TRUE POPOLUTATION MODEL (UNKNOWN / NOT OBSERVABLE)

$$Y_i = a + bX_i + \varepsilon_i \quad \text{REGRESSION MODEL}$$

Y_i, X_i : Random Variables - OBSERVED (FROM A SAMPLE)

ε_i : Random Variable - NOT OBSERVED (ERRORTERM)

a, b : UNKNOWN PARAMETERS TO BE ESTIMATED FROM A SAMPLE

GOAL: $\hat{a}, \hat{b}$: ESTIMATES of a and b FROM A SAMPLE $i = 1, .., N$

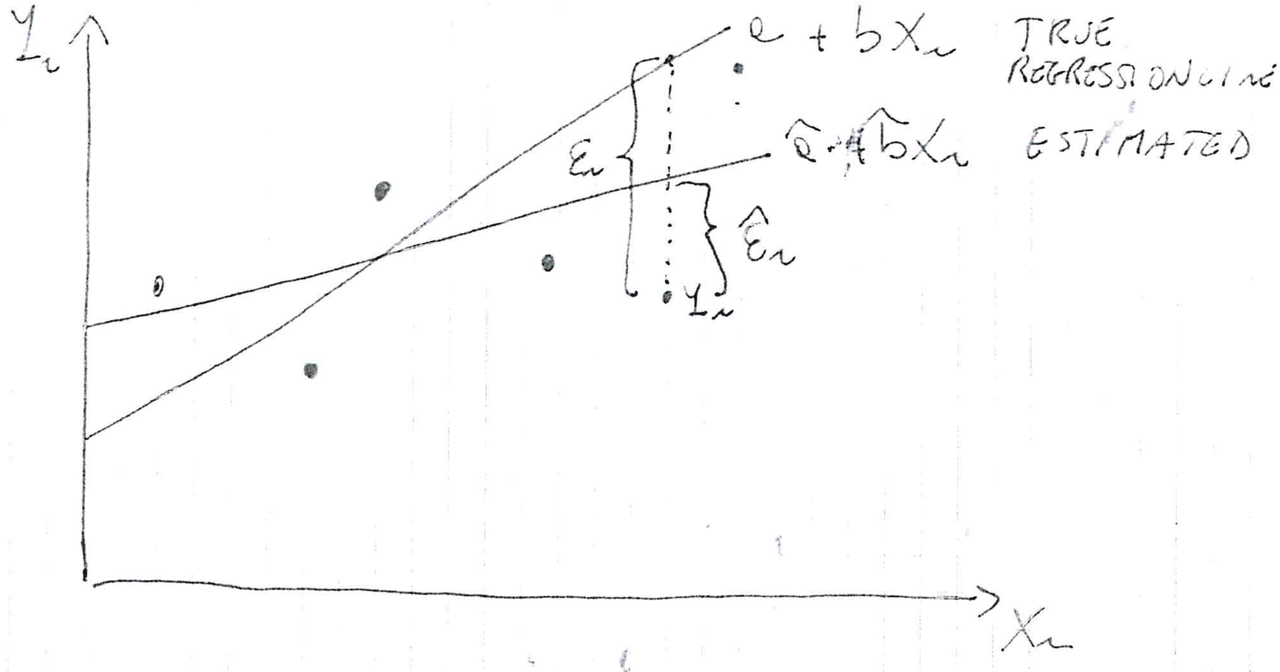
by definition: $\varepsilon_i = Y_i - a - bX_i$

LET US DEFINE: $\hat{Y}_i = \hat{a} + \hat{b} X_i$

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

$$= Y_i - \hat{a} - \hat{b} X_i$$

$$= \text{RESIDUAL}$$



HOW TO ESTIMATE a and b ?

OLS

ORDINARY LEAST SQUARE

I have defined $\hat{\varepsilon}_i = Y_i - \hat{a} - \hat{b} X_i$

OLS are defined as:

$$\underset{\hat{a},\hat{b}}{\mathrm{Min}} \sum_{i=1}^{N} \hat{\varepsilon}_i^2$$

$$\underset{\hat{a},\hat{b}}{\mathrm{Min}} \sum_{i=1}^{N} (Y_i - \hat{a} - b X_i)^2$$

F.O.C.

$$\begin{cases} ① \ \dfrac{\partial \sum_{i=1}^{N} (Y_i - \hat{a} - \hat{b} X_i)^2}{\partial \hat{a}} = 0 \\ ② \ \dfrac{\partial \sum_{i=1}^{N} (Y_i - \hat{a} - \hat{b} X_i)^2}{\partial \hat{b}} = 0 \end{cases}$$

① $-2 \sum_{i=1}^{N} (Y_i - \hat{a} - \hat{b} X_i) = 0$

② $-2 \sum_{i=1}^{N} X_i (Y_i - \hat{a} - \hat{b} X_i) = 0$

(1) $$\sum_{i=1}^{N} (Y_i - \hat{a} - \hat{b} X_i) = 0$$

$$\sum_{i=1}^{N} Y_i - \hat{a} N - \hat{b} \sum_{i=1}^{N} X_i = 0$$

$$\hat{a} = \frac{1}{N} \sum_{i=1}^{N} Y_i - \hat{b} \frac{1}{N} \sum_{i=1}^{N} X_i$$

$$\Rightarrow \hat{a} = \underbrace{\overline{Y}}_{\text{MEAN } Y} - \hat{b} \underbrace{\overline{X}}_{\text{MEAN } X}$$

also: $\overline{Y} = \hat{a} + \hat{b} \overline{X}$

by (initial) definition:

$$Y_i = \hat{a} + \hat{b} X_i + \hat{\varepsilon}_i$$

from the definition of $\hat{a} = \overline{Y} - \hat{b} \overline{X}$:

$$Y_i = \overline{Y} - \hat{b} \overline{X} + \hat{b} X_i + \hat{\varepsilon}_i$$

$$(Y_i - \overline{Y}) = \hat{b} (X_i - \overline{X}) + \hat{\varepsilon}_i$$

AXES TRANSLATION:

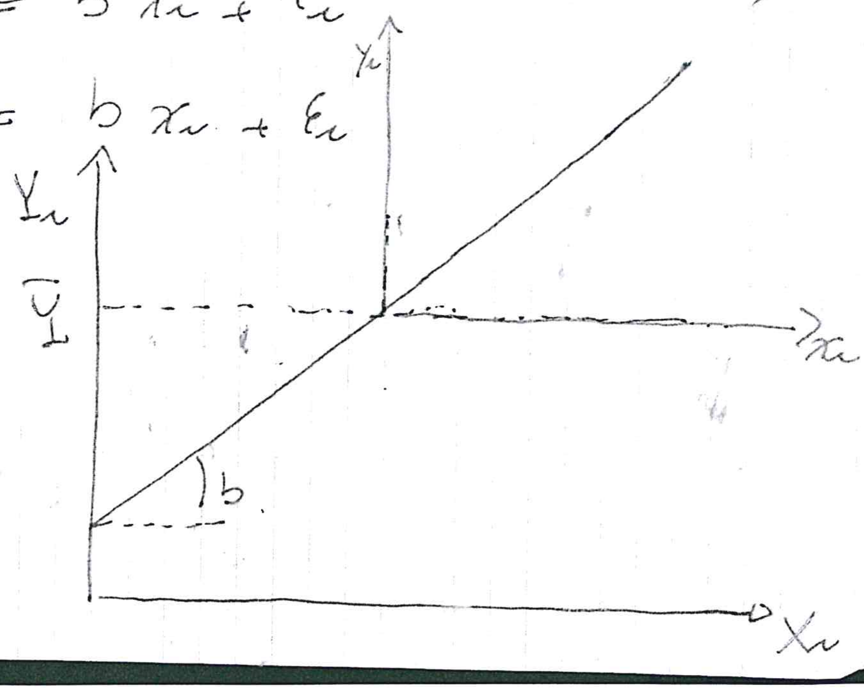
$$y_i = \hat{b} x_i + \hat{\varepsilon}_i$$

and

$$y_i = b x_i + \varepsilon_i$$

$y_i = (Y - \overline{Y})$

$x_i = (X - \overline{X})$



I

$$y_i = b x_i + \varepsilon_i$$

(2) $$-2 \sum_{i=1}^{N} x_i (y_i - \hat{b} x_i) = 0$$

$$\sum_{i=1}^{N} x_i (y_i - \hat{b} x_i) = 0$$

$$\sum_{i=1}^{N} x_i y_i - \hat{b} \sum_{i=1}^{N} x_i^2 = 0$$

$$\hat{b} = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2}$$

$$= \frac{\frac{1}{N} \sum_{i=1}^{N} \left[(X_i - \bar{X})(Y_i - \bar{Y}) \right]}{\frac{1}{N} \sum_{i=1}^{N} (X_i - \bar{X})^2}$$

$$= \frac{COV(X, Y)}{VAR(X)}$$

I

F.O.C. OLS

(1)
$$\frac{\partial \sum_{i=1}^{N} (Y_i - \hat{a} - \hat{b} X_i)^2}{\partial \hat{a}} = 0$$

$$\Rightarrow \sum_{i=1}^{N} (Y_i - \hat{a} - \hat{b} X_i) = 0$$

$$\Rightarrow \sum_{i=1}^{N} \hat{\varepsilon}_i^2 = 0$$

(2)
$$\frac{\partial \sum_{i=1}^{N} (Y_i - \hat{a} - \hat{b} X_i)^2}{\partial \hat{b}} = 0$$

$$\Rightarrow \sum_{i=1}^{N} X_i (Y_i - \hat{a} - \hat{b} X_i) = 0$$

$$\Rightarrow \sum_{i=1}^{N} X_i \hat{\varepsilon}_i = 0$$

$$\begin{cases} \sum_{i=1}^{N} \hat{\varepsilon}_i = 0 \\ \sum_{i=1}^{N} X_i \hat{\varepsilon}_i = 0 \end{cases}$$

NORMAL EQUATIONS SYSTEM

S.O.C

HESSIAN MATRIX DEFINITE POSITIVE
(for being minimizing)

$$H = \begin{pmatrix} \dfrac{\partial^2 \sum_{i=1}^{N} \hat{\varepsilon}_i^2}{\partial \hat{a}^2} & \dfrac{\partial^2 \sum_{i=1}^{N} \hat{\varepsilon}_i^2}{\partial \hat{a} \hat{b}} \\ \dfrac{\partial^2 \sum \hat{\varepsilon}_i^2}{\partial \hat{a} \hat{b}} & \dfrac{\partial^2 \sum_{i=1}^{N} \hat{\varepsilon}_i^2}{\partial \hat{b}^2} \end{pmatrix}$$

$$H = \begin{pmatrix} 2N & 2\sum_{i=1}^{N} x_i \\ 2\sum_{i=1}^{N} x_i & 2\sum_{i=1}^{N} x_i^2 \end{pmatrix}$$

In general, we have to study eigenvalues, they should be > 0 for a definite POSITIVE MATRIX

2x2 Matrix : — The elements in the diagonal > 0

— Det > 0

- Elements in the diagonal are > 0
- $|H| = 2N\,2\sum_{i=1}^{N} x_i^2 - 4\left(\sum_{i=1}^{N} x_i\right)^2 = 4N\sum_{i=1}^{N} x_i^2 - 4(N\bar{x})^2$

$$= 4N\left(\sum_{i=1}^{N} x_i^2 - N\bar{x}^2\right) = 4N\sum_{i=1}^{N}(x_i - \bar{x})^2 > 0$$

II

$Y_i = a + b X_i + \varepsilon_i$ ...TRUE (POPULATION) REGRESSION MODEL

$\hat{Y}_i = \hat{a} + \hat{b} X_i$ ESTIMATED REGRESSION

$\hat{\varepsilon}_i = Y_i - \hat{Y}_i$ RESIDUALS

$= Y_i - \hat{a} - \hat{b} X_i$

OLS Objective function:

$$\min_{\hat{a}, \hat{b}} \sum_{i=1}^{N} \hat{\varepsilon}_i^2$$

$$\min_{\hat{a}, \hat{b}} \sum_{i=1}^{N} (Y_i - \hat{a} - b X_i)^2$$

F.O.C

(1) $\dfrac{\partial \sum_{i=1}^{N} \hat{\varepsilon}_i^2}{\partial \hat{a}} = 0 \implies \sum_{i=1}^{N} \hat{\varepsilon}_i = 0$

(2) $\dfrac{\partial \sum_{i=1}^{N} \hat{\varepsilon}_i^2}{\partial \hat{b}} = 0 \implies \sum_{i=1} X_i \hat{\varepsilon}_i = 0$

NORMAL EQUATIONS SYSTEM

S.O.C HESSIAN DEFINITE POSITIVE

$$H = \begin{pmatrix} \dfrac{\partial^2 \sum_{i=1}^{N} \hat{\varepsilon}_i^2}{\partial \hat{a}^2} & \dfrac{\partial^2 \sum_{i=1}^{N} \hat{\varepsilon}_i^2}{\partial \hat{a} \partial \hat{b}} \\ \dfrac{\partial^2 \sum_{i=1}^{N} \hat{\varepsilon}_i^2}{\partial \hat{a} \partial \hat{b}} & \dfrac{\partial^2 \sum_{i=1}^{N} \hat{\varepsilon}_i^2}{\partial \hat{b}^2} \end{pmatrix} = \begin{pmatrix} 2N & 2\sum_{i=1}^{N} X_i \\ 2\sum_{i=1}^{N} X_i & 2\sum_{i=1}^{N} X_i^2 \end{pmatrix}$$

DIAGONAL ELEMENTS $\left(2N, 2\sum_{i=1} X_i^2\right) > 0$

$|H| > 0$

So H IS DEFINITE POSITIVE

$$\hat{a} = \bar{Y} - \hat{b} \bar{X}$$

$$\hat{b} = \frac{\sum_{i=1}^{N} [(X_i - \bar{X})(Y_i - \bar{Y})]}{\sum_{i=1}^{N} (X_i - \bar{X})^2} = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} y_i^2}$$

II

HYPHOTESIS

SIMPLE LINEAR REGRESSION MODEL

(1) $Y_i = a + b X_i + \varepsilon_i$ "TRUE" MODEL

$\Rightarrow y_i = b x_i + \varepsilon_i$

(2) $\mathbb{E}(\varepsilon_i | X_i) = \mathbb{E}(\varepsilon_i)$

(3) $\mathbb{E}(\varepsilon_i) = 0$

(2)+(3) $\mathbb{E}(\varepsilon_i | X_i) = 0$ ZERO CONDITIONAL MEAN ASSUMPTION

(4) $\mathbb{E}(\varepsilon_i)^2 = \sigma^2$ OMOSCHEDASTICITY

(Variance does not depend on i)

(5) $\mathbb{E}(\varepsilon_i \varepsilon_j) = 0 \ \forall i \neq j$ (CROSS-SECTIONAL INDEPENDENCE)

Main goal: estimation of b, which is the SLOPE of the regression line and also

$$\frac{\partial Y_i}{\partial X_i} = b$$

CETERIS PARIBUS INTERPRETATION

The estimation of a is generally less important

WHAT ABOUT THE OLS ESTIMATOR $\hat{b}$?

RECAP: $\hat{b} = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_i x_i^2}$ is the OLS estimator of b

$\hat{b}$ = "a given value" is the OLS estimate of b

$$\hat{b} = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2}$$

HP $\underline{\underline{2 \text{ BIS}}}$: X DETERMINISTIC VARIABLE (NON STOCHASTIC)

LET US DEFINE WEIGHTS: $w_i = \frac{x_i}{\sum_{i=1}^{N} x_i^2}$

I CAN REWRITE:

$$\hat{b} = \sum_{i=1}^{N} w_i y_i = \sum_{i=1}^{N} w_i (b x_i + \varepsilon_i) \Rightarrow \text{from } 1$$

$\hat{b}$ = WEIGHTED SUMMATION OF THE OBSERVATIONS, y_i

$$\hat{b} = b \sum_{i=1}^{N} w_i x_i + \sum_{i=1}^{N} w_i \varepsilon_i$$

$$\hat{b} = b \frac{\sum_{i=1}^{N} x_i^2}{\sum_{i=1}^{N} x_i^2} + \sum_{i=1}^{N} w_i \varepsilon_i$$

$$\hat{b} = b + \sum_{i=1}^{N} w_i \varepsilon_i$$

$\hat{b}$ is a random variable which - ex ante - depends on

- b (a parameter)
- a weighted summation of random variables

WHAT ALLOWS UNBIASEDNESS OF $\hat{b}$

Recap: a "generic" estimator $\hat{b}$ is unbiased if

$$\mathbb{E}(\hat{b}) = b$$

$$\text{BIAS}(\hat{b}) = \mathbb{E}(\hat{b}) - b$$

$$\hat{b} = b + \sum_{i=1}^{N} w_i \varepsilon_i$$

$$\mathbb{E}(\hat{b}) = b + \mathbb{E}\left(\sum_{i=1}^{N} w_i \varepsilon_i\right)$$

\# HP 2BIS : $w_i = \dfrac{x_i}{\sum_{i=1}^{N} x_i^2}$ are FIXED WEIGHTS (NOT RANDOM)

$$\mathbb{E}(\hat{b}) = b + \sum_{i=1}^{N} w_i \mathbb{E}(\varepsilon_i)$$

\# HP 3 : $\mathbb{E}(\varepsilon_i) = 0$

$$\mathbb{E}(\hat{b}) = b + \sum_{i=1}^{N} w_i \cdot 0$$

$$\mathbb{E}(\hat{b}) = b$$

$\hat{b}$ IS AN UNBIASED ESTIMATOR OF b

Variance of $\hat{b}$

$$Var(\hat{b}) = \mathbb{E}(\hat{b} - \mathbb{E}(\hat{b}))^2$$
$$= \mathbb{E}(\hat{b} - b)^2$$

We also know that: $\hat{b} = b + \sum_{i=1}^{N} w_i \varepsilon_i$

$$(\hat{b} - b) = \sum_{i=1}^{N} w_i \varepsilon_i$$

$$Var(\hat{b}) = \mathbb{E}\left(\sum_{i=1}^{N} w_i \varepsilon_i\right)^2$$

$$= \mathbb{E}\left[w_1^2 \varepsilon_1^2 + w_2^2 \varepsilon_2^2 + \dots + w_n^2 \varepsilon_n^2 + 2(w_1 \varepsilon_1 w_2 \varepsilon_2 + \dots + w_{N-1} \varepsilon_{N-1} w_N \varepsilon_N)\right]$$

HP 4 $\rightarrow \mathbb{E}(\varepsilon_i)^2 = \sigma^2$

HP 5 $\rightarrow \mathbb{E}(\varepsilon_i \varepsilon_j) = 0$

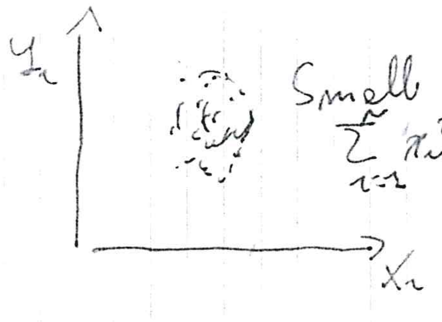
$$Var(\hat{b}) = \sigma^2 \sum_{i=1}^{N} w_i^2 + 0$$

$$Var(\hat{b}) = \sigma^2 \sum_{i=1}^{N} w_i^2$$

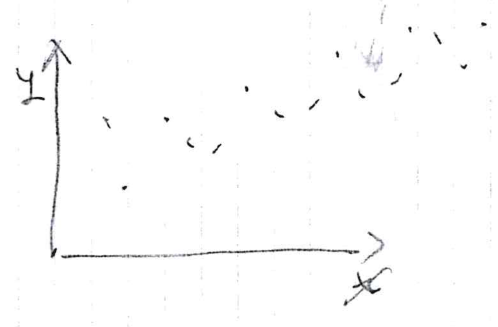
What about $\sum_{i=1}^{N} w_i^2$?

$$\sum_{i=1}^{N} w_i^2 = \sum_{i=1}^{N} \left(\frac{x_i}{\sum_{i=1}^{N} x_i^2}\right)^2 = \frac{\cancel{\sum_{i=1}^{N} x_i^2}}{\left(\sum_{i=1}^{N} x_i^2\right)^{\cancel{2}}} = \frac{1}{\sum_{i=1}^{N} x_i^2}$$

$$Var(\hat{b}) = \frac{\sigma^2}{\sum_{i=1}^{N} x_i^2}$$



Small $\sum_{i=1}^{N} x_i^2$



1 (6)

OLS

$$\min \sum_{i=1}^{N} \hat{\varepsilon}_i^2$$

(A) $\hat{b} = \sum_{i=1}^{N} w_i y_i$ $\qquad w_i = \dfrac{x_i}{\sum_{i=1}^{N} x_i^2}$

(B) $E(\hat{b}) = b$ $\qquad$ HP 2 + 3

(C) $Var(\hat{b}) = \dfrac{\sigma^2}{\sum_{i=1}^{N} x_i^2}$ $\qquad$ HP 4 + 5

GAUSS - MARKOV THEOREM:

UNDER ALL THE HYPOTHESES WE HAVE DONE, THE ESTIMATOR $\hat{b}$ IS BLUE

BEST LINEAR UNBIASED ESTIMATOR

BEST ??

↳ MINIMUM VARIANCE (AMONG THE UNBIASED ESTIMATORS)

II 1

ALTERNATIVE PROOF OF $\mathbb{E}(\hat{b}) = b$

UNDER HP2 INSTEAD OF HP2BIS

We already know that:

$$\hat{b} = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2} \qquad y_i = b x_i + \varepsilon_i$$

$$\hat{b} = b + \sum_{i=1}^{N} k_i \varepsilon_i \qquad k_i = \frac{x_i}{\sum_{i=1}^{N} x_i^2}$$

$$\hat{b} = b + \frac{\sum_{i=1}^{N} x_i \varepsilon_i}{\sum_{i=1}^{N} x_i^2}$$

$$\mathbb{E}(\hat{b}) = b + \mathbb{E}\left(\frac{\sum_{i=1}^{N} x_i \varepsilon_i}{\sum_{i=1}^{N} x_i^2}\right)$$

USE LAW OF ITERATED EXPECTATIONS (LIE)

X = R.V ; x_i = REALISATION

$$\mathbb{E}(Y) = \underbrace{\sum_{i=1}^{N} \mathbb{E}(Y \mid X = x_i) \cdot \underbrace{\Pr(X = x_i)}}$$

$\mathbb{E}(Y)$ = WEIGHTED AVERAGE - WITH WEIGHTS EQUAL TO THE PROBABILITY OF A GIVEN REALISATION OF X - OF CONDITIONAL EXPECTED VALUES OF Y GIVEN $X = x_i$

$$\mathbb{E}(Y) = \mathbb{E}\left[\mathbb{E}(Y \mid X)\right]$$

$\mathbb{E}(Y)$ = EXPECTED VALUE OF THE EXPECTED VALUE OF $Y \mid X$

II

LIE: $\mathbb{E}(Y) = \mathbb{E}[\mathbb{E}(Y|X)]$ 8

Y is a "generic" random variable

I can write:

$$\mathbb{E}(\varepsilon) = \mathbb{E}[\mathbb{E}(\varepsilon|X)]$$

Consequently, from $\mathbb{E}(\hat{b}) = b + \mathbb{E}\left(\frac{\sum_{i=1}^{N} x_i \varepsilon_i}{\sum_{i=1}^{N} x_i^2}\right)$:

$$\mathbb{E}\left(\frac{\sum_{i=1}^{N} x_i \varepsilon_i}{\sum_{i=1}^{N} x_i^2}\right) = \mathbb{E}\left(\frac{\sum_{i=1}^{N} x_i \mathbb{E}(\varepsilon_i|x_i)}{\sum_{i=1}^{N} x_i^2}\right)$$

From HP 2 and 3:

$$\mathbb{E}(\varepsilon_i | x_i) = 0$$

$$\mathbb{E}(\hat{b}) = b + 0$$

$$\mathbb{E}(\hat{b}) = b$$

PROOF THAT $\hat{b}$ IS BLUE

We have to show that $\hat{b}$ has the lowest variance

- Let us define a "generic" linear ESTIMATOR:

$$\tilde{b} = \sum_{i=1}^{N} c_i y_i \qquad y_i = x_i b + \varepsilon_i$$

$$= b \sum_{i=1}^{N} c_i x_i + \sum_{i=1}^{N} c_i \varepsilon_i$$

c_i are deterministic WEIGHTS (NON STOCHASTIC)

$\tilde{b}$ IS a RANDOM VARIABLE !! Why??

$$\mathbb{E}(\tilde{b}) = b \sum_{i=1}^{N} c_i x_i + \cancel{\sum_{i=1}^{N} c_i \mathbb{E}(\varepsilon_i)}$$

(c_i DETERMINISTIC WEIGHTS) $\qquad \mathbb{E}(\varepsilon_i) = 0$ HP3

So, To have $\mathbb{E}(\tilde{b}) = b$

We need: $\boxed{\sum_{i=1}^{N} c_i x_i = 1}$ CONDITION FOR UNBIASEDNESS

$$Var(\tilde{b}) = \mathbb{E}\left(\tilde{b} - \mathbb{E}(\tilde{b})\right)^2$$

$$= \mathbb{E}\left(\sum_{i=1}^{N} c_i \varepsilon_i\right)^2$$

$$= \sigma^2 \sum_{i=1}^{N} c_i^2$$

HP(4)

II IN ORDER $\hat{b}$ TO BE BLUE:

$$\left(Var(\hat{b}) = \sigma^2 \sum_{i=1}^{n} c_i^2 \right)$$

$$\text{Min} \sum_{i=1}^{n} c_i^2$$

$$\text{u.c.} \sum_{i=1}^{n} c_i x_i = 1$$

CONDITION FOR $E(\hat{b}) = b$

$$\mathcal{L} = \sum_{i=1}^{n} c_i^2 - 2\lambda \left(\sum_{i=1}^{n} c_i x_i - 1 \right)$$

↳ only for simplify algebra

F.O.C

$$\frac{\partial \mathcal{L}}{\partial c_i} = 0 \Rightarrow 2c_i - 2\lambda x_i = 0$$

$$c_i = \lambda x_i$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 0 \Rightarrow -2\left(\sum_{i=1}^{n} c_i x_i - 1 \right) = 0$$

$$\sum_{i=1}^{n} c_i x_i = 1$$

Solving the system:

$$\sum_{i=1}^{n} c_i x_i = 1$$

$$\sum_{i=1}^{n} (\lambda x_i) x_i = 1$$

$$\lambda \sum_{i=1}^{n} x_i^2 = 1 \Rightarrow \lambda = \frac{1}{\sum_{i=1}^{n} x_i^2}$$

$$\boxed{c_i = \frac{x_i}{\sum_{i=1}^{n} x_i^2}} \Rightarrow c_i = w_i \quad !!!$$

If $c_i = w_i \Rightarrow \hat{b}$ is BLUE

and $Var(\hat{b}) = \sigma^2 / \sum_{i=1}^{n} x_i^2$

III GEOMETRIC REPRESENTATION OF OLS

GOAL: TO DEFINE OLS FROM ANOTHER POINT OF VIEW: FROM THE SPACE $\mathbb{R}^N$ WHICH IS CALLED THE SPACE OF OBSERVATIONS

Recap: • let $\underset{N\times 1}{X}$ be a vector belonging from $\mathbb{R}^N$

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix}$$

• $\|X\|^2$ is the square of the EUCLIDEAN NORM

$$\|X\| = (X'X)^{1/2} \quad \text{EUCLIDEAN NORM}$$

$$\|X\|^2 = X'X = \sum_{i=1}^{N} x_i^2$$

• TWO "NAIVE" DEFINITIONS:

→ VECTOR SPACE: COLLECTION OF VECTORS WHICH CAN BE ADDED TOGHETER AND MULTIPLIED BY SCALARS

$\mathbb{R}^N$ IS A VECTOR SPACE

→ VECTOR SUBSPACE: VECTOR SPACE THAT IS A SUBSET OF ANOTHER - HIGHER DIMENSION - VECTOR SPACE

III THEOREM

Let $\mathcal{L}$ be a VECTOR SUBSPACE belonging from $\mathbb{R}^N$, the minimum distance in terms of euclidean norm between a vector $Y \in \mathbb{R}^N$ and $\mathcal{L}$ is given by the ORTHOGONAL PROJECTION OF Y ONTO $\mathcal{L}$

- VECTOR REPRESENTATION OF THE SIMPLE LINEAR REGRESSION MODEL

Let define the following vectors belonging from $\mathbb{R}^N$

$$\underset{N\times 1}{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \quad \underset{N\times 1}{X} = \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} \quad \underset{N\times 1}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix} \quad \underset{N\times 1}{\hat{Y}} = \begin{pmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_N \end{pmatrix} \quad \underset{N\times 1}{\hat{\varepsilon}} = \begin{pmatrix} \hat{\varepsilon}_1 \\ \vdots \\ \hat{\varepsilon}_N \end{pmatrix}$$

Let define also:

$$\underset{(N\times 1)}{\iota_N} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

- I can rewrite $y_i = a + b x_i + \varepsilon_i$ as:

$$\underset{N\times 1}{Y} = \underset{N\times 1}{\iota_N} \underset{1\times 1}{a} + \underset{N\times 1}{X} \underset{1\times 1}{b} + \underset{N\times 1}{\varepsilon}$$

$$\underset{N\times 1}{\hat{Y}} = \underset{N\times 1}{\iota_N} \underset{1\times 1}{\hat{a}} + \underset{N\times 1}{X} \hat{b}$$

$$\underset{N\times 1}{Y} = \underset{N\times 1}{\hat{Y}} + \underset{N\times 1}{\hat{\varepsilon}}$$

$$\underset{N\times 1}{\hat{\varepsilon}} = \underset{N\times 1}{Y} - \underset{N\times 1}{\hat{Y}}$$

$\hat{a}$, $\hat{b}$ OLS ESTIMATES

- "Implicit" hypothesis of OLS (from $Y_i = b x_i + \varepsilon_i$)
  - $\sum_{i=1} x_i^2 \neq 0$ (for the existence of $\hat{b}$)

this because $\hat{b} = \dfrac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2}$

- Let rewrite this hypothesis in a different way:

$$\nexists \lambda \in \mathbb{R} \text{ such that } X = \lambda 1_N$$

the VECTORS X and 1_N are LINEARLY INDEPENDENT

- If this hypothesis is true:

→ X and 1_N define a vector subspace from $\mathbb{R}^N$ of dimension 2
that is, a <u>plan</u> noted as $\mathcal{L}(X, 1_N)$

→ $\hat{Y}$ is by definition a linear combination

$$\hat{Y} = 1_N \hat{a} + X \hat{b}$$

thus:

$$\hat{Y} \in \mathcal{L}(X, 1_N)$$

$$Y \in \mathbb{R}^N$$

- GOAL: TO MINIMIZE THE DISTANCE BETWEEN Y and $\hat{Y}$ USING THE METRIC OF THE EUCLIDEAN NORM (OR EQUALLY, THE SQUARE OF THE NORM)

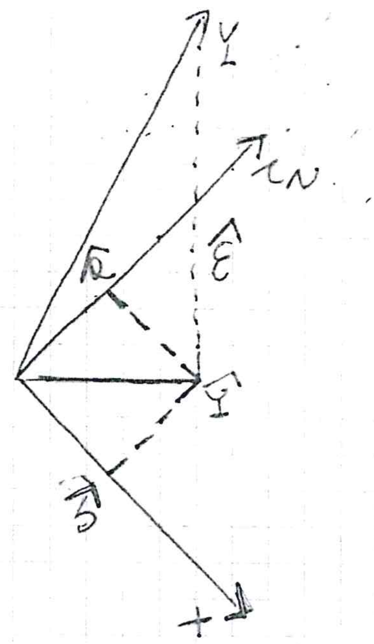
$$\mathrm{Min}\ \| \underline{Y} - \hat{\underline{Y}} \|$$
$$\mathrm{Min}\ \| \hat{\varepsilon} \|$$

or

$$\mathrm{Min}\ \| \underline{Y} - \hat{\underline{Y}} \|^2$$
$$\mathrm{Min}\ \| \hat{\varepsilon} \|^2 \Longleftrightarrow \mathrm{Min} \sum_{i=1}^{N} \hat{\varepsilon}_i^2$$

then, I can apply the previous THEOREM:

- THE MINIMUM DISTANCE ($\mathrm{Min}\ \| \hat{\varepsilon} \|^2$) is OBTAINED DEFINING $\hat{Y}$ as ORTHOGONAL PROJECTION OF $\underline{Y}$ ONTO $\mathcal{L}(X, 1_N)$



The COORDINATES ON The PLAN $\mathcal{L}(X, 1_N)$ are $\hat{b}$ and $\hat{a}$

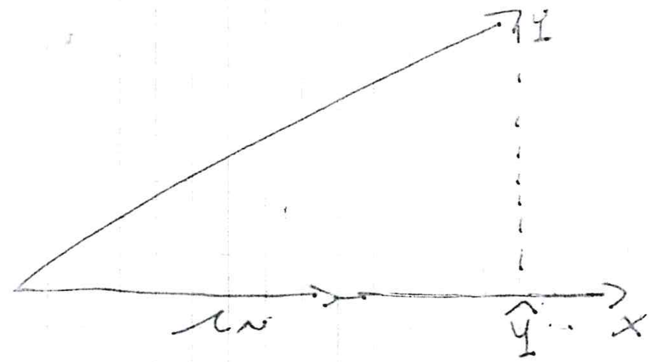
$\hat{Y}$ is the ORTHOGONAL PROJECTION OF $\underline{Y}$ ($\underline{Y} \in R^N$) ONTO THE PLAN $\mathcal{L}(X, 1_N)$

$$\hat{Y} = 1_N \hat{a} + X \hat{b}$$
$$Y = \hat{Y} + \hat{\varepsilon}$$
$$\hat{\varepsilon} \perp \mathcal{L}(X, 1_N) \Longrightarrow \begin{array}{l} \hat{\varepsilon} \perp X \\ \hat{\varepsilon} \perp 1_N \end{array}$$

$$\begin{cases} \hat{\varepsilon}' 1_N = 0 \\ \hat{\varepsilon}' X = 0 \end{cases}$$ NORMAL EQUATIONS SYSTEM

III

- What happens if $\exists \lambda$ such that $X = \lambda 1_N$



$\hat{a}$ and $\hat{b}$ are NOT DETERMINED

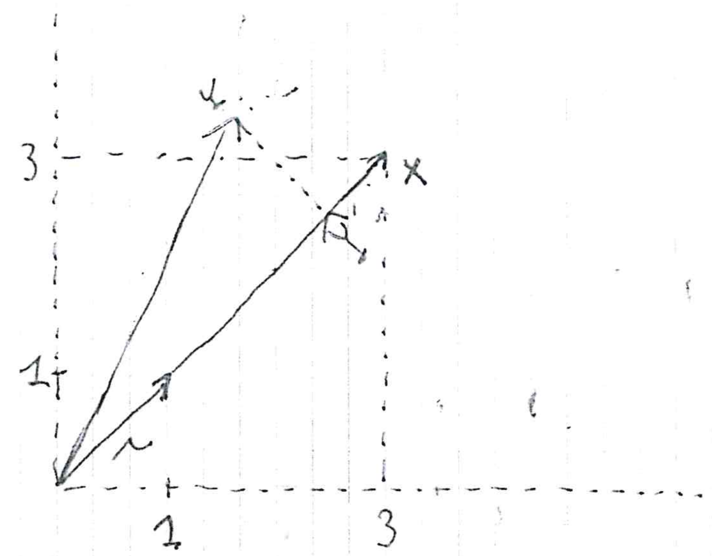
- WHEN IT HAPPENS ??

X IS CONSTANT (Remember $\sum_{i=1}^{N} x_i^2 = 0$ $\Rightarrow X_i$ constant)

EXAMPLE in $\mathbb{R}^2$

$$1_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad X = \begin{pmatrix} 3 \\ 3 \end{pmatrix}$$

$$X = 3 \, 1_2$$



(IV

$\hat{b}$ (OLS) IS BLUE

Min. Variance among all the UNBIASED estimators

POSSIBLE TRADE-OFF BETWEEN UNBIASEDNESS AND EFFICIENCY

UNBIASED ESTIMATOR : $\mathbb{E}(\hat{b}) = b$

BIAS : $\mathbb{E}(\hat{b}) - b \neq 0$

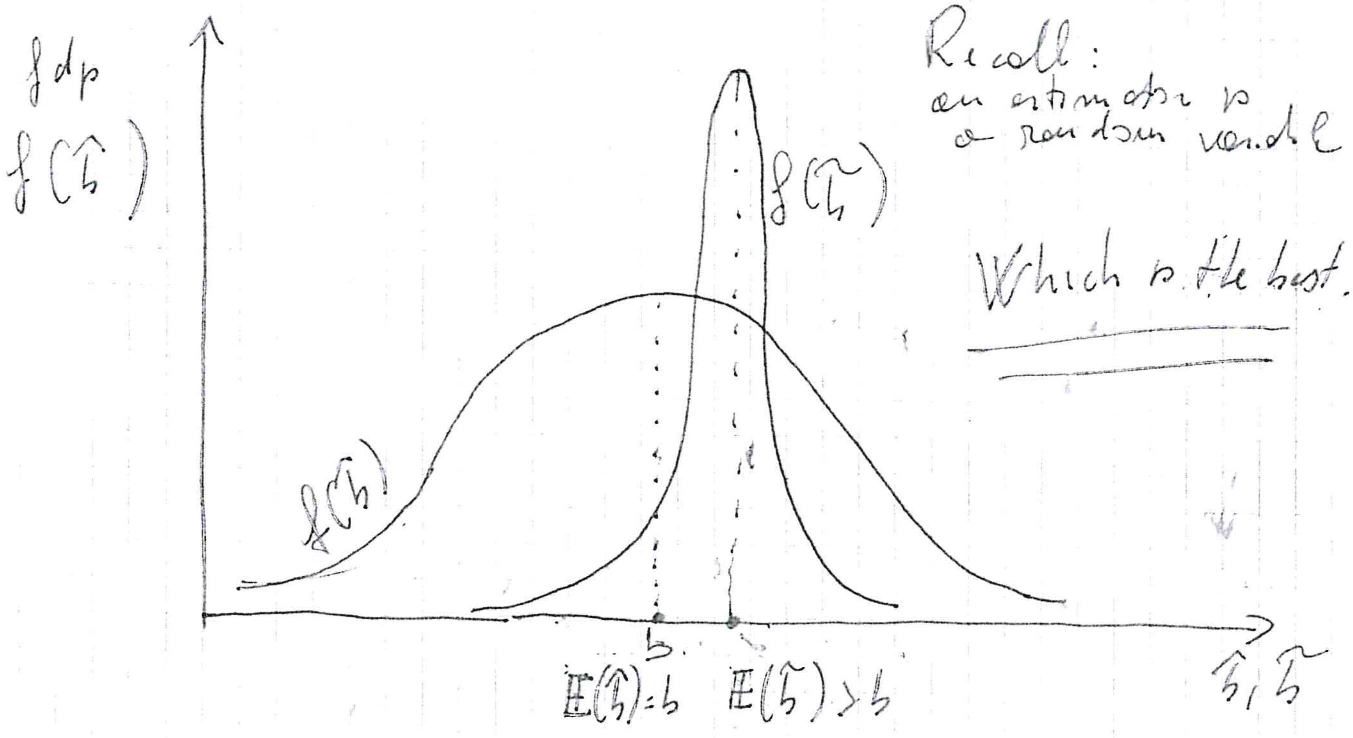
$$\mathbb{E}(\hat{b}) - b \begin{cases} > 0 & \text{OVER ESTIMATE} \\ < 0 & \text{UNDER ESTIMATE} \end{cases}$$

Let suppose we have two alternative estimators of b:

$\hat{b}$ is UNBIASED $\Rightarrow \mathbb{E}(\hat{b}) = b$

$\tilde{b}$ is BIASED (OVER ESTIMATE) $\Rightarrow \mathbb{E}(\tilde{b}) > b$

but $Var(\hat{b}) > Var(\tilde{b})$



Recall: an estimator is a random variable

Which is the best?

Let introduce a Loss function
I call this loss function as Mean Square Error

$$MSE(\tilde{b}) = \mathbb{E}(\tilde{b} - b)^2$$

Please note that it is close to the variance:

$$Var(\tilde{b}) = \mathbb{E}(\tilde{b} - \mathbb{E}(\tilde{b}))^2$$

A possible way to choose between $\hat{b}$ and $\tilde{b}$ is choosing the one with lower MSE
According to the MSE a biased estimator can be better than an unbiased one but with lower variance:

I can rewrite:

$$MSE(\tilde{b}) = \mathbb{E}(\underbrace{\tilde{b} - \mathbb{E}(\tilde{b})} + \underbrace{\mathbb{E}(\tilde{b}) - b}_{bias})^2$$

$$= Var(\tilde{b}) + bias^2(\tilde{b}) + 2\,bias(\tilde{b})\underbrace{[\mathbb{E}(\tilde{b}) - \mathbb{E}(\tilde{b})]}_{0}$$

$$MSE(\tilde{b}) = Var(\tilde{b}) + bias^2(\tilde{b})$$

IV $\hat{b}_{(OLS)}$

$$MSE(\hat{b}) = Var(\hat{b}) + 0$$

$\hat{b}$ has minimum variance among all the unbiased estimators

$\hookrightarrow \hat{b}$ has minimum MSE among all the unbiased estimators

However...

there may exist an estimator $\tilde{b}$ which is biased but:

$$MSE(\tilde{b}) < MSE(\hat{b})$$

this happens if

$$Var(\tilde{b}) < Var(\hat{b}) - bias^2(\tilde{b})$$

N

Test of HYPOTHESIS
in the LINEAR SIMPLE REGRESSION Model

(1) $Y_i = a + b X_i + \varepsilon_i$ or $y_i = b x_i + \varepsilon_i$

(2) (3) (4) (5) $\Rightarrow \hat{b}$ (OLS) BLUE

(6) $\varepsilon_i \sim_i N(0, \sigma^2)$ NORMALITY

$\Rightarrow$ INDIPENDENTLY

It means that every ε_i is ..
Normal an all the i are
independent, that is $\mathbb{E}(\varepsilon_i, \varepsilon_j) = 0$

We already know that:

$$\hat{b} = b + \sum_{i=1}^{N} w_i \varepsilon_i \qquad w_i = \frac{x_i}{\sum_{i=1}^{N} x_i^2}$$

A feature of the Normal distribution is that:
A linear combination of Normal Random variables
is also a Normal distribution:

$\Rightarrow$ If $\varepsilon_i \sim N \Rightarrow \hat{b} = b + \sum_{i=1}^{N} w_i \varepsilon_i \sim N$

$$\hat{b} \sim N\left(b, \frac{\sigma^2}{\sum_{i=1}^{N} x_i^2}\right)$$

$$(\hat{b} - b) \sim N\left(0, \frac{\sigma^2}{\sum_{i=1}^{N} x_i^2}\right)$$

IV WHAT ABOUT σ^2

σ^2 is the variance of ε_i

$$\mathbb{E}(\varepsilon_i)^2 = \mathbb{E}(\varepsilon_i - \mathbb{E}(\varepsilon_i))^2$$
$$= \mathbb{E}(\varepsilon_i - 0)^2$$
$$= \sigma^2$$

HOWEVER ε_i IS NOT KNOWN
thus σ^2 is not observable

A UNBIASED ESTIMATOR OF σ^2 is s^2

$$s^2 = \frac{\sum_{i=1}^{N} \hat{\varepsilon}_i^2}{N-2}$$

$N-2$ degrees of freedom
2 constraints

$$\sum_{i=1}^{N} \hat{\varepsilon}_i = 0$$

$$\sum_{i=1}^{N} x_i \hat{\varepsilon}_i = 0$$

$$\mathbb{E}(s^2) = \sigma^2$$

s => standard error of regression

IV

TEST OF HYPOTHESIS about b

Anno prox:
Prendere appunti
statistica z

$H_0: b = b_0$

$H_1: b \neq b_0$

OLS ESTIMATION

$\Rightarrow \hat{b}$ $\qquad \hat{b} = \sum_{i=1}^{N} x_i y_i \Big/ \sum_{i=1}^{N} x_i^2$

$\Rightarrow s_{\hat{b}}$

$$\sigma^2_{\hat{b}} = \frac{\sigma^2}{\sum_{i=1}^{N} x_i^2}$$ IS NOT KNOWN, WHILE

$$s^2_{\hat{b}} = \frac{s^2}{\sum_{i=1}^{N} x_i^2}$$ IS KNOWN

then, the test is

$$\frac{\hat{b} - b_0}{s_{\hat{b}}} \sim t_{N-2}$$

Generally, we want to test the following

$H_0: b = 0$

$H_1: b \neq 0$

$$\Rightarrow \frac{\hat{b}}{s_{\hat{b}}} \sim t_{N-2}.$$

EXTENDING THE SIMPLE LINEAR REGRESSION MODEL (SLRM) TO THE CASE OF MULTIPLE EXPLANATORY VARIABLES

- LET US START WITH 2 EXPLANATORY VARIABLES

(1)
→ RELEVANT OMITTED VARIABLES IN THE ESTIMATED MODEL

→ IRRELEVANT INCLUDED VARIABLES IN THE ESTIMATED MODEL

(2) EASIER UNDERSTANDING OF THE OLS ESTIMATOR WITH RESPECT TO THE CASE OF MORE THAN 2 X, WHILE KEEPING THE MAIN FEATURES

Let us write the model:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

Define $y_i = Y_i - \bar{Y}$; $x_{2i} = X_{2i} - \bar{X}_2$; $x_{3i} = X_{3i} - \bar{X}_3$

$$\bar{Y} = \beta_1 + \beta_2 \bar{X}_2 + \beta_3 \bar{X}_3 + 0$$

$$Y_i - \bar{Y} = \cancel{\beta_1 - \beta_1} + \beta_2 (X_{2i} - \bar{X}_2) + \beta_3 (X_{3i} - \bar{X}_3) + \varepsilon_i$$

$$y_i = \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i}$$

$\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$: OLS ESTIMATES

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

V

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

$$\left(y_i = \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i \right)$$

Objective function: $\min\limits_{\hat\beta_1, \hat\beta_2, \hat\beta_3} \sum_{i=1}^{N} \hat\varepsilon_i^2$

$$\left(\hat\beta_j = \underset{\beta_j}{\text{ARG MIN}} \sum_{i=1}^{N} \left(Y_i - \beta_1 - \beta_2 X_{2i} - \beta_3 X_{3i} \right)^2 \right) \quad j = 1,2,3$$

F.O.C.

1) $$\frac{\partial \sum_{i=1}^{N} \hat\varepsilon_i^2}{\partial \hat\beta_1} = 0$$

$$\Rightarrow -2 \sum_{i=1}^{N} \left(Y_i - \hat\beta_1 - \hat\beta_2 X_{2i} - \hat\beta_3 X_{3i} \right) = 0$$

$$\Rightarrow \sum_{i=1}^{N} \hat\varepsilon_i = 0 \iff \hat\varepsilon' 1_N = 0$$

As in the SLRM: EX POST residuals sum to zero !!

$$\sum_{i=1}^{N} Y_i - N\hat\beta_1 - \hat\beta_2 \sum_{i=1}^{N} X_{2i} - \hat\beta_3 \sum_{i=1}^{N} X_{3i} = 0$$

$$\hat\beta_1 = \frac{\sum_{i=1}^{N} Y_i}{N} - \hat\beta_2 \frac{\sum_{i=1}^{N} X_{2i}}{N} - \hat\beta_3 \frac{\sum_{i=1}^{N} X_{3i}}{N}$$

$$\hat\beta_1 = \bar Y - \hat\beta_2 \bar X_2 - \hat\beta_3 \bar X_3$$

Recap: In the SLRM $\Rightarrow \hat a = \bar Y - \hat b \bar X$

$\hat{\beta}_2$ and $\hat{\beta}_3$?

To simplify calculation, let us focus on y_i

2) $$\frac{\partial \sum_{i=1}^{n} \hat{\varepsilon}_i^2}{\partial \hat{\beta}_2} = 0 \Rightarrow -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i}) x_{2i} = 0$$

$$\Rightarrow \sum_{i=1}^{n} \hat{\varepsilon}_i x_{2i} = 0 \Leftrightarrow \hat{\varepsilon}' X_2 = 0$$

3) $$\frac{\partial \sum_{i=1}^{n} \hat{\varepsilon}_i^2}{\partial \hat{\beta}_3} = 0 \Rightarrow -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i}) x_{3i} = 0$$

$$\Rightarrow \sum_{i=1}^{n} \hat{\varepsilon}_i x_{3i} = 0 \Leftrightarrow \hat{\varepsilon}' X_3 = 0$$

$$\begin{cases} \sum_{i=1}^{n} \hat{\varepsilon}_i = 0 \\ \sum_{i=1}^{n} \hat{\varepsilon}_i x_{2i} = 0 \\ \sum_{i=1}^{n} \hat{\varepsilon}_i x_{3i} = 0 \end{cases}$$

Normal equations system

2) $$\sum_{i=1}^{n} y_i x_{2i} - \hat{\beta}_2 \sum_{i=1}^{n} x_{2i}^2 - \hat{\beta}_3 \sum_{i=1}^{n} x_{3i} x_{2i} = 0$$

3) $$\sum_{i=1}^{n} y_i x_{3i} - \hat{\beta}_2 \sum_{i=1}^{n} x_{2i} x_{3i} - \hat{\beta}_3 \sum_{i=1}^{n} x_{3i}^2 = 0$$

2) $$\hat{\beta}_2 \sum_{i=1}^{n} x_{2i}^2 = \sum_{i=1}^{n} y_i x_{2i} - \hat{\beta}_3 \sum_{i=1}^{n} x_{3i} x_{2i}$$

$$\Rightarrow \hat{\beta}_2 = \frac{\sum_{i=1}^{n} y_i x_{2i} - \hat{\beta}_3 \sum_{i=1}^{n} x_{3i} x_{2i}}{\sum_{i=1}^{n} x_{2i}^2}$$

V

$$\boxed{\hat{\beta}_2 = \frac{\sum_{i=1}^{N} y_i x_{2i} - \hat{\beta}_3 \sum_{i=1}^{N} x_{2i} x_{3i}}{\sum_{i=1}^{N} x_{2i}^2}} \quad *$$

We can put now the expression above of $\hat{\beta}_2$ in equation 3) $\sum_{i=1}^{N} y_i x_{3i} - \hat{\beta}_2 \sum_{i=1}^{N} x_{2i} x_{3i} - \hat{\beta}_3 \sum_{i=1}^{N} x_{3i}^2$

We get:

$$\sum_{i=1}^{N} y_i x_{3i} - \left(\frac{\sum_{i=1}^{N} y_i x_{2i} - \hat{\beta}_3 \sum_{i=1}^{N} x_{2i} x_{3i}}{\sum_{i=1}^{N} x_{2i}^2} \right) \sum_{i=1}^{N} x_{2i} x_{3i} - \hat{\beta}_3 \sum_{i=1}^{N} x_{3i}^2 = 0$$

Solving for $\hat{\beta}_3$ we get:

$$\hat{\beta}_3 = \frac{\sum_{i=1}^{N} y_i x_{3i} \sum_{i=1}^{N} x_{2i}^2 - \sum_{i=1}^{N} y_i x_{2i} \sum_{i=1}^{N} x_{2i} x_{3i}}{\sum_{i=1}^{N} x_{2i}^2 \sum_{i=1}^{N} x_{3i}^2 - \left(\sum_{i=1}^{N} x_{2i} x_{3i} \right)^2}$$

Substituting $\hat{\beta}_3$ in $*$

$$\hat{\beta}_2 = \frac{\underbrace{\sum_{i=1}^{N} y_i x_{2i}}_{Cov(Y, X_2)} \overbrace{\sum_{i=1}^{N} x_{3i}^2}^{Var(X_3)} - \overbrace{\sum_{i=1}^{N} y_i x_{3i}}^{Cov(Y, X_3)} \overbrace{\sum_{i=1}^{N} x_{2i} x_{3i}}^{Cov(X_2, X_3)}}{\underbrace{\sum_{i=1}^{N} x_{2i}^2}_{Var(X_2)} \underbrace{\sum_{i=1}^{N} x_{3i}^2}_{Var(X_3)} - \underbrace{\left(\sum_{i=1}^{N} x_{2i} x_{3i} \right)^2}_{(Cov(X_2, X_3))^2}}$$

So we can also write: (Just notation)

$$\hat{\beta}_2 = \frac{\hat{\sigma}_{12} \hat{\sigma}_{33} - \hat{\sigma}_{13} \hat{\sigma}_{23}}{\hat{\sigma}_{22} \hat{\sigma}_{33} - \hat{\sigma}_{23}^2}$$

$\hat{\sigma}_{12}$ = sample covariance of (Y, X_2)

$\hat{\sigma}_{22}$ = sample variance (X_2)

...
..

POPULATION PARAMETERS AND SAMPLE STATISTICS

1) $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$

TRUE POPULATION MODEL

$y_i = \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$

2) JOINT NORMALITY of Y_i, X_2, X_3

or alternatively of y, x_2, x_3

$$\begin{bmatrix} y_i \\ x_{2i} \\ x_{3i} \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & [\sigma_{12} & \sigma_{13}] \\ \sigma_{12} & \begin{bmatrix} \sigma_{22} & \sigma_{23} \\ \sigma_{23} & \sigma_{33} \end{bmatrix} \\ \sigma_{13} & \end{pmatrix}\right]$$

We also know that the conditional expectation $\mathbb{E}(y_i|x_{2i}, x_{3i})$:

$$\mathbb{E}(y_i|x_{2i}, x_{3i}) = \beta_2 x_{2i} + \beta_3 x_{3i}$$

Recap:

$$\mathbb{E}(Y) = \int Y f(Y)\, dy$$

$f(y)$: p.d.f.

joint distribution

$$\mathbb{E}(Y|x) = \int Y \frac{f(x, Y)}{f(x)}\, dy$$

→ marginal distribution of x

Where $f(x) = \int f(x, Y)\, dy$

V

Using the above definition of conditional expectation it can shown that:

$$\mathbb{E}(Y_t \mid X_{2t}, X_{3t}) = \underset{1\times 2}{\begin{pmatrix} \sigma_{12} & \sigma_{13} \end{pmatrix}} \underbrace{\underset{2\times 2}{\overbrace{\begin{pmatrix} \sigma_{22} & \sigma_{23} \\ \sigma_{23} & \sigma_{33} \end{pmatrix}}^{\Sigma}}^{-1} \underset{2\times 1}{\begin{pmatrix} X_{2t} \\ X_{3t} \end{pmatrix}}}_{1\times 1}$$

---

theorem:

the matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible only if $|A| \neq 0$.

In such a case: $$A^{-1} = \frac{1}{|A|} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

With $|A| = ad - cb$

So: $$\Sigma^{-1} = \frac{1}{\sigma_{22}\sigma_{33} - \sigma_{23}^2} \begin{pmatrix} \sigma_{33} & -\sigma_{23} \\ -\sigma_{23} & \sigma_{22} \end{pmatrix}$$

Since we have that:

$$\mathbb{E}(Y_t \mid X_{2t}, X_{3t}) = \beta_2 X_{2t} + \beta_3 X_{3t}$$

and also:

$$\mathbb{E}(Y_t \mid X_{2t}, X_{3t}) = \begin{pmatrix} \sigma_{12} & \sigma_{13} \end{pmatrix} \left(\frac{1}{\sigma_{22}\sigma_{33} - \sigma_{23}^2} \begin{pmatrix} \sigma_{33} & -\sigma_{23} \\ -\sigma_{23} & \sigma_{22} \end{pmatrix} \right) \begin{pmatrix} X_{2t} \\ X_{3t} \end{pmatrix}$$

So

$$\beta_2 X_{2t} + \beta_3 X_{3t} = \quad //$$

Finally, I can get the expressions of β_2 and β_3

V. Population PARAMETERS

$$\beta_2 = \frac{\sigma_{12}\sigma_{33} - \sigma_{13}\sigma_{23}}{\sigma_{23}\sigma_{33} - \sigma_{23}^2}$$

$$\beta_3 = \frac{\sigma_{13}\sigma_{22} - \sigma_{12}\sigma_{23}}{\sigma_{23}\sigma_{33} - \sigma_{23}^2}$$

β_2, β_3 : UNKNOWN POPULATION PARAMETERS
they are expressed as LINEAR COMBINATIONS of other unknown population parameters, variances and covariances – $\sigma_{12}, \sigma_{13}, \sigma_{23}, \sigma_{22}, \sigma_{33}$

- - - - - - - - - - - - - - - -

OLS ESTIMATORS

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{N} y_i x_{2i} \sum_{i=1}^{N} x_{3i}^2 - \sum_{i=1}^{N} y_i x_{3i} \sum_{i=1}^{N} x_{2i} x_{3i}}{\sum_{i=1}^{N} x_{2i}^2 \sum_{i=1}^{N} x_{3i}^2 - \left(\sum_{i=1}^{N} x_{2i} x_{3i}\right)^2}$$

$$\hat{\beta}_3 = \frac{\sum_{i=1}^{N} y_i x_{3i} \sum_{i=1}^{N} x_{2i}^2 - \sum_{i=1}^{N} y_i x_{2i} \sum_{i=1}^{N} x_{2i} x_{3i}}{\sum_{i=1}^{N} x_{2i}^2 \sum_{i=1}^{N} x_{3i} - \left(\sum_{i=1}^{N} x_{2i} x_{3i}\right)^2}$$

OLS estimators $\hat{\beta}_2, \hat{\beta}_3$ are expressed as LINEAR COMBINATIONS of SAMPLE STATISTICS – – variances and covariances – $\hat{\sigma}_{12}, \hat{\sigma}_{13}, \hat{\sigma}_{23}, \hat{\sigma}_{22}, \hat{\sigma}_{33}$

VI OMISSION OF A RELEVANT VARIABLE

Estimated model: $y_i = \beta_2^* x_{2i} + \varepsilon_i$

"True" model: $y_i = \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$

We know that the OLS estimator of β_2^* is

$$\hat{\beta}_2^* = \frac{\sum_{i=1}^{n} y_i x_{2i}}{\sum_{i=1}^{n} x_{2i}^2}$$

What about $\hat{\beta}_2^*$?

Let us substitute y_i with its "TRUE" expression

$y_i = \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$:

$$\hat{\beta}_2^* = \frac{\sum_{i=1}^{n} x_{2i} (\beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i)}{\sum_{i=1}^{n} x_{2i}^2}$$

$$\hat{\beta}_2^* = \beta_2 \cancel{\frac{\sum_{i=1}^{n} x_{2i}^2}{\sum_{i=1}^{n} x_{2i}^2}} + \beta_3 \frac{\sum_{i=1}^{n} x_{2i} x_{3i}}{\sum_{i=1}^{n} x_{2i}^2} + \frac{\sum_{i=1}^{n} x_{2i} \varepsilon_i}{\sum_{i=1}^{n} x_{2i}^2}$$

$$\mathbb{E}(\hat{\beta}_2^*) = \beta_2 + \beta_3 \frac{\sum_{i=1}^{n} x_{2i} x_{3i}}{\sum_{i=1}^{n} x_{2i}^2} + 0$$

↳ x_{2i}, x_{3i} are "FIXED VARIABLES" so:

$\mathbb{E}(x_{2i}) = x_{2i}$

$\mathbb{E}(x_{3i}) = x_{3i}$

$$\text{BIAS}(\hat{\beta}_2^*) = \mathbb{E}(\hat{\beta}_2^*) - \beta_2$$

$$= \beta_3 \frac{\sum_{i=1}^{n} x_{2i} x_{3i}}{\sum_{i=1}^{n} x_{2i}}$$

$$BIAS(\hat{\beta}_2^*) = \beta_3 \frac{\sum_{i=1}^{N} x_{2i} x_{3i}}{\sum_{i=1}^{N} x_{2i}^2}$$

What this expression is saying to us?

$$BIAS(\hat{\beta}_2^*) = f(\overset{+}{\beta_3}, \overset{+}{Cov(x_2,x_3)}, \overset{-}{Var(x_2)})$$

- If $\beta_3 = 0 \Rightarrow BIAS(\hat{\beta}_2^*) = 0$

  TRIVIAL: In such a case, there is not an omission of a relevant variable

  More the omitted variable is "relevant" (high β_3) and more the bias will be high

- If $Cov(x_2, x_3) = 0 \Rightarrow BIAS(\hat{\beta}_2^*) = 0$

  In such a case, x_{3i} "enters" ε_i

  And we know that if $\mathbb{E}(\varepsilon_i | x_{2i}) = 0$

  $\Rightarrow \hat{\beta}_2^*$ has no bias

  Generally it is not the case

- Again having a big variance of x_{2i} "help" the estimation; in this specific case, it helps to reduce the omitted variable bias

OMITTED VARIABLE BIAS

Example : wage equation

Estimated model $WAGE_i = \beta_1 + \beta_2^* EDUCATION_i + \varepsilon_i$

"TRUE" model : $WAGE_i = \beta_1 + \beta_2 EDUCATION_i + \beta_3 ABILITY_i + \varepsilon_i$

$$BIAS(\hat{\beta}_2^*) = \beta_3 \frac{\sum_{i=1}^{N} education_i \, ability_i}{\sum_{i=1}^{N} education_i^2}$$

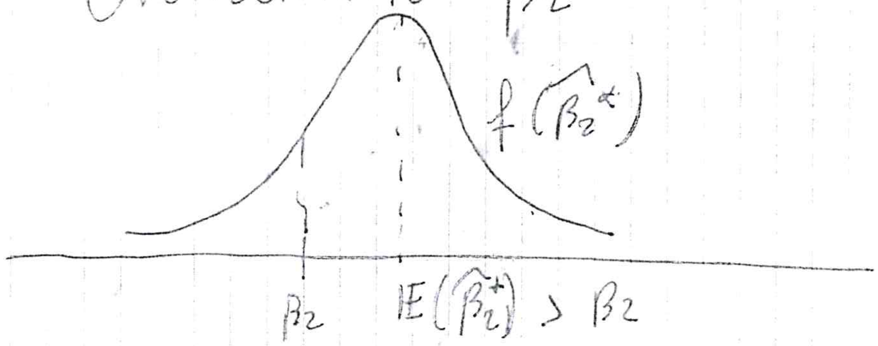
EX ANTE, what is the "expected" sign of $BIAS(\hat{\beta}_2^*)$?
think a bit....

I am expecting that:

- $\beta_3 > 0$
- $\sum_{i=1}^{N} education_i \, ability_i > 0$

$\Rightarrow$ $BIAS(\hat{\beta}_2^*) > 0$

OVERESTIMATE β_2



INCLUSION OF AN IRRELEVANT VARIABLE

ESTIMATED MODEL : $Y_t = \beta_2^* X_{2t} + \beta_3^* X_{3t} + \varepsilon_t$

"TRUE" MODEL : $Y_t = \beta_2 X_{2t} + \varepsilon_t$

We already know that:

$$\widehat{\beta}_2^* = \frac{\sum_{t=1}^{\tilde{N}} Y_t X_{2t} \sum_{t=1}^{\tilde{N}} X_{3t}^2 - \sum_{t=1}^{\tilde{N}} Y_t X_{3t} \sum_{t=1}^{\tilde{N}} X_{2t} X_{3t}}{\sum_{t=1}^{\tilde{N}} X_{2t}^2 \sum_{t=1}^{\tilde{N}} X_{3t}^2 - \left(\sum_{t=1}^{\tilde{N}} X_{2t} X_{3t}\right)^2}$$

Again, let us substitute Y_t with its true expression, now: $Y_t = \beta_2 X_{2t} + \varepsilon_t$

$$\widehat{\beta}_2^* = \frac{\sum_{t=1}^{\tilde{N}} X_{2t}(\beta_2 X_{2t} + \varepsilon_t) \sum_{t=1}^{\tilde{N}} X_{3t}^2 - \sum_{t=1}^{\tilde{N}} X_{2t} X_{3t} \sum_{t=1}^{\tilde{N}} X_{3t}(\beta_2 X_{2t} + \varepsilon_t)}{\sum_{t=1}^{\tilde{N}} X_{2t}^2 \sum_{t=1}^{\tilde{N}} X_{3t}^2 - \left(\sum_{t=1}^{\tilde{N}} X_{2t} X_{3t}\right)^2}$$

$$\widehat{\beta}_2^* = \frac{\beta_2\left[\sum_{t=1}^{\tilde{N}} X_{2t}^2 \sum_{t=1}^{\tilde{N}} X_{3t}^2 - \left(\sum_{t=1}^{\tilde{N}} X_{2t} X_{3t}\right)^2\right] + \sum_{t=1}^{\tilde{N}} X_{3t}^2 \sum_{t=1}^{\tilde{N}} X_{2t}\varepsilon_t - \sum_{t=1}^{\tilde{N}} X_{2t} X_{3t} \sum_{t=1}^{\tilde{N}} X_{3t}\varepsilon_t}{\sum_{t=1}^{\tilde{N}} X_{2t}^2 \sum_{t=1}^{\tilde{N}} X_{3t}^2 - \left(\sum_{t=1}^{\tilde{N}} X_{2t} X_{3t}\right)^2}$$

$$\widehat{\beta}_2^* = \beta_2 \cancel{\frac{\sum_{t=1}^{\tilde{N}} X_{2t}^2 \sum_{t=1}^{\tilde{N}} X_{3t}^2 - \left(\sum_{t=1}^{\tilde{N}} X_{2t} X_{3t}\right)^2}{\sum_{t=1}^{\tilde{N}} X_{2t}^2 \sum_{t=1}^{\tilde{N}} X_{3t}^2 - \left(\sum_{t=1}^{\tilde{N}} X_{2t} X_{3t}\right)^2}} + \frac{\sum_{t=1}^{\tilde{N}} X_{3t}^2 \sum_{t=1}^{\tilde{N}} X_{2t}\varepsilon_t - \sum_{t=1}^{\tilde{N}} X_{2t} X_{3t} \sum_{t=1}^{\tilde{N}} X_{3t}\varepsilon_t}{\sum_{t=1}^{\tilde{N}} X_{2t}^2 \sum_{t=1}^{\tilde{N}} X_{3t}^2 - \left(\sum_{t=1}^{\tilde{N}} X_{2t} X_{3t}\right)^2}$$

$\mathbb{E}\left(\widehat{\beta}_2^*\right) = \beta_2 + 0 \longrightarrow$ Simply because $\mathbb{E}(\varepsilon_t) = 0$

$\mathbb{E}\left(\widehat{\beta}_3^*\right) = 0$ Please show this result at home

VI

there is no bias from including an irrelevant variable, but there is still a PRICE to pay:

$$Var\left(\hat{\beta}_2^*\right) = \frac{\sigma^2}{\sum_{i=1}^{n} x_{2i}^2 (1-\hat{\rho}^2)}$$

with $\hat{\rho}^2 = \left(\frac{\sum_{i=1}^{n} x_{2i} x_{3i}}{\sqrt{\sum_{i=1}^{n} x_{2i}^2 \sum_{i=1}^{n} x_{3i}^2}} \right)^2$

While, as we know, when estimating the "true" regression model $Y_i = \beta_2 x_{2i} + \varepsilon_i$, we have:

$$Var\left(\hat{\beta}_2^*\right) = \frac{\sigma^2}{\sum_{i=1}^{n} x_{2i}^2}$$

☒ Including an irrelevant variable $\rightarrow$ LOOSE EFFICIENCY

$$Var\left(\hat{\beta}_2^*\right) \geq Var\left(\hat{\beta}_2\right)$$

$$Var\left(\hat{\beta}_2^*\right) = Var\left(\hat{\beta}_2\right) \text{ if } \rho = 0$$

☒ Including an irrelevant variable is a less important SPECIFICATION ERROR than omitting a relevant variable

↳ MAIN MESSAGE:

Melius est abundare quam deficere !!

VII MULTIPLE LINEAR REGRESSION MODEL (MLRM) 1

K-1 EXPLANATORY VARIABLES

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_K X_{Ki} + \varepsilon_i$$

$i = 1, \dots, N$

$\beta_1, \beta_2, \dots, \beta_K$: UNKNOWN PARAMETERS TO BE ESTIMATED

ε_i : RANDOM VARIABLE - NOT OBSERVABLE

Y_i : RANDOM VARIABLE - OBSERVABLE

$X_{2i}, \dots, X_{Ki}$: RANDOM - OR FIXED - VARIABLES - OBSERVABLE

REWRITE THE MODEL:

$i=1$: $Y_1 = \beta_1 + \beta_2 X_{21} + \dots + \beta_K X_{K1} + \varepsilon_1$

$i=2$: $Y_2 = \beta_1 + \beta_2 X_{22} + \quad + \beta_K X_{K2} + \varepsilon_2$

$\vdots$

$i=N$: $Y_N = \beta_1 + \beta_2 X_{2N} + \quad + \beta_K X_{KN} + \varepsilon_N$

$Y, X_2, \dots, X_K$ VARY ACROSS CROSS-SECTIONS

$\beta_1, \dots, \beta_K$ ARE THE SAME FOR ALL i

WHY?

- this "MAKES SENSE" ??
- SOLUTIONS : USE OF DUMMY VARIABLES
  PANEL DATA

VII REWRITE THE MODEL AGAIN

$$Y = \iota_N \beta_1 + X \beta_2 + \varepsilon$$

$$\underset{N\times 1}{\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}}_{Y}} = \underset{N\times 1}{\underbrace{\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}}_{\iota_N}} \underset{1\times 1}{\beta_1} + \underset{N\times (K-1)}{\underbrace{\begin{pmatrix} x_{21} & x_{K1} \\ x_{22} & x_{K2} \\ & \\ x_{2N} & x_{KN} \end{pmatrix}}_{X}} \underset{(K-1)\times 1}{\underbrace{\begin{pmatrix} \beta_2 \\ \beta_3 \\ \vdots \\ \beta_K \end{pmatrix}}_{\beta_2}} + \underset{N\times 1}{\underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \\ \varepsilon_N \end{pmatrix}}_{\varepsilon}}$$

ALTERNATIVELY:

$$Y = X\beta + \varepsilon$$

$$\underset{N\times 1}{\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}}_{Y}} = \underset{N\times K}{\underbrace{\begin{pmatrix} 1 & x_{21} & x_{K1} \\ 1 & x_{22} & x_{K2} \\ \vdots & \vdots & \vdots \\ 1 & x_{2N} & x_{KN} \end{pmatrix}}_{X}} \underset{K\times 1}{\underbrace{\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix}}_{\beta}} + \underset{N\times 1}{\underbrace{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \\ \varepsilon_N \end{pmatrix}}_{\varepsilon}}$$

HYPOTHESES

- $$\underset{N\times 1}{Y} = \underset{N\times K}{X}\,\underset{K\times 1}{\beta} + \underset{N\times 1}{\varepsilon}$$

- $$\underbrace{\mathbb{E}(\underset{N\times 1}{\varepsilon})}_{N\times 1} = \underset{N\times 1}{0} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \qquad \begin{matrix} \mathbb{E}(\varepsilon_1) = 0 \\ \mathbb{E}(\varepsilon_2) = 0 \\ \vdots \\ \mathbb{E}(\varepsilon_N) = 0 \end{matrix}$$

- $$\underbrace{\mathbb{E}(\underset{N\times 1}{\varepsilon | X})}_{N\times 1} = \underset{N\times 1}{0} = \begin{pmatrix} 0 \\ 0 \\ \\ 0 \end{pmatrix} \qquad \begin{matrix} \mathbb{E}(\varepsilon_1 | X) = 0 \\ \mathbb{E}(\varepsilon_2 | X) = 0 \\ \\ \mathbb{E}(\varepsilon_N | X) = 0 \end{matrix}$$

- $$\underset{N\times N}{\mathbb{E}(\underset{N\times 1}{\varepsilon}\,\underset{1\times N}{\varepsilon'})} = \underset{1\times 1}{\sigma^2}\,\underset{N\times N}{I}$$

I : IDENTITY MATRIX, $I = \begin{pmatrix} 1 & 0 & & 0 \\ 0 & 1 & & \\ & & 1 & \\ & & & 0 \\ 0 & & & 1 \end{pmatrix}$

$$\mathbb{E}(\varepsilon\varepsilon') = \sigma^2 I = \begin{pmatrix} \sigma^2 & & & 0 \\ 0 & \sigma^2 & & \\ & & \sigma^2 & \\ & & & 0 \\ 0 & & & \sigma^2 \end{pmatrix}$$

$\mathbb{E}(\varepsilon\varepsilon') =$ IS EQUIVALENT TO

$\mathbb{E}(\varepsilon_i)^2 = \sigma^2$ OMOSCHEDASTICITY

$\mathbb{E}(\varepsilon_i \varepsilon_j) = 0$ CROSS-SECTIONAL INDEPENDENCE

New Hypothesis

- RANK $(X) = K$ (full rank)
  (and $K < N$)

Recap. Rank (X): maximum number of (column) vectors linearly INDIPENDENT

2 vectors X_1 and X_2 are linearly independent if:

$$\nexists\ \lambda \in \mathbb{R} \text{ such that } X_1 = \lambda X_2$$

3 vectors X_1, X_2, X_3 are linearly independent if:

$$\nexists\ \lambda \in \mathbb{R} \text{ such that } X_1 = \lambda_1 X_2 + \lambda_2 X_3$$

Link with the SLRM

$$\underset{N\times1}{Y} = \underset{1\times1}{\beta_1} \underset{N\times1}{\iota_N} + \underset{1\times1}{\beta_2} \underset{N\times1}{X_2} + \underset{N\times1}{\varepsilon}$$

$$\nexists\ \lambda \in \mathbb{R} \text{ such that } X = \lambda \iota_N$$

(or equivalently : $\sum_{i=1}^{N} x_{2i}^2 \neq 0$ $\quad y_i = \beta_1 x_{2i} + \varepsilon_i$)

$$\underset{N\times1}{Y} = \underset{N\times2}{X} \underset{2\times1}{\beta} + \underset{N\times1}{\varepsilon}$$

$$X = \begin{pmatrix} 1 & X_{21} \\ 1 & X_{22} \\ \vdots & \\ 1 & X_{2N} \end{pmatrix}$$

Rank $(X) = 2$ if
$\nexists\ \lambda \in \mathbb{R}$ such that $X_2 = \lambda \iota_N$
if $X_2 = \lambda \iota_N \Rightarrow$ Rank $(X) = 1$

OLS

$$\underset{N\times 1}{Y} = \underset{N\times K}{X}\underset{K\times 1}{\beta} + \underset{N\times 1}{\varepsilon}$$

$$\hat{\varepsilon} = Y - X\hat{\beta}$$

OLS $\quad \text{MIN} \underset{1\times 1}{\underset{1\times N}{\hat{\varepsilon}'}\underset{N\times 1}{\hat{\varepsilon}}}$

DEFINE $\quad S(\hat{\beta}) = \hat{\varepsilon}'\hat{\varepsilon} = (Y - X\hat{\beta})'(Y - X\hat{\beta})$

1 2 3 4

$$S(\hat{\beta}) = \underset{1\times 1}{\underset{1\times N}{Y'}\underset{N\times 1}{Y}} - \underset{1\times 1}{\underset{1\times N}{\underset{1\times K}{\hat{\beta}'}\underset{K\times N}{X'}}\underset{N\times 1}{Y}} - \underset{1\times 1}{\underset{1\times K}{\underset{1\times N}{Y'}\underset{N\times K}{X}}\underset{K\times 1}{\hat{\beta}}} + \underset{1\times 1}{\underset{1\times N}{\underset{1\times K}{\hat{\beta}'}\underset{K\times N}{X'}}\underset{N\times 1}{\underset{N\times K}{X}\underset{K\times 1}{\hat{\beta}}}}$$

$$= \underset{1\times 1}{Y'Y} - 2\underset{1\times 1}{\underset{1\times N}{\underset{1\times K}{\hat{\beta}'}\underset{K\times N}{X'}}\underset{N\times 1}{X}} + \underset{1\times 1}{\hat{\beta}'X'X\beta}$$

Recap Derivatives of Matrices

Let X and X be two column vectors $N\times 1$

$\underset{1\times 1}{\underset{1\times N}{c'}\underset{N\times 1}{X}}$ is a scalar and $\dfrac{\partial \underset{1\times 1}{c'X}}{\partial \underset{N\times 1}{X}} = \underset{N\times 1}{c}$

also $\dfrac{\partial \underset{1\times 1}{c'X}}{\partial \underset{N\times 1}{c}} = \underset{N\times 1}{X}$

If A is a matrix

$$\frac{\partial AX}{\partial X} = A'$$

COME BACK TO OLS

$$\operatorname{Min} S(\hat{\beta})$$

with $S(\hat{\beta}) = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}$

$$\frac{\partial S(\hat{\beta})}{\partial \hat{\beta}} = 0$$

$$\frac{\partial S(\hat{\beta})}{\partial \hat{\beta}} = \frac{\partial (Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta})}{\partial \hat{\beta}}$$

$$= 0 - 2\frac{\partial \hat{\beta}'X'Y}{\partial \hat{\beta}} + \frac{\partial \hat{\beta}'X'X\hat{\beta}}{\partial \hat{\beta}}$$

$$= 0 - 2\underbrace{\underset{K\times N}{X'}\underset{N\times 1}{Y}}_{K\times 1} + 2\underbrace{\underbrace{\underset{K\times N}{X'}\underset{N\times K}{X}}_{K\times K}\underset{K\times 1}{\hat{\beta}}}_{K\times 1}$$

$$-2X'Y + 2X'X\hat{\beta} = 0$$

$$X'Y - X'X\hat{\beta} = 0$$

$$\hat{\beta} = \underbrace{\underbrace{(\underset{K\times N}{X'}\ \underset{N\times K}{X})^{-1}}_{K\times K}\ \underbrace{\underset{K\times N}{X'}\ \underset{N\times 1}{Y}}_{K\times 1}}_{K\times 1}$$

$\hat{\beta}$ is a $K\times 1$ vector: $\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_K \end{pmatrix}$

NOTE: $\underbrace{\underbrace{\underset{1\times K}{\hat{\beta}'}\ \underset{K\times N}{X'}}_{1\times N}\ \underbrace{\underset{N\times K}{X}\ \underset{K\times 1}{\hat{\beta}}}_{N\times 1}}_{1\times 1}$ IS A QUADRATIC FORM, AND BY DEFINITION IS A SCALAR.

THE $K\times 1$ VECTOR OF DERIVATIVES OF THIS TERM WITH RESPECT TO $\underset{K\times 1}{\hat{\beta}}$ IS EQUAL TO $2X'X\hat{\beta}$ (PROOF NOT PROVIDED HERE)

- $\hat{\beta}$ exists if $(X'X)$ can be inverted

$(X'X)$ can be inverted if $Rank(X) = K$

$Rank(X) = K \to X$ invertible

$Rank(X) = K \to (X'X)$ INVERTIBLE

- PROOF OF MINIMUM

If the MATRIX X HAS RANK K, the HESSIAN MATRIX:

$$\frac{\partial^2 S(\hat{\beta})}{\partial \hat{\beta} \partial \hat{\beta}'} = 2X'X$$

IS A POSITIVE DEFINITE MATRIX

$\underset{K\times N}{X'}\underset{N\times K}{X}$ is a $K \times K$ MATRIX

DEFINE THE QUADRATIC FORM

$$q = \underbrace{\underset{1\times K}{C'}\underset{K\times N}{X'}}_{1\times N}\underbrace{\underset{N\times K}{X}\underset{K\times 1}{C}}_{N\times 1}$$

C is a NON ZERO VECTOR $K\times 1$

If $q > 0 \Rightarrow X'X$ DEFINITE POSITIVE

DEFINE $\underset{N\times 1}{V} = \underset{N\times 1}{XC}$ (JUST NOTATION)

$$q = V'V$$

$\to q > 0$ in all cases exept when all the elements of $V = \begin{pmatrix} v_1 \\ v_2 \\ v_N \end{pmatrix}$ are <u>ZERO</u>

$\to$ this happens only if $Rank(X) < K$

$$\mathbb{E}(\hat{\beta}) = ?$$

$$\hat{\beta} = (X'X)^{-1}X'Y$$

$$\mathbb{E}(\hat{\beta}) = \mathbb{E}\left[(X'X)^{-1}X'(X\beta + \varepsilon)\right]$$

$$= \mathbb{E}\left[\cancel{(X'X)^{-1}}\,\cancel{X'X}\beta + (X'X)^{-1}X'\varepsilon\right]$$

$$= \mathbb{E}\left[\beta + (X'X)^{-1}X'\varepsilon\right]$$

$$= \beta + \mathbb{E}(X'X)^{-1}X'\varepsilon$$

$$\mathbb{E}(\varepsilon) = 0$$

$$\mathbb{E}(\hat{\beta}) = \beta$$

VII OLS 9

$$Var(\hat{\beta}) = ?$$

$$= \mathbb{E}\left[\left(\hat{\beta} - \mathbb{E}(\hat{\beta})\right)\left(\hat{\beta} - \mathbb{E}(\hat{\beta})\right)'\right]$$

$$= \mathbb{E}\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right]$$

$$= \mathbb{E}\left[\left((X'X)^{-1}X'\underbrace{(X\beta + \varepsilon)}_{Y} - \beta\right)\left((X'X)^{-1}X'(X\beta + \varepsilon) - \beta\right)'\right]$$

$$= \mathbb{E}\left[\left(\cancel{(X'X)^{-1}X'X}\beta + (X'X)^{-1}X'\varepsilon - \beta\right)\left(\cancel{(X'X)^{-1}X'X}\beta + (X'X)^{-1}X'\varepsilon - \beta\right)'\right]$$

$$= \mathbb{E}\left[\left(\cancel{\beta - \beta} + (X'X)^{-1}X'\varepsilon\right)\left(\cancel{\beta - \beta} + (X'X)^{-1}X'\varepsilon\right)'\right]$$

$$= \mathbb{E}\left[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}\right]$$

$$\mathbb{E}(\varepsilon\varepsilon') = \sigma^2 I$$

$$= \cancel{(X'X)^{-1}}\,\cancel{X'X}\,(X'X)^{-1}\sigma^2$$

$$= \sigma^2 (X'X)^{-1}$$

True $gdp_i = \beta_2 \, corruption_i + \beta_3 \, institutions_i + \varepsilon_i$

Estimated model case 1 $gdp_i = \beta_2^* \, corruption_i + \varepsilon_i$

$BIAS(\hat{\beta}_2^*)$?

Estimated model case 2 $gdp_i = \beta_3^* \, institutions_i + \varepsilon_i$

$BIAS(\hat{\beta}_3^*)$?

Expected signs : $\beta_2 < 0$, $\beta_3 > 0$

$\sum_{i=1}^{N} corruption_i \, institutions_i < 0$

CASE 1 $BIAS(\hat{\beta}_2^*) = \underset{>0}{\beta_3} \frac{\overset{<0}{\sum corruption_i \, institutions_i}}{\sum corruption_i^2}$

< 0

CASE 2 $BIAS(\hat{\beta}_3^*) = \underset{<0}{\beta_2} \frac{\overset{<0}{\sum corruption_i \, institutions_i}}{\sum institutions_i^2}$

$\rightarrow 0$

$y'y$

$y = X\beta + \varepsilon$

$y' = \beta' X' + \varepsilon'$

$(\beta' X' + \varepsilon')(X\beta + \varepsilon)$

$$\frac{\partial}{\partial \beta} \; \beta' X' X \beta + \beta' X' \varepsilon + \varepsilon' X \beta + \varepsilon' \varepsilon =$$

$2X'X\beta + X'\varepsilon + \varepsilon' X +$

RANDOM VARIABLE

X : RANDOM VARIABLE

x : REALIZATION OF X

X → DISCRETE
X → CONTINUOUS

X DISCRETE

X : x_1 x_2 x_3 ... x_m

Probability: p_1 p_2 p_3 -- p_m

PROBABILITY FUNCTION $\quad p_i = P(X = x_i)$

- $\sum_{i=1}^{m} p_i = 1$
- $p_i \geq 0 \quad \forall i$

IT IS a FUNCTION ASSOCIATING AT EACH VALUE OF X THE CORRESPONDING PROBABILITY

CUMULATIVE PROBABILITY FUNCTION
(FUNZIONE DI RIPARTIZIONE)

$$F(x) = P(X \leq x)$$

$$F(x) = \sum_{i : x_i \leq x}$$

EXPECTATION of a R.V.

$$\mathbb{E}(X) = \sum_{i=1}^{n} x_i p_i$$

Recall : "descriptive mean"

$$M(X) = \sum_{i=1}^{n} x_i w_i$$

SIMPLE ARITHMETIC MEAN: $w_i = \frac{1}{n}$

$\mathbb{E}(X)$ is the EXPECTED VALUE,
"expected" if the experiment
<u>will</u> <u>be</u> <u>performed</u> <u>many</u> <u>times</u>

VARIANCE of a R.V.

$$V(X) = \mathbb{E}\left[X - \mathbb{E}(X)\right]^2$$

$$= \sum_{i=1}^{n} \left[x_i - \mathbb{E}(X)\right]^2 p_i$$

Sometime useful to know:

$$V(X) = \mathbb{E}(X^2) - \left[\mathbb{E}(X)\right]^2$$

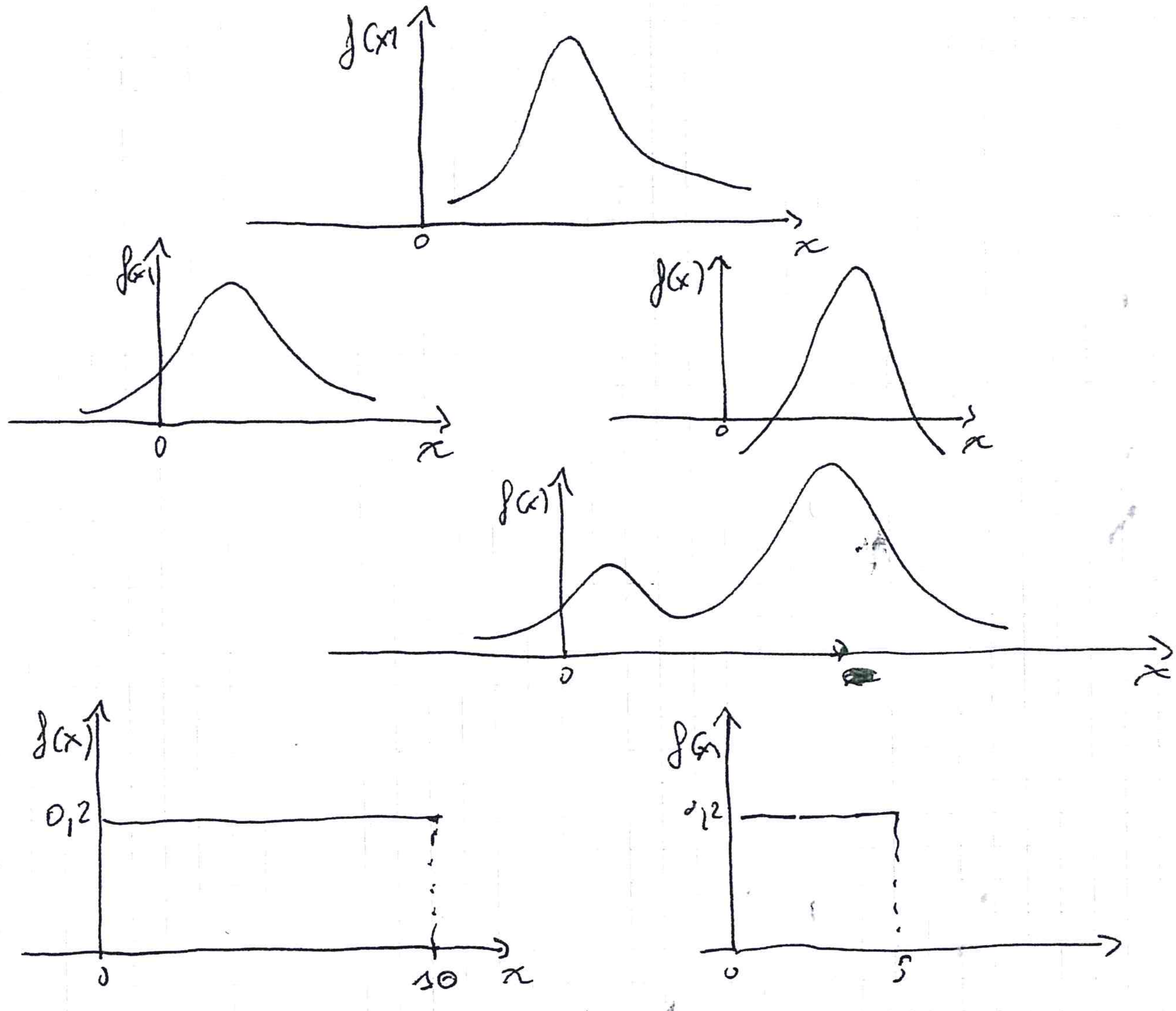
$$= \sum_{i=1}^{n} x_i^2 p_i - \left[\mathbb{E}(X)\right]^2$$

CONTINUOUS RANDOM VARIABLES

$f(x)$: (probability) density function (pdf)

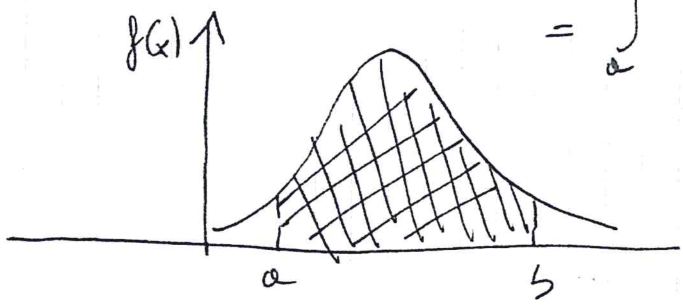
- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$

Question: the following functions would be (in theory) pdfs ??



$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)\,dx$$

$$P(a \leq X \leq b) = F(b) - F(a) = \int_{a}^{b} f(x)\,dx$$



$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x)\,dx$$

$$V(X) = \mathbb{E}[X - \mathbb{E}(X)]^2 = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - [\mathbb{E}(X)]^2$$

Recap: GAUSSIAN DISTRIBUTION

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]$$

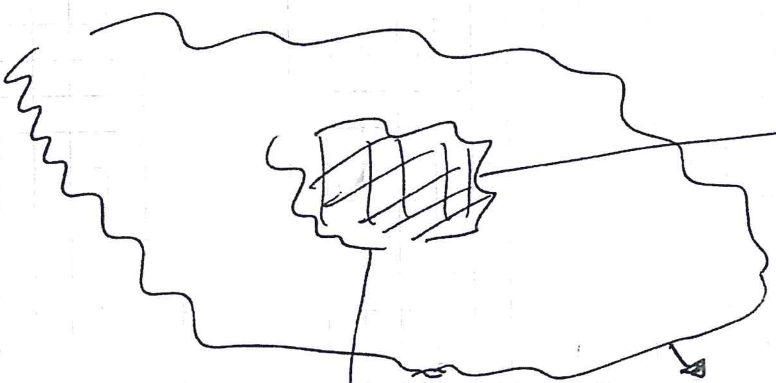
$\mathbb{E}(X) = \mu$

$V(X) = \sigma^2$

$$X \sim N(\mu, \sigma^2)$$

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

RANDOM SAMPLING AND CLT



RANDOM SAMPLE

POPULATION **UNKNOWN** !!

UNKNOWN PARAMETERS μ, σ^2

STATISTICS CAN BE CALCULATED FROM THE SAMPLE

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

CLT : IF THE POPULATION IS NORMAL (GAUSSIAN) **OR** if we have a **BIG SAMPLE** (even if the population is not Gaussian), we have:

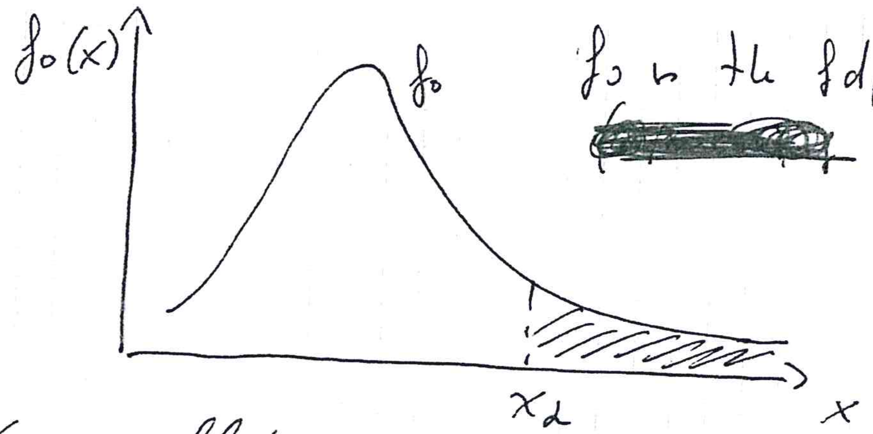
$$\bar{X} \sim N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$$

TEST OF HYPOTHESIS

$H_0 : \mu = c$

$H_1 : \mu > c$

Fix α, which is called SIGNIFICATIVITY LEVEL, ex: $0{,}05$



f_0 is the fdp under H_0

x_α is called CRITICAL VALUE. It is the value such that:

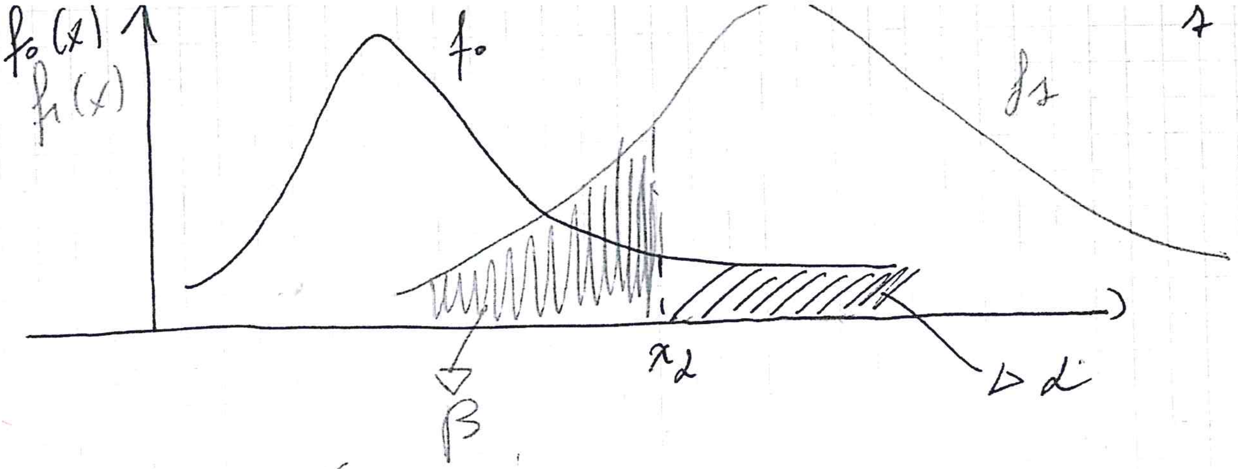
$$\int_{x_\alpha}^{\infty} f_0(x)\,dx = \alpha$$

$$= P(\text{reject } H_0 \mid H_0)$$

$$\int_{-\infty}^{x_\alpha} f(x)\,dx = (1-\alpha)$$

$$= P(\text{not reject } H_0 \mid H_0)$$

α error of first type



$$\int_{-\infty}^{x_\alpha} f_1(x)\,dx = \beta$$

$$= P(\text{not reject } H_0 \mid H_1)$$

$$\int_{x_\alpha}^{\infty} f_1(x)\,dx = 1-\beta$$

$$= P(\text{Reject } H_0 \mid H_1)$$

β : second type error

$(1-\beta)$: POWER OF A TEST

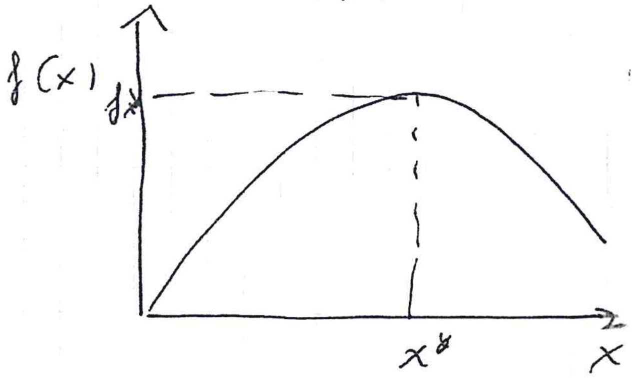
MAX / MIN

1 VARIABLE

$f: \mathbb{R} \rightarrow \mathbb{R}$

x^* : max / min

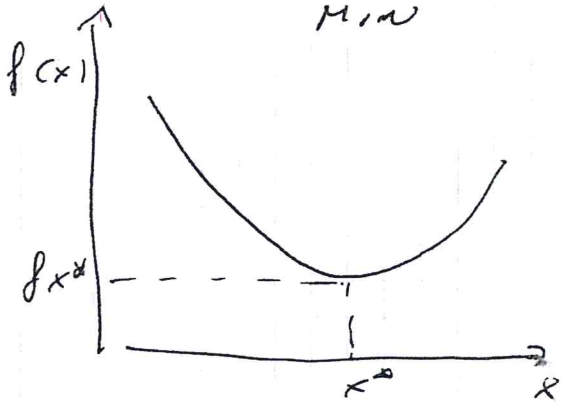
MAX



F.O.C. $f'(x) = 0$

S.O.C. $f''(x) < 0$

MIN



F.O.C. $f'(x) = 0$

S.O.C. $f''(x) > 0$

MULTIVARIATE CASE

2 VARIABLES

- $\max_{x_1, x_2} f(x_1, x_2)$
- $\min_{x_1, x_2} f(x_1, x_2)$

F.O.C. :

$$\frac{\partial f(x_1, x_2)}{\partial x_1} = 0$$

$$\frac{\partial f(x_1, x_2)}{\partial x_2} = 0$$

n Variables

GRADIENT VECTOR (vector of partial derivatives)

$$Df(X) = \left(\frac{\partial f(x_1, .. x_n)}{\partial x_1}, \ldots, \frac{\partial f(x_1, .. x_n)}{\partial x_n} \right)$$

F.O.C : $Df(X) = 0$

: $\frac{\partial f(x_1, .. x_n)}{\partial x_1} = 0, \ldots, \frac{\partial f(x_1, .. x_n)}{\partial x_n} = 0)$

S.O.C HESSIAN MATRIX (second derivatives)

2 variables case

$$f_{xy} = \frac{\partial^2 f}{\partial x_1 x_2}$$

$$H = \begin{pmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{pmatrix}$$

MAX : H definite NEGATIVE

MIN : H " POSITIVE

$$\underset{n \times n}{\overset{A}{(\quad)}} \quad \underset{n \times 1}{\overset{X}{(\,)}} \quad \underset{1 \times n}{\overset{X'}{(\quad)}}$$

QUADRATIC FORM

$$\underset{1 \times n}{X'} \; \underset{n \times n}{A} \; \underset{n \times 1}{X} = \underset{1 \times 1}{\text{scalar}}$$

$$\overset{X'}{(x_1 \; x_2)} \overset{A}{\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}} \overset{X}{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}$$

$$= a_{11} x_1 + (a_{12} + a_{21}) x_1 x_2 + a_{22} x_2^2$$

$X'AX > 0 \rightarrow$ A Definite POSITIVE

$X'AX \geq 0 \rightarrow$ A SEMI DEFINITE POSITIVE

$X'AX < 0 \rightarrow$ A DEFINITE NEGATIVE

$X'AX \leq 0 \rightarrow$ A SEMI DEFINITE NEGATIVE

A : MATRIX 2X2

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

A DEFINITE POSITIVE IF $\det A > 0$

$$\det(A) = a_{11} a_{22} - a_{21} a_{12}$$

In general, to determine if a symetric matrix is definite POSITIVE or NEGATIVE, we have to study EIGENVALUES

Given a symetric matrix $\underset{m \times m}{A}$,

Z ($m\times 1$ vector) and λ (scalar) are the EIGENVECTOR and EIGENVALUE of A if

$$\underset{\substack{m\times m \quad m\times 1 \\ m\times 1}}{A Z} = \underset{\substack{1\times 1 \quad m\times 1 \\ m\times 1}}{\lambda Z}$$

The Vector AZ is proportional to Z according to λ.

We search all the λs for which

$$AZ - \lambda Z = 0$$

$$(A - \lambda I) Z = 0$$

Theorem:

$\lambda_1, \dots, \lambda_m > 0 \rightarrow$ A definite POSITIVE

$\lambda_1, \dots, \lambda_m \geq 0 \rightarrow$ A SEMI definite POSITIVE

$\lambda_1, \dots, \lambda_m < 0 \rightarrow$ A definite NEGATIVE

$\lambda_1, \dots, \lambda_m \leq 0 \rightarrow$ A SEMI definite NEGATIVE