

## Decision Tree

CHAID - Classification ( $y \rightarrow$  categorical)

↳ Chi-Squared

CART - Classification and Regression  
( $y \rightarrow$  categorical)      ( $y \rightarrow$  continuous)

Chi-Square Test

Obs

	Mkt	HR	Fin	Prod.	
M	6	2	3	10	21
F	4	5	1	3	13
	10	7	4	13	34

Claim

- Which dept. you  
work in &  
which gender you  
have has a  
relationship

$$P(F) = 13/34$$

$$P(Prod) = 13/34$$

If being female and being in prod.  
dept is not correlated with each  
 other independent

$$P(F \text{ and Prod}) = P(F \cap \text{Prod})$$

$$= P(F) \cdot P(\text{Prod})$$

$$= \frac{13}{34} \times \frac{13}{34}$$

$$N(F \text{ and Prod}) = P(F \text{ and Prod}) \times N$$

$$= \frac{13}{34} \times \frac{13}{\cancel{34}} \times \cancel{34}$$

Assumption  
 is that gender  
 & dept has no  
 relationship

$$= \frac{13 \times 13}{34}$$

$$\text{Expected} = \frac{RT \times CT}{GT}$$

(ΣRP)

	Mkt	HR	Fin	Prod
M		$\frac{21 \times 7}{34}$	$\frac{21 \times 4}{34}$	$\frac{21 \times 13}{34}$
F				$\frac{13 \times 13}{34}$

(Obs)

	Mkt	HR	Fin	Prod	
M	6	2	3	10	21 ✓
F	4	5	1	3	13
	10	7	4	13	34

When the observed table is close to the expected table we can conclude that there is no relationship.

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

$\chi^2 \uparrow$  the chance that

they are same (p)  $\downarrow$  (P < 0.05)

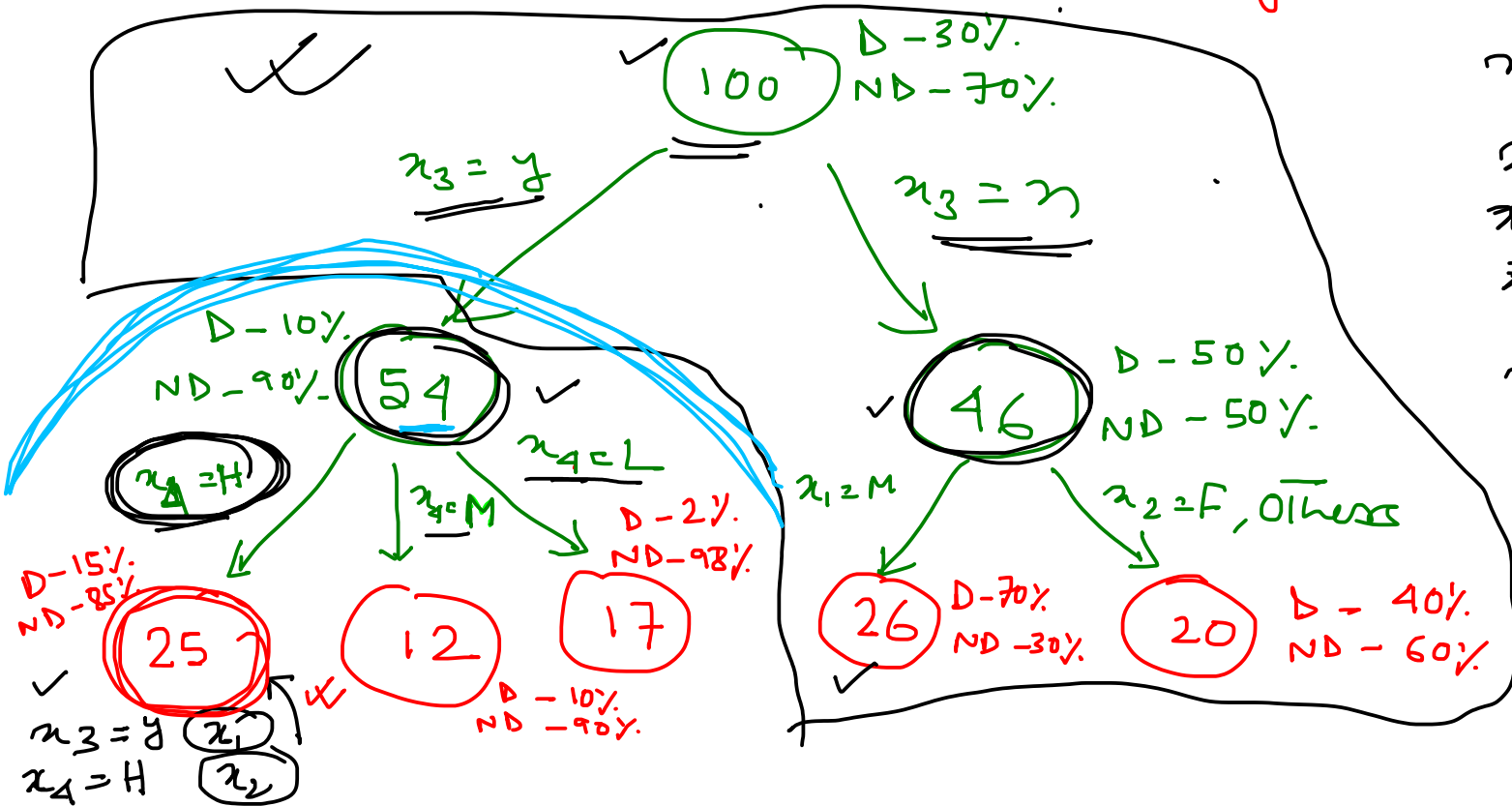
$\chi^2 \uparrow$  the prob. of having no relationship  $\downarrow$



X



- until when all points in the last node are having same  $y$



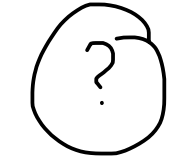
new case

$$x_1 = M$$

$$x_2 = H$$

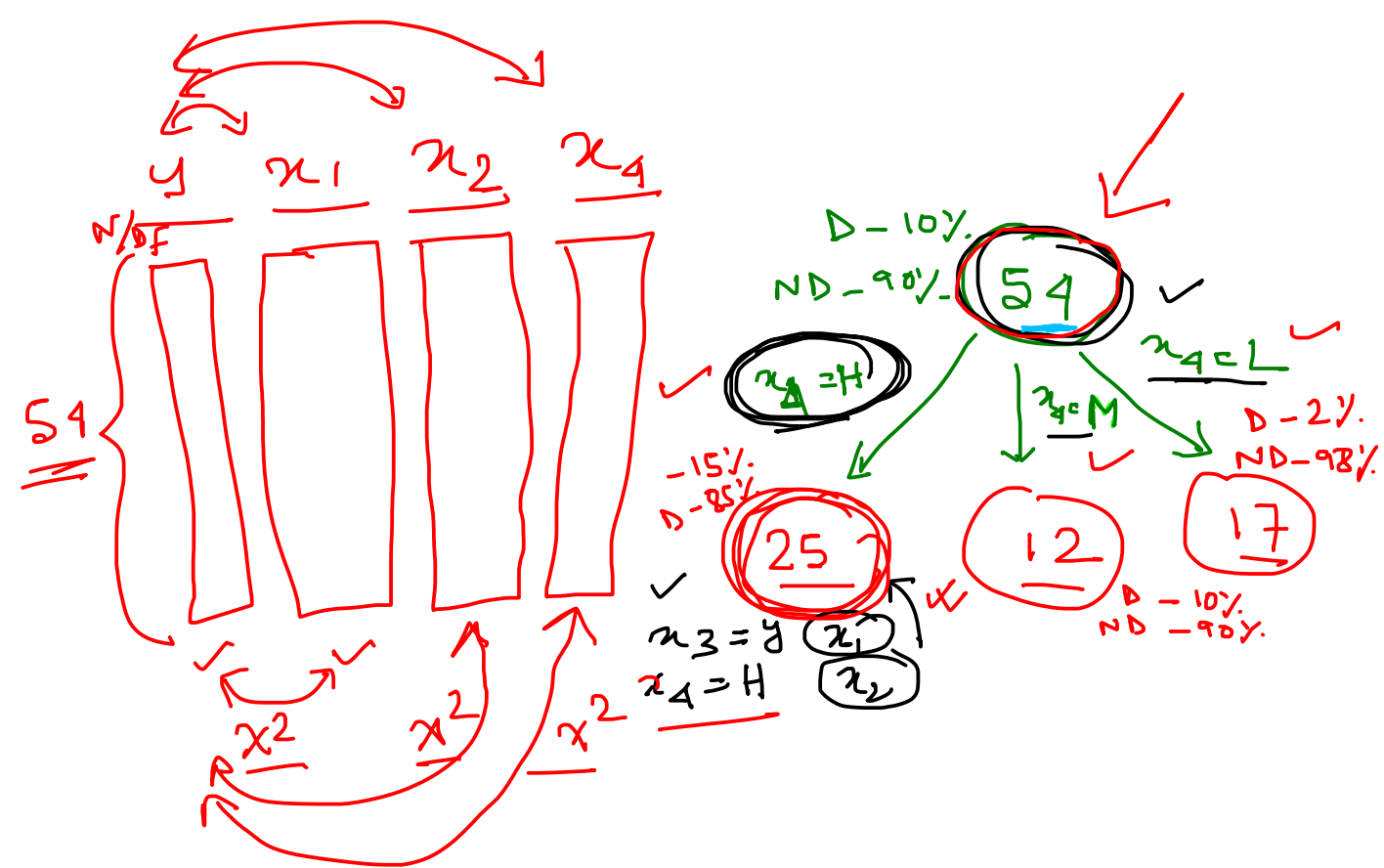
$$x_3 = n$$

$$x_4 = L$$



$P(D) =$

with  $x_1$  or  $x_2$  is related to  $y$  for these 25 obs.



$$\underline{\underline{fit}} = \text{rpart}(Y \sim x_1 + x_2 + x_3, \text{data} = \text{data})$$



(1-7)

Brand  
Preference towards  
maggie

$y$

(Income)

$X_1$

(Gender)

$X_2$

(R/U)

$X_3$

CART

100 obs

$$\sum (y - \bar{y})^2 = TSS$$

$\bar{X}$

100 obs

M

54

F

46

$\bar{X}_M$

$\bar{X}_F$

↑  
- If the groups are  
away from  
each other  
(between group diff)

↓  
- If the groups are  
within  
very cohesive themselves  
(within group diff)

TSS =  $BSS + WSS$

$$BSS = \sum (\bar{y} - \bar{y}_i)^2 \quad \uparrow$$

$$WSS = \sum_i \sum_j (y_{ij} - \bar{y})^2 \quad \downarrow$$

$$\left( \text{BPrd}(\cdot) \right) \underline{y} \quad \left( \text{M/F} \right) \underline{x_1} \quad \left( \text{Income} \right) \underline{x_2} \quad \xrightarrow{\quad} \left( \text{R/U} \right) \underline{x_3} \quad \boxed{0.010001}$$
