

Feature Engineering

Friday, 19 November 2021 8:00 PM

Hypothesis Testing:

Rs 110

Rs 99

Buy Now → Yellow color
→ Red color
(1%)

Credit Card

VISA : S1
MasterCard :

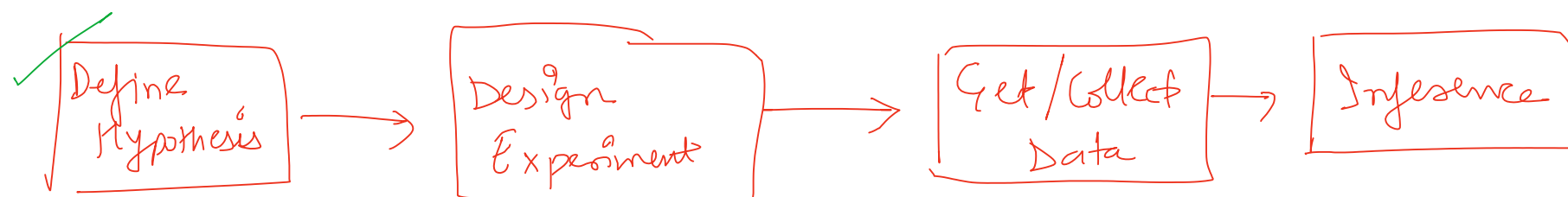
Claim

Hypothesis : Customers having VISA Card spend more than other

[Null Hypothesis : H_0 : No difference b/w ^{customers} VISA & ^{customers} MasterCard. [Assumption]]

✓ Alternate Hypo : $H_1 (H_a)$: Customers VISA spend more than mastercard.

Test - Statistic : $X = \mu_v - \mu_m$



S1 : Customers Spending VISA ($n=100$) 1M
 S2 : Customer Spending MasterCard ($n=100$)

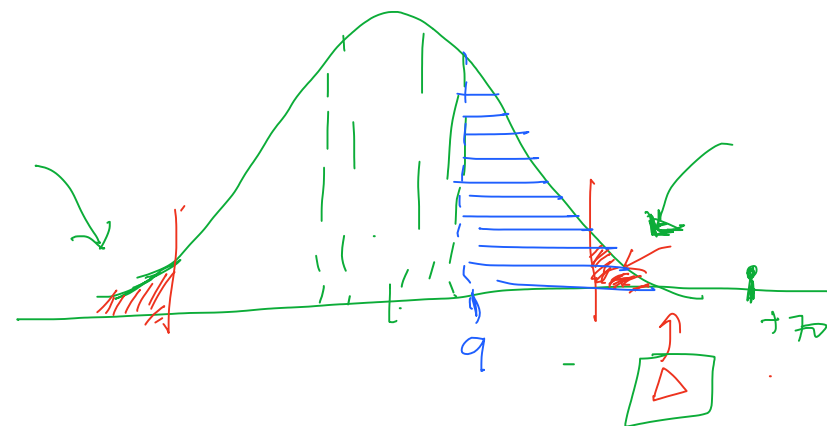
S1 : [500, 1100, 729, ...,]₁₀₀ $\mu_V = 559$ 650
 S2 : [3000, 199, ...,]₁₀₀ $\mu_M = 550$
 500

Observation = $559 - 550 \Rightarrow 9$

$\begin{cases} \mu_V = 552 \\ \mu_M = 555 \end{cases}$ $\begin{matrix} 562 \\ 551 \end{matrix}$

test-statistic

$\boxed{\bar{x}} = -20, \dots, +50$
 $\mu_V - \mu_M$



0.00005

$P\text{-value} = P(\bar{x} \geq \Delta \mid H_0) \Rightarrow$

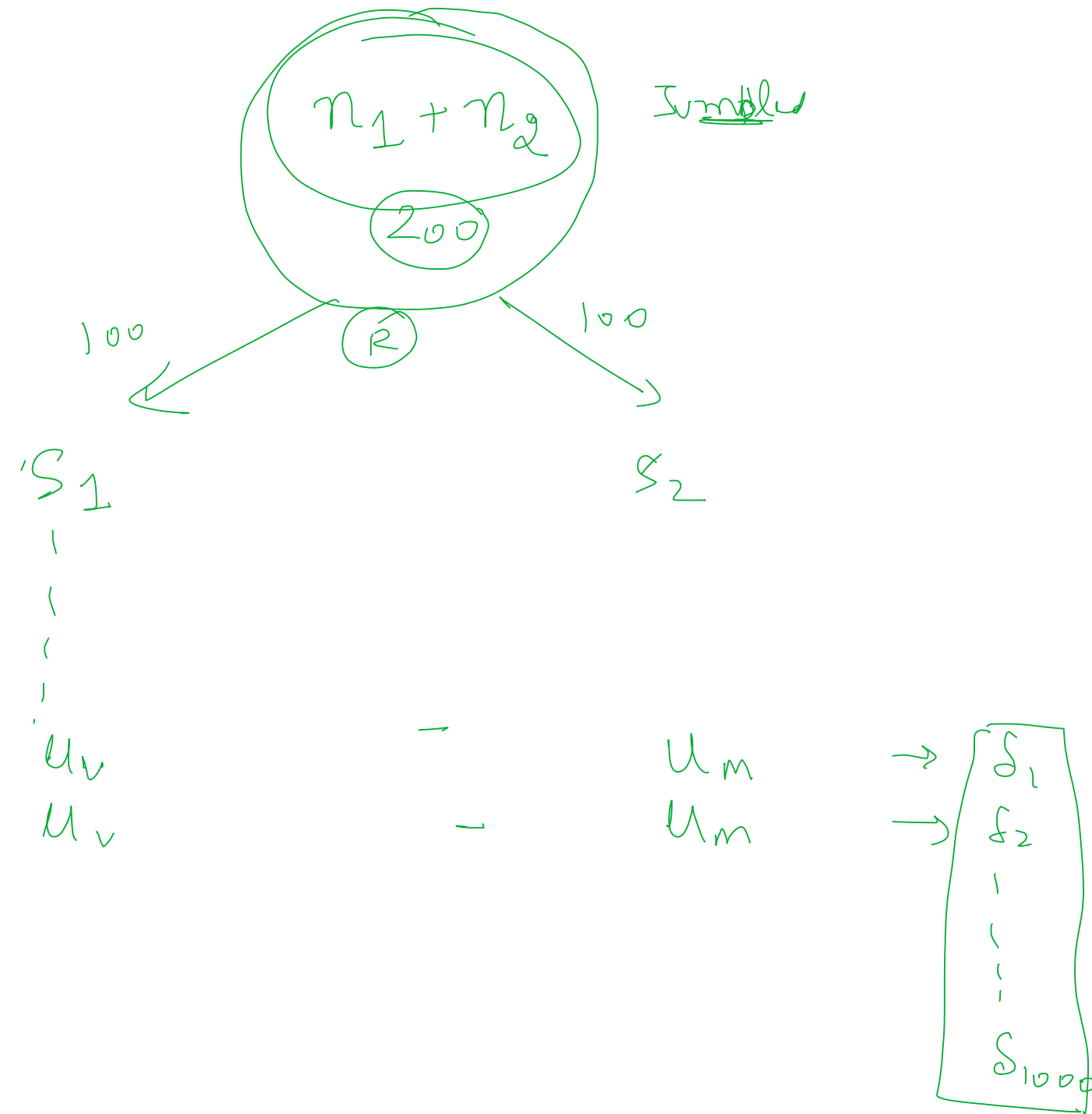
$P\text{-value} < 5\% \rightarrow \text{Reject } H_0$
 Accept H_a

$P\text{-value} \geq 5\% \rightarrow \text{fail to reject } H_0$

$\alpha = 5\%$

Significance level

Permutation Test



$\delta_1, \delta_2, \delta_3, \dots, \delta_{1000}$
 \downarrow sort
 $\delta'_1, \delta'_2, \delta'_3, \dots, \delta'_{1000}$
 Δ (difference)
 $\alpha\%$ (significance level)
 δ'_{1000} (critical value)

$\checkmark \rightarrow \text{cal}$

$$p\text{-value} = P(X \geq \Delta | H_0) \rightarrow \underline{2\%}$$

$\alpha \rightarrow 1\%$
medical application

$$\cancel{P\text{-value}} = \cancel{P(H_0)}$$

Feature Binning (Bins)

Income	Income-bins
1500	$[0, 2500] \rightarrow \text{Low}$
7500	$[2500, 4000] \rightarrow \text{Avg.}$
55000	$[4000, 6000] \rightarrow \text{High}$
49000	$[6000, 81000] \rightarrow \text{V. high}$
80000	

Numeric \rightarrow Categorical

Low	Poor
Average	Middle
High	Rich
V. high	

Generic f.o.E

①

Arithmetic feature \rightarrow

x_1 & x_2

$$x_1 + x_2 \rightarrow x_1 * x_2, x_1 / x_2$$

$$x_1^2, x_2^2, x_1^2 + x_2^2$$

V. Imp.

(2) log / exponential \rightarrow $\log(x_1)$, $\log(x_2)$

(3) Trigonometry \rightarrow $\sin(x_1)$, $\cos(x_1)$

(4) Feature Binning