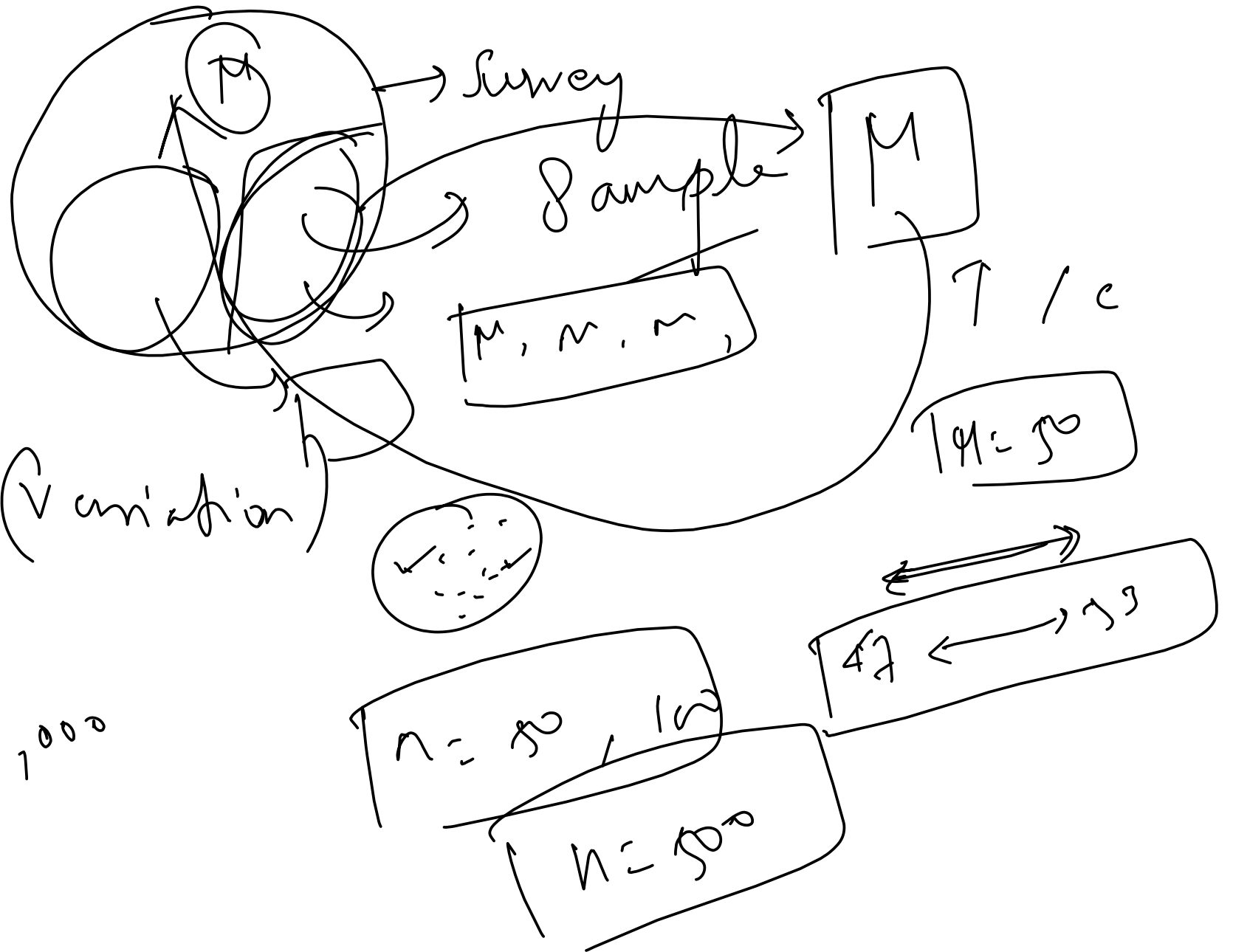


1. Sampling
2. Confidence Interval
3. Margin of Error
4. Hypothesis testing

Sampling:



Sampling error:

(Variation)

Confidence Interval:

- Express the range to predict / estimate the population parameter

Factors affecting the C.I.?

① \hookrightarrow Variation (population)

② \hookrightarrow The sample size.

C.D: \rightarrow
4

$$\bar{x} \pm t \frac{s}{\sqrt{n}}$$

$$95\%$$

$$n = 15$$

$$\bar{x} \rightarrow 149.3$$

$$std \rightarrow 4.75$$

$$\frac{s}{\sqrt{n}} \Rightarrow \text{Standard error}$$

$$df \Rightarrow \text{sample size} - 1$$

$$t = 2.145$$

$$\rightarrow 149.3 \pm$$

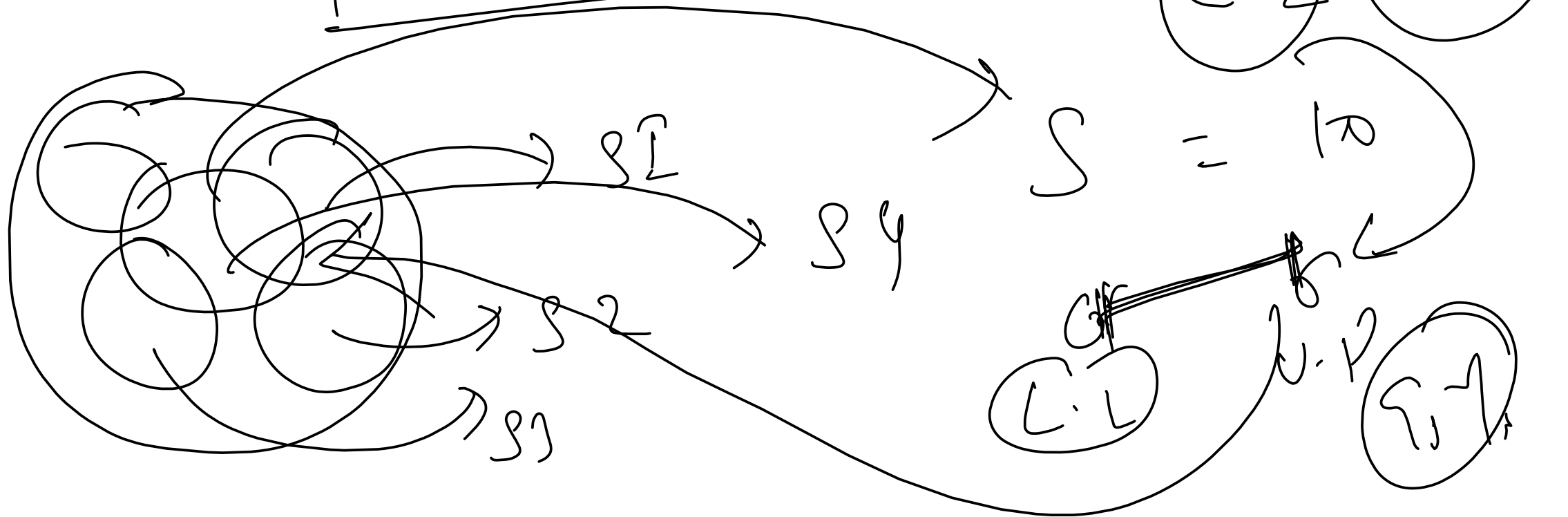
$$\rightarrow 149.3 \pm 2.631$$

$$2.145 * \left(\frac{4.75}{\sqrt{n}} \right)$$

$$\Rightarrow 146.7 - 151.9$$

$$t * \frac{std}{\sqrt{n}} \Rightarrow \text{Margin of Error}$$

$$148.2 - 151.9$$



$$n = 1$$

10

$$n = 145 - 155$$

$M = 150$

1950

201

2007

$$n = 2$$

$$145 - 152$$

2

$$n = 3$$

$$144 - 151$$

3

17.5

90

9

prime
number

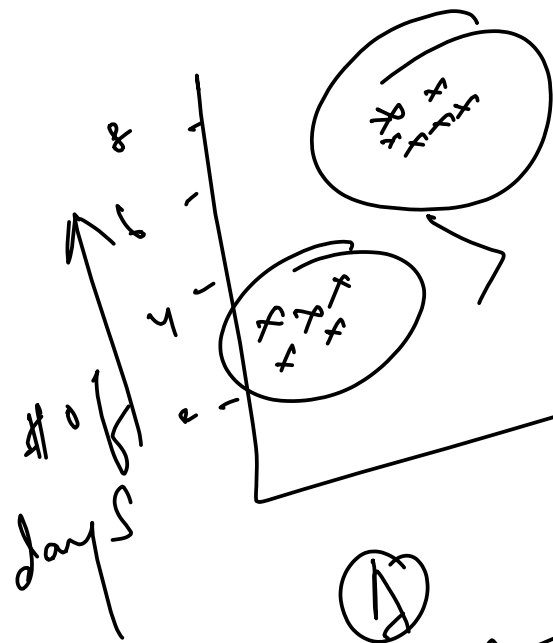
130

160

Hypothesis Testing:-

Drug (A)
Advil

$n=10$

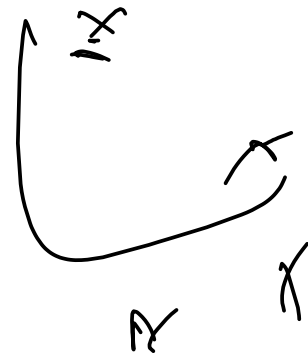
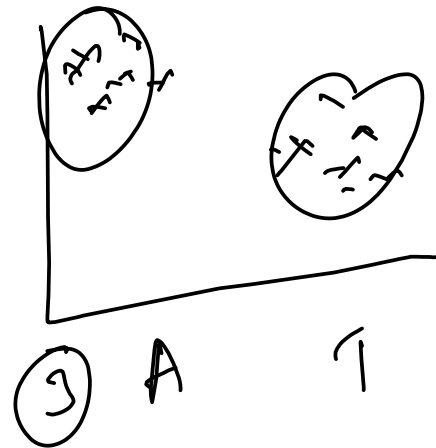
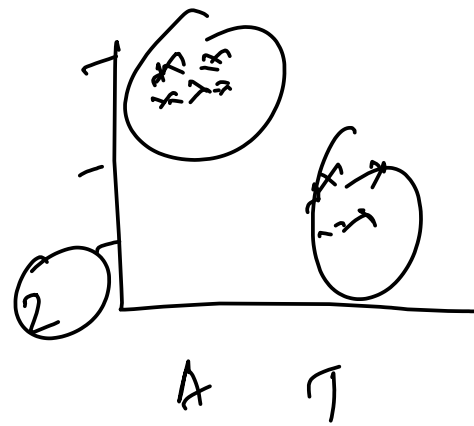


①

②

Null hypothesis

Drug (T)
Tylenol



Accept if : (M_0) is close enough to the true value (M)

Reject if : " too far from the true value (M)

Drug A (100)

Cured	Not
97	3

Cured	Not
73	25

37%

Drug B

Cured	Not
3	97

Cured	Not
59	131

31%

Summary statistics:

L.R

	conf	p-val	std
Var 1			
Var 2			
Var 3			

$\alpha = 0.05$

Null hypothesis:

$A \equiv B$

~~$A \equiv B$~~

$0 < P < 1$

$P = 0.01$

0.05

0.1 ~~0.025~~ ~~0.01~~ more confident

\Rightarrow Closer the p-value to zero, both drugs are different

$\alpha = 0.05$

~~\times~~

5%

$\alpha \rightarrow$ significance level

$P = 0.05$

4-7

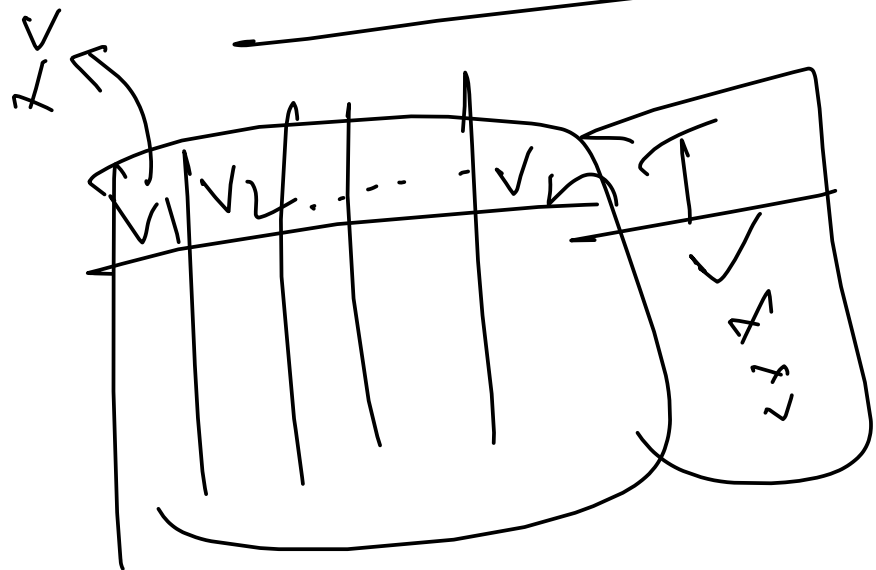
One tail test \Rightarrow

H_0

$\hookrightarrow f/a \gg H_a$

$H_a < H_0$

Two tail test \Rightarrow



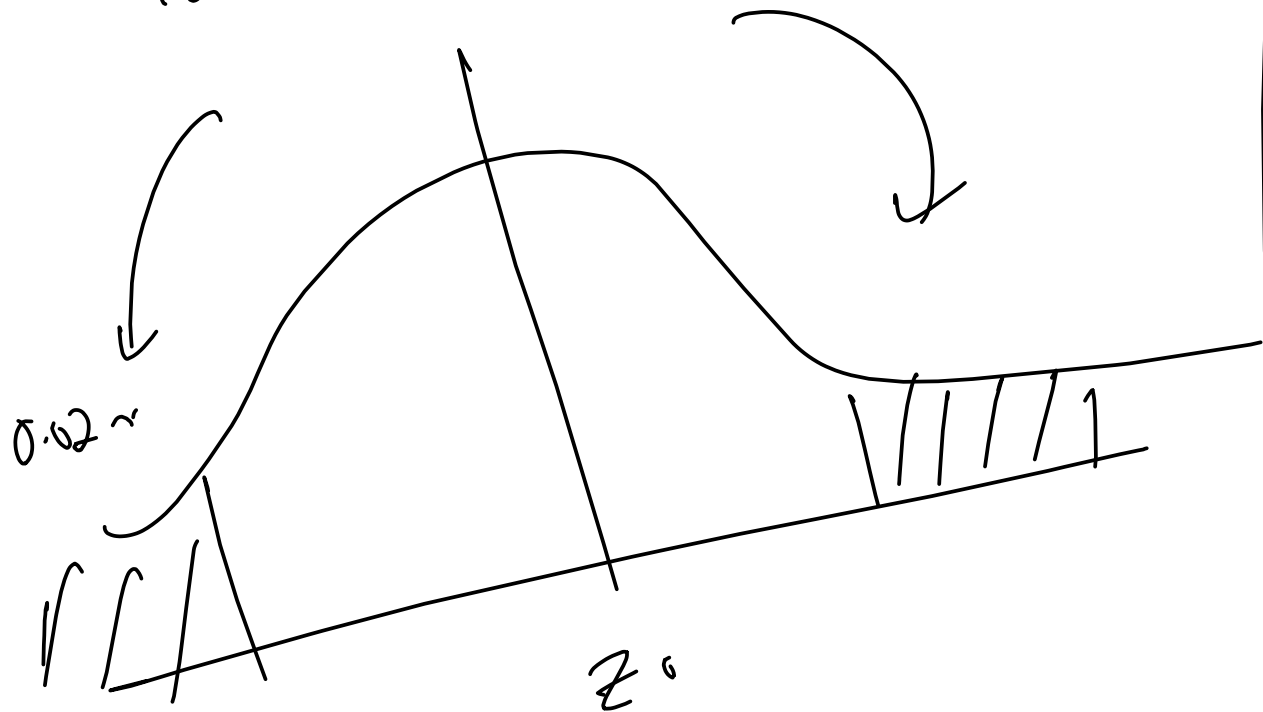
	$H_0 = H_a$
V_1	\times
V_2	\times
V_3	\times
	$H_0 \neq H_a$

(or)

p-val

99%

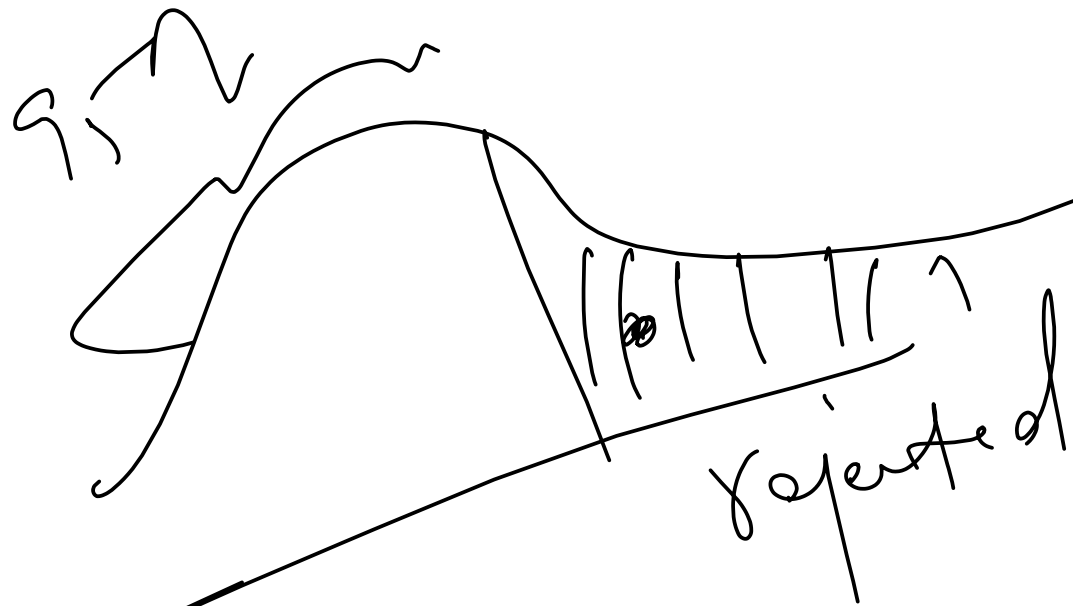
$95 \rightarrow 0.025$



$\alpha = 0.05$
 $\rightarrow 5\%$

$H_0 > H_a$

$H_0 < H_a$ 95%



$p < 0.05$; we have an evidence of an effect
(b/w Groups)

$p > 0.05$; No evidence of an effect

Chocolate:

200g \rightarrow 70g of nuts

$H_0 \Rightarrow 70g$;

$H_a < 70g$

$\alpha = 0.05$

Avg weight of nuts (μ) = 68.7g

$p = 0.18$

\Rightarrow the prob of getting a mean

IF

68.7 (or) less is 18%
the mean wt of peanuts is 70gm (or) more

99%

Summary statistics

50-70

157

Coeff

p-value

t-val

std
err

R^2

Var 1

Var 2

Var 3

Var 4

0.00025

0.01

0.25

0.05

0.10

a

$R^2 = 0.75$

0.75

$0 < R^2 < 1$

Significance level

$\alpha = 0.05$

0.1
0.05

We can't accept a variable if it is same

$H_0 \Rightarrow$ Every variable is the same ~~XXXX~~

$H_a \Rightarrow$ Not same ✓

$p \rightarrow \text{Var 1} \rightarrow 0.00025$

$p \rightarrow \text{Var} \rightarrow 0.25$ Drop

$\alpha = 0.05$
↑

Pre-processing:

① $[5] \rightarrow \text{scalar}$

② $[5, 7, 8, 9, 10, \dots] \rightarrow \text{Vector}$

③ $\begin{bmatrix} 1 & 5 \\ 6 & 9 \end{bmatrix} \rightarrow \text{Matrix}$

④ $\left[\begin{bmatrix} 1 & 5 \\ 2 & 6 \end{bmatrix} \quad \begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix} \right] \Rightarrow \text{Tensor}$
 $\left[\begin{bmatrix} 2 & 4 \\ 3 & 8 \end{bmatrix} \quad \begin{bmatrix} 2 & 4 \\ 3 & 7 \end{bmatrix} \right]$

$\text{pd} \rightarrow \text{np}$ ✓

07/20

np.var

$[1, 5, \dots]$

$\text{np.var}()$

matmul

D.L

Tf

Data cleaning

3 types:

→ Numeric

(eg: income, age)

→ Categorical

(eg: gender, nationality)

→ Ordinal

(low / medium / high)

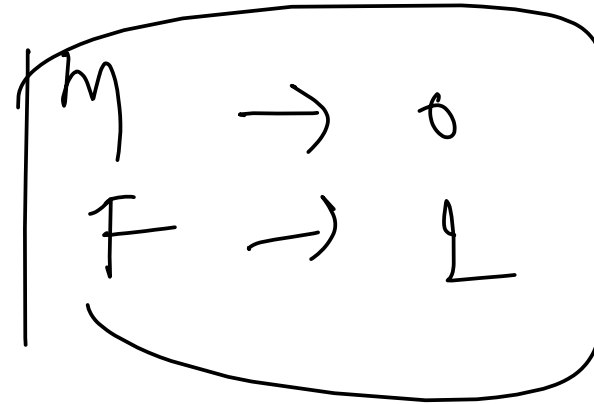
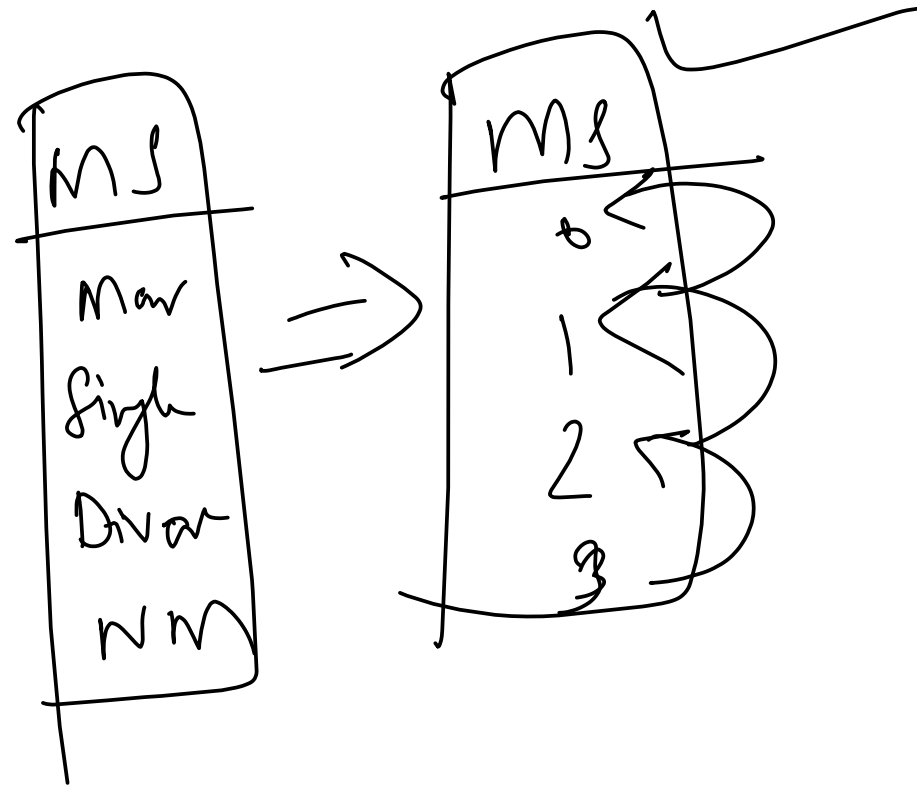
eg: ~~Online~~ F Inertia

One-hot encoding:

Sex $\begin{cases} \rightarrow \text{Male} \\ \rightarrow \text{Female} \end{cases}$

Sex							
0	male	1	0	0	0	0	0
1	f	0	1	0	0	0	0
2	m	1	0	0	0	0	0
3	f	0	1	0	0	0	0
4	f	0	1	0	0	0	0
5	m	1	0	0	0	0	0

Label Encoding:



relationship

Single ✓

married ✓

Date

	rel
6	Single
1	married
2	Single
3	Single
4	married

Date



Single	married
1	0
0	1
1	0
0	0

Hypothesis Testing:

- Define Null hypothesis

- Sample

→ Define (α)

→ Calculate p-value ✓

→ Finalize decision

→ Quick Recap.

Accept H_0
Reject H_0