS THIS INTRINSIC RANDOMNESS.
4. POOR PROXY VARIABLES.

EX: MILTON FREEDMAN'S CONSUMPTION FUNCTION:

$$Y^P = f(X^P)$$

PERMANENT INCOME

$$Y^P = f(X^P)$$

PERMANENT CONSUMPTION OVER TIME

PERMANENT INCOME OVER TIME

SINCE Y^P AND X^P ARE NOT DIRECTLY OBSERVED. IN PRACTICE, ~~YOU HAVE~~ TO USE PROXY VARIABLES.

EX: CURRENT CONSUMPTION TO BE A PROXY OF Y^P

CURRENT INCOME TO BE A PROXY OF X^P.

SINCE CURRENT CONSUMPTION, OBSERVABLE, OF COURSE MAY NOT EQUAL TO Y^P

AND CURRENT INCOME MAY NOT EQUAL TO X^P.

THEREFORE, WE HAVE GOT A PROBLEM OF ERRORS OF MEASUREMENT.

u MAY, IN THIS CASE, REPRESENT THE ERRORS OF MEASUREMENT.

(5) WRONG FUNCTIONAL FORM.

ACTUALLY WE MAY NOT KNOW THE CORRECT FUNCTIONAL FORM OF THE RELATIONSHIP. — (1)

EX: $Y_i = \beta_1 + \beta_2 X_i + u_i$

OR $Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + u_i$ — (2)

(6) PRINCIPLE OF PARSIMONY: KEEP YOUR MODEL AS SIMPLE AS POSSIBLE.

REMARK: WE SHOULD NOT EXCLUDE RELEVANT AND IMPORTANT VARIABLES JUST FOR THE SAKE OF KEEPING THE REGRESSION MODEL SIMPLE!

(7) VAGUENESS OF THEORY

EVEN THOUGH MR. KEYNES POSTULATES THAT INCOME PLAYS AN IMPORTANT ROLE IN EXPLAINING CONSUMPTION EXPENDITURE, BUT WE MIGHT BE UNSURE ABOUT THE OTHER VARIABLES AFFECTING Y.

AS A RESULT, u_i MAY BE USED AS A SUBSTITUTE FOR ALL THE EXCLUDED OR OMITTED VARIABLES FROM THE MODEL.