→ Z-test

→ t-test

↓

$n < 30$

no pop S.D

use-case ✓

↳ compare categorical vs numeric ✓ data.

A	B
1 ✓	2 ✓
2 ✓	3 ✓
3 ✓	4 ✓
μ_1	μ_1

categorical vs categorical [Chi-squared test]

↓

India

Sachin

century no-century

win loss

DoF (Degree of Freedom)

Salary	S_1	S_2	S_3	S_4
	40	30	40	?

$$\boxed{\underline{avg} = 40}$$

S_4 ↓ 50

How many known salaries required to find the missing salary values given we know M salary.

n

DoF : $(n-1)$

	H	W	
5	73	85	
	68	73 ✓	
	74	96	
	71	82	(8)
	✓ 62	70	
	(71)	(81)	

DoF : $(n_1 - 1) + (n_2 - 1)$

$=$

n_1 n_2

win

Sachin Century

	F	T ✓	
F	x	$314 - x$	314 ✓
T ✓	$176 - x$		(46) ✓

176	184	360

✓ ✓

(1)

4-1

	A	B	C	D	
X	p ✓	r	t	✓	349
Y	q ✓	s	u	✓	151
Z	✓	✓	✓	✓	150
	150 ✓	150 ✓	200 ✓	150 ✓	650

3-1

(1) (3)

(6) ✓✓

(4)

DoF:

C1 → 3

C2 → 4

$(3-1) \times (4-1)$

$= 6$

$$(r-1) \times (m-1)$$

Chi-Squared (Goodness of Fit)

50 trials

contingency matrix
cross-tab

H_0 : fair coin

H_a : not a fair coin

	H	T	
observed	28	25	50
expected			

→ Actual

(28) ✓	(22) ✓	(50) ✓

9 9 DoF ✓✓

$(O-E)^2$

$(3)^2 + (3)^2$ (1) ✓✓

observed - expected $= +(3)$

-3

total error

$= 3^2 + 3^2$

$= (18)$ ✓✓

$(O-E)^2 \rightarrow$ +ve error

✓ 1 cr ✓ 100 cr ✓

A

Rs 10

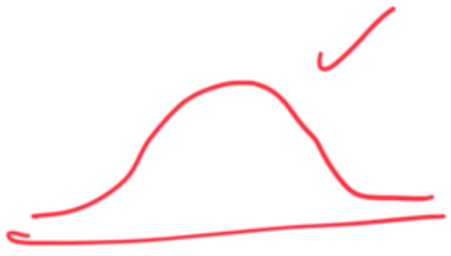
15

50 cr

$$\text{Error} = \frac{(O-E)^2}{E}$$

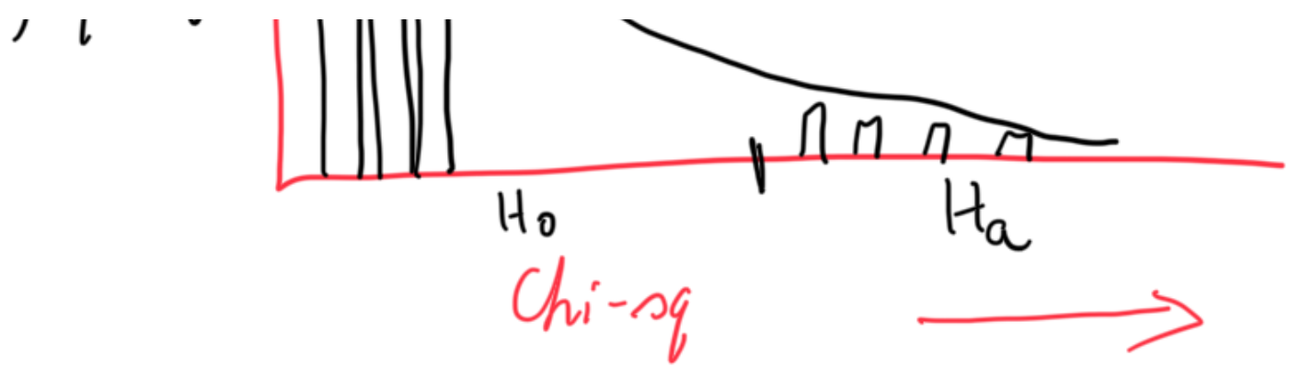
Chi-sq statistic

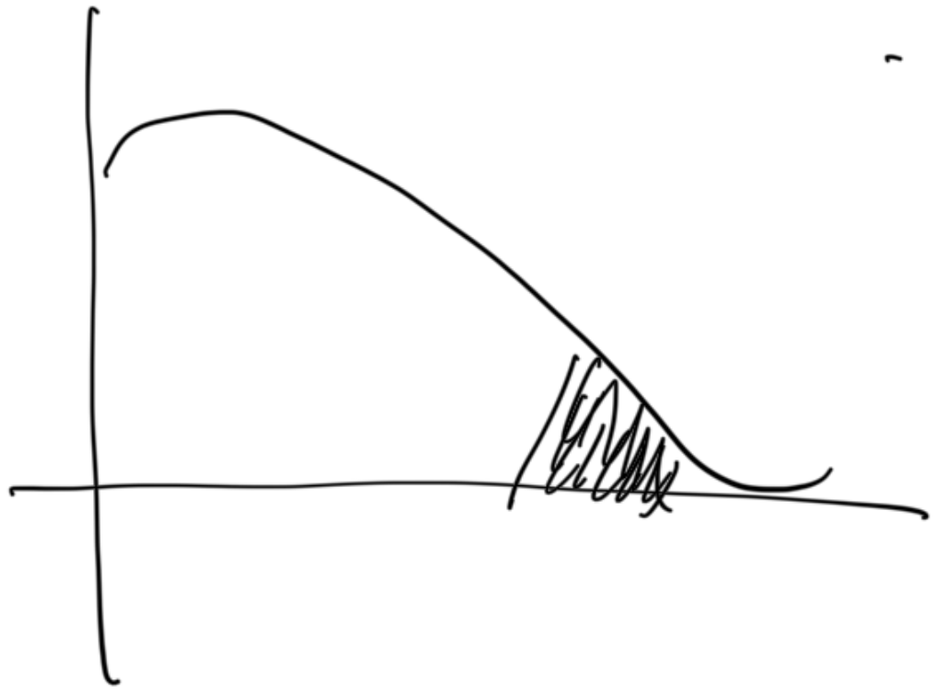
χ^2



Distribution of χ^2 test.

Coin is fair





Test of Independence.

Q: is gender related to online vs offline shopping behaviour

Shopping

observed values

	M	F	
online	206 249	59 102	308
offline	527 484	72 115	599
	733	174	907

34%

599/907 ≈ 66%

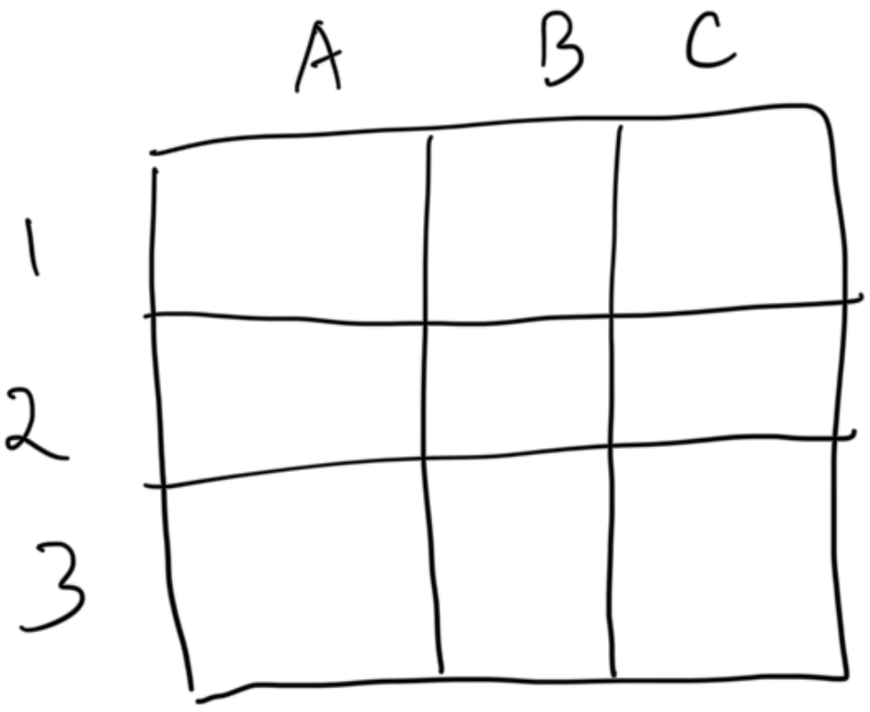
66% of 733 =

$$\frac{(206-249)^2}{249} +$$

Using relation b/w gender and shopping style

H_0 : [illegible]

H_a : there is a relation

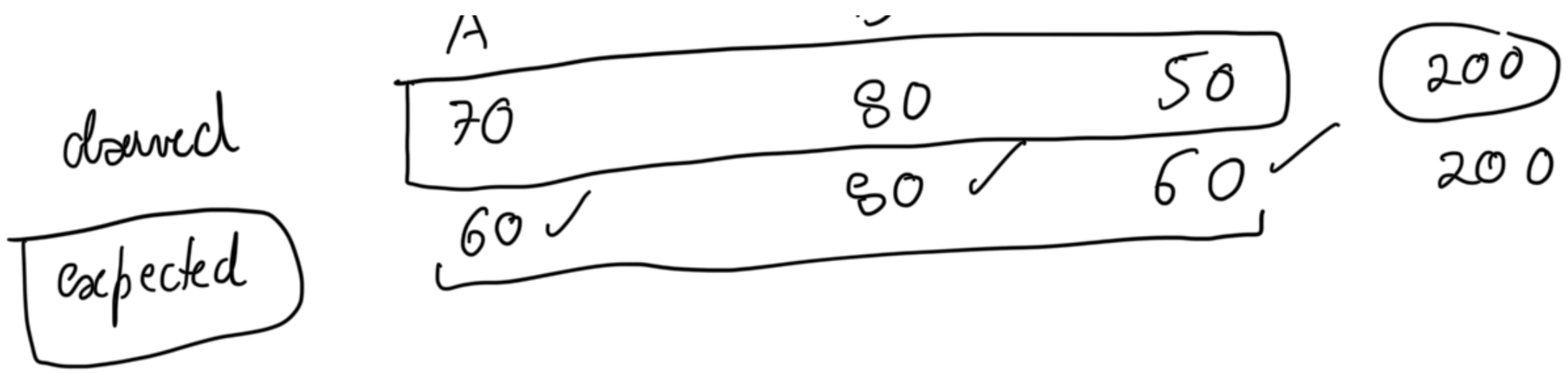


$$(n-1) \times (m-1)$$

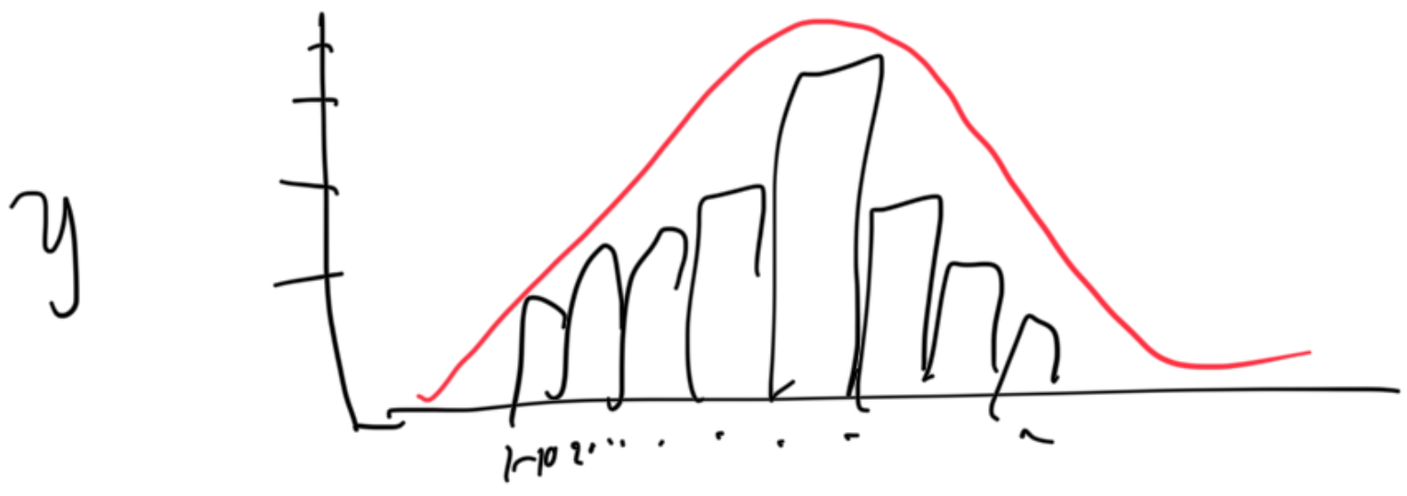
$$2 \times 2 = 4$$

$3-1 = 2$

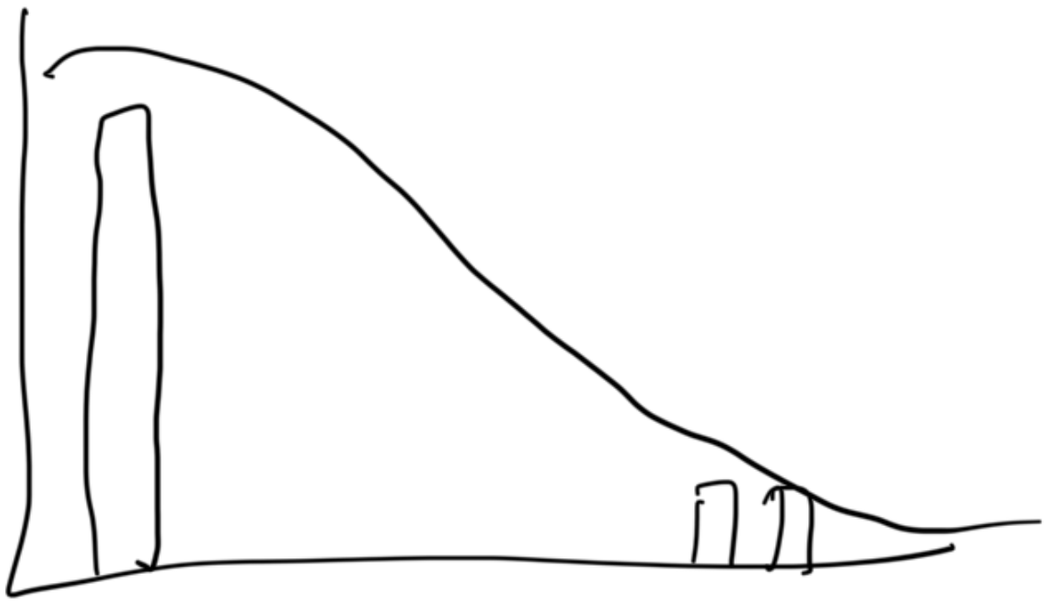
A ✓ B ✓ C ✓



GOF: for a categorical var
if obs and expected match or not.



ɔC
age



Chi-