AGENDA – DAY 2 – 16-NOV-2025 (SUN)

- **REACP – DAY 1 + DOUBT CLEARING – MAX 10 MINUTES**

- **DAY 2**
  - **Supervised Learning – Regression (Contd…)**
  - **Linear Regression**
    - **Types – SLR, MLR**
  - **Overfitting & Underfitting**
  - **Non-linear Regression**
    - **Polynomial Regression**
    - **Model Performance Metrics**
  - **Regularisation**
    - **Lasso, Ridge, Elastic-Net**
  - **Model Optimisation**
  - **Hands-On Demo**
- **Q & A**
- **SUMMARY, HEADS-UP FOR DAY 3 & CLOSURE**

REACP – DAY 1 + DOUBT CLEARING – MAX 10 MINUTES

- AI ML DL intro , How they are different
- Set, Subset of AI MLs
- Regression , and different type and use
- Linear & Non linear Regression, ANOVA - Nature of X, Nature of Y
- Regression basics
- Null and Alternate hypothesis
- supervised semi supervised non supervised
- Corelation basics
- Model and use case
- Diff type of que-and corresponding model/algo - ML algos

1. WHAT IS THE DIFFERENCE BETWEEN REGRESSION & CORRELATION?
2. WHAT IS REGRESSION TRYING TO ACHIEVE?
3. WHAT DOES THE SLOPE MEAN IN SLR?
4. WHAT IS THE ROLE OF THE INTERCEPT?

"SATIRIS PARISUS"

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_i x_i + \epsilon \quad \text{MLR} \quad \frac{\partial y}{\partial x} \quad i = 1, \ldots, k$$

$\beta_0 = 10$ ⇒ BASELINE

$x_1 \rightarrow$ AGE

$Y \rightarrow$ SALES

HOUSE PRICE
- x_1 SQ FOOTAGE
- x_2 AGE
- x_3 LOCATION...

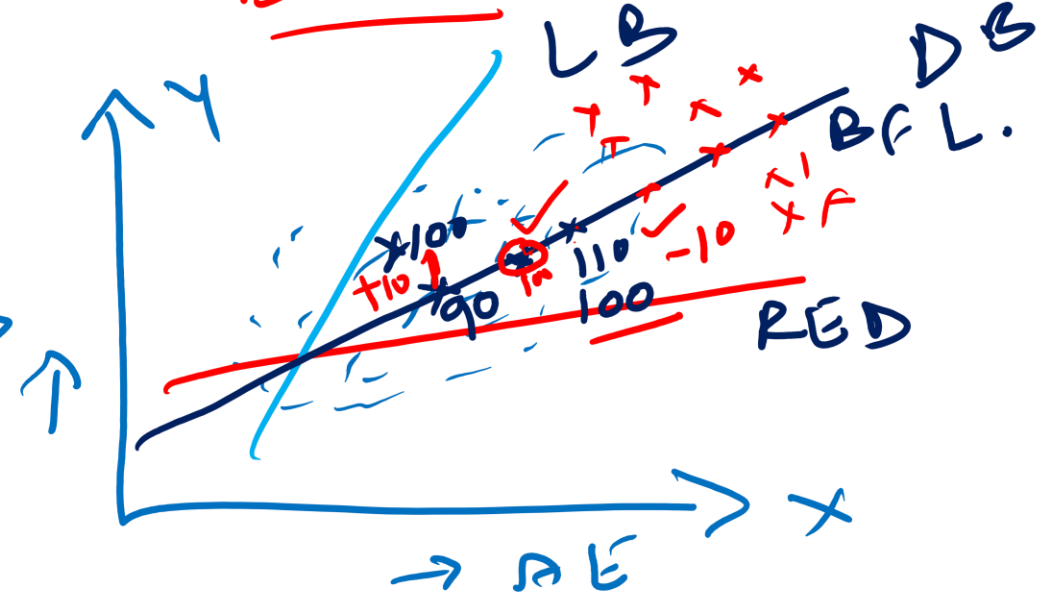
$HP = f(SF, A, L, \ldots)$

$HP = 0.2 + 0.3 * SF + 0.2 * A^2 + 0.25 * L + \epsilon$

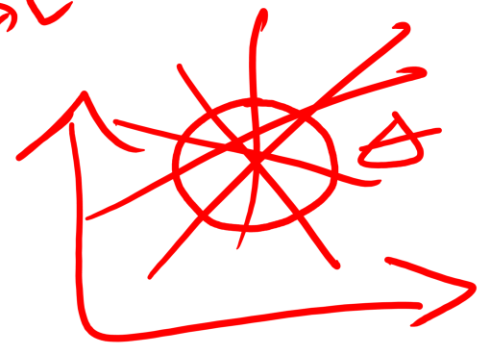
WHY REGRESSION INSTEAD OF JUST PLOTTING A LINE THRO' POINTS?

"BLUE"

S ACT	PS	AE →	
100	90	+10	100
100	110	−10	100
100	100	0	0
			SSE



→ BALOON OVAL SHAPE.



"IDIYOSYNCRATIC"

MODEL → DS

LEVERAGE VALUES.

$\leftarrow Y = \beta_0 + \beta_1 x_1 + \epsilon \rightarrow$ SLR

LINEAR ?

$x_1 \rightarrow$ VARIABLE

$\beta_1 \rightarrow$ PARAM / CO-EFF

$\leftarrow Y = \beta_0 + \beta_1 x_1^2 + \epsilon \rightarrow$ LINEAR / NON-LINEAR ?

$Y = \beta_0 + \beta_1 x_1 + \beta_2^2 x_2^2 + \beta_3^3 x^3 + \epsilon \rightarrow$ a^m $\log(a^m)$ $n \log a$

$Y = \beta_0 + \beta_1 e^{x^2/10}$ $Y = \beta_0 + \beta_1 e^{x^2} x_2 + \beta_3 + -$

$Y = \beta_0 + \beta_1^3 x$

$Y = \dfrac{\beta_1}{1 + e^{-\beta_1 x}}$

$\dfrac{a x^2 + b x + c}{a\Sigma + b\Sigma + c}$

$x^2 = 2$

NON-LINEAR IN PARAMETER.

β_1 MLR $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon$

$\beta_0, \beta_1, \cdots \beta_k \rightarrow$ CO-EFFI / BETA CO-EFFI

PARAMETERS OF REGRESSION.

β_1 $\times 0.3 + \beta$

β_1 $S = 0.25$

PARAMETER,

HYPER PARAMETER.

1, 2, 3, 4, 5

LIKERT

NATURE OF Y	NATURE OF X	MODEL
1. CONTINUOUS	ALL CATEGORICAL	~ DUMMY REGRN
2. CONTINUOUS	MIXED VARIABLES [NUM + CAT]	→ MULT REGRN WITH DUMMY
3. CONT (IDEAL CASE)	CONT	MULT REGRN

S = f(D, P, ...)

SS
1st
2nd

	G	G-M	G-F
1	M	1	0
2	F	0	1

RISK
H 0 3
M 1 2
L 2 1

L 1
M 2
H 3

VARIATION → MEASURE

TOTAL VARIATION ← SST = SSR + SSE

SST → TOTAL SUM OF SQUARES

SSR → EXPLAINED VAR → REGRN SUM OF SQUARES

SSE → UNEXPLAINED VAR → ERROR SUM OF SQUARES

$$\rightarrow SST = \sum (y_i - \bar{y})^2$$

$$\rightarrow SSR = \sum (\hat{y}_i - \bar{y})^2$$

$$\rightarrow SSE = \sum (y_i - \hat{y}_i)^2$$

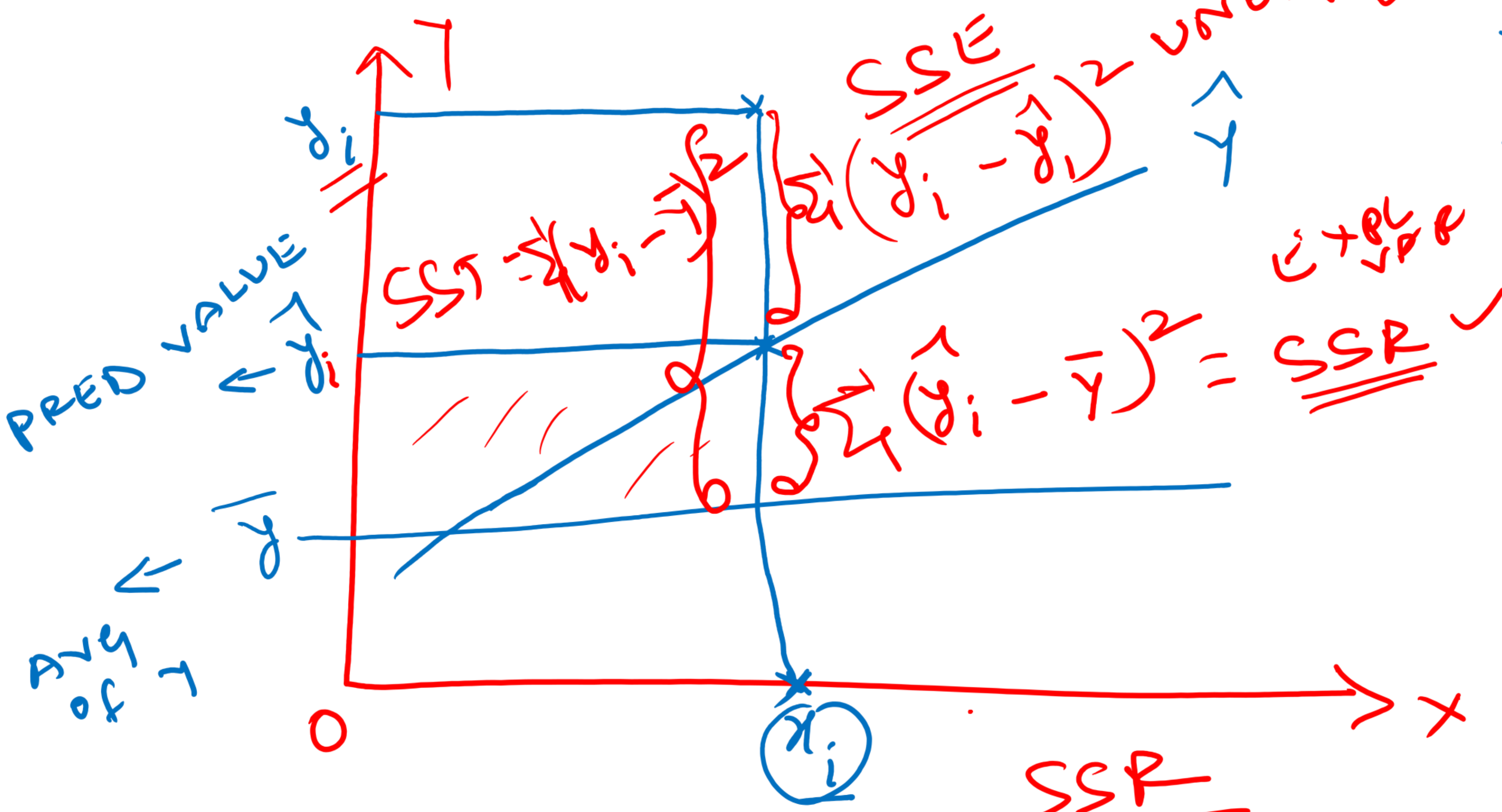
$\bar{y}$ → DV AVERAGE

y_i → OBS VAL of DV

$\hat{y}_i$ → PRED VALUE of 'y' for x_i value

A	P	E
100	80	+20
101	85	+16
99		

100 → 80



X	y (ACT)	$\hat{y}$ (PRED)
	5	
	16	

Avg = 105.5

COEFF OF DETERMIN $\Leftarrow$ $r^2 = \frac{SSR}{SST}$

$0 \le r^2 \le 1$

100%

$S = f(PY, D ...)$

75% 25% PRE OF

poc 300 mg
295 305

$H_0: \mu_1 \leq \mu_2$

60 ✓

50

H_1 or $H_a: \mu_1 > \mu_2$ ✓

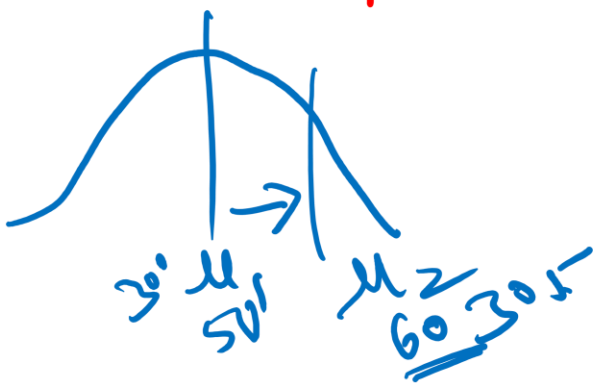
50

40

1

H_0:

H_1:



STATUS QUO

"EQUALITY" →

INEQUALITY

$\mu_1 = \mu_2$

2

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

2 TAILED

3

$H_0: \mu_1 \leq \mu_2$

$H_1: \mu_1 > \mu_2$

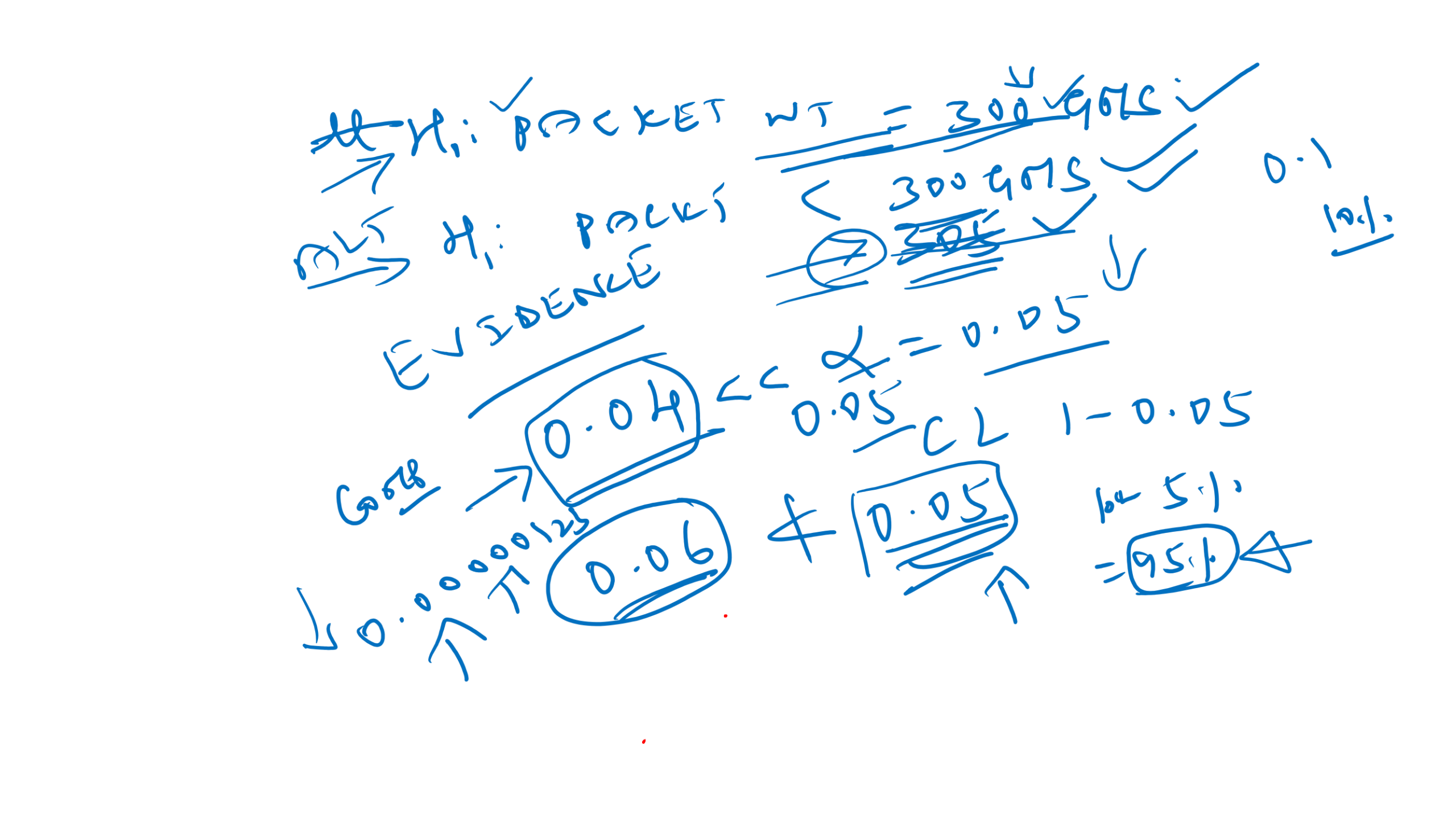
RT-TAILED TEST

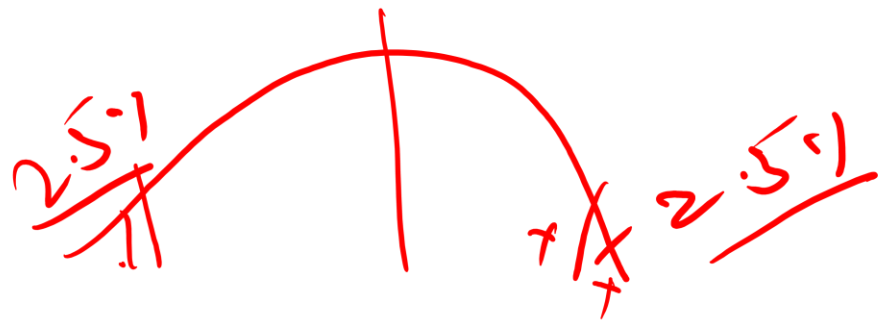
3

$H_0: \mu_1 \geq \mu_2$

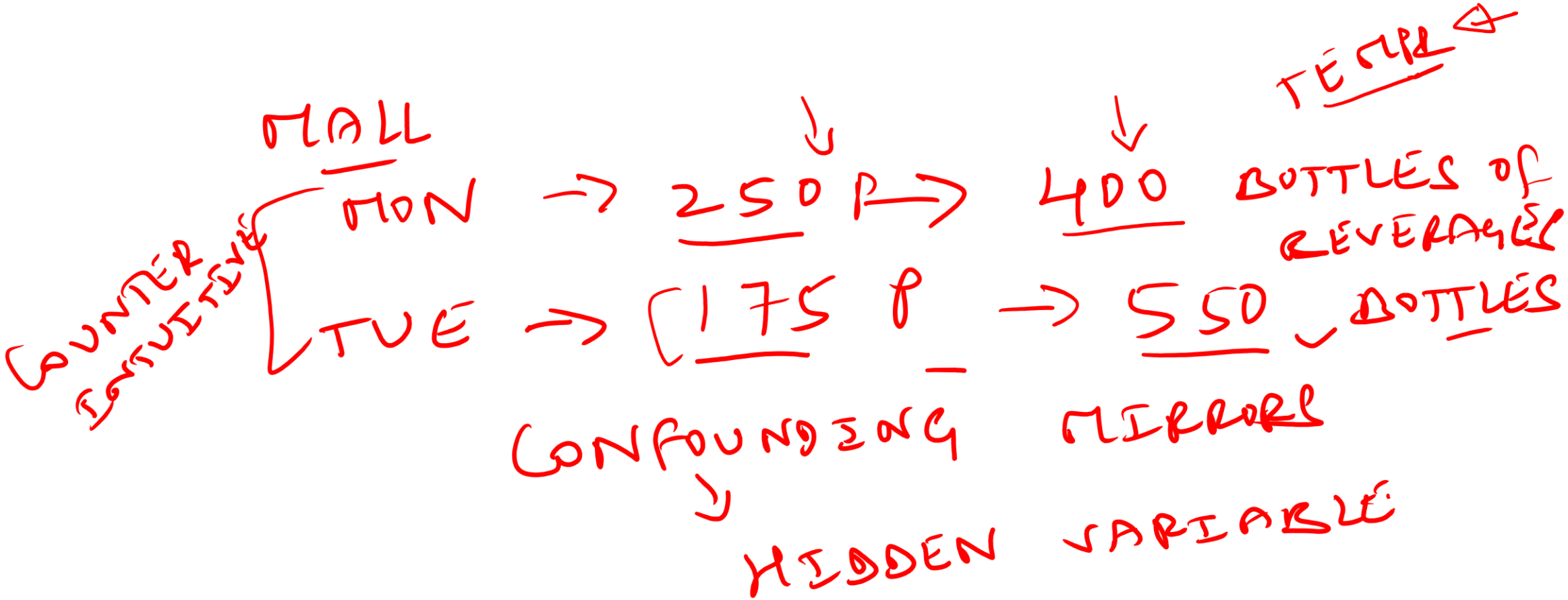
$H_1: \mu_1 < \mu_2$

LF LOWE


H1: POCKET WT = 300 GMS
H1: POCKET WT < 300 GMS
EVIDENCE
0.04
0.06
0.00000123
α = 0.05
0.05
CL 1 - 0.05
0.05
95%



$$\frac{0.04}{2} = 0.02$$


MALL
MON → 250 P → 400 BOTTLES OF BEVERAGES
TUE → 175 P → 550 BOTTLES
COUNTER INTUITIVE
TEMP
CONFOUNDING
HIDDEN VARIABLE
MIRROR