



will start at 9:05 PM

- Chi-square test
- ANOVA

2 - Remedial version

sql ↔ Remedial version

Chi-square Test
"ki"



Card-squared → chinese

⇒ Product Mgrs

$x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6 \quad x_7 \quad x_8 \quad x$

$$y = a x_1^2 + \sqrt{b} x_2^3 + \dots +$$

⇒ 50 → toss a coin

25 H 25 T ⇒ fair coin
20 H 22 T ⇒ unfair coin

23 H 27 T \Rightarrow fair coin
10 H 40 T \Rightarrow biased coin

$(x_1' - x_1)^2 + (x_2' - x_2)^2$

$$\chi^2 = \sum \frac{(\text{obs} - \text{exp})^2}{\text{exp}}$$

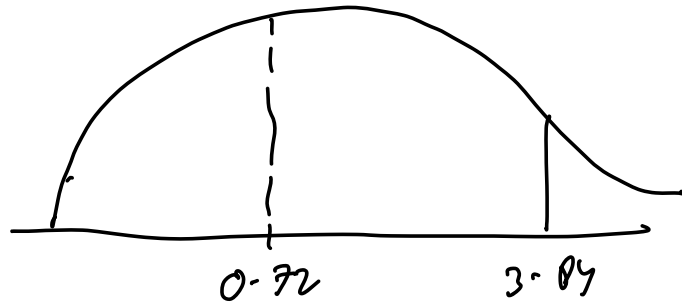
	Head	Tail
Expected Res (€)	25	25
Observed Val (€)	28	22

$$x^2 = \frac{(28 - 25)^2}{25} + \frac{(22 - 25)^2}{25}$$
$$= 0.72$$

$$\text{DOF} = (\# \text{ rows} - 1) * (\# \text{ col} - 1)$$
$$(2-1) * (2-1)$$
$$\Rightarrow 1$$

significance level $\neq 95\%$

\downarrow
 3.84



	Head	Tail
Expected Res (E)	25	25
Observed Val (O)	45	5

$$\begin{aligned} \chi^2 &= \frac{(45 - 25)^2}{25} + \frac{(5 - 25)^2}{25} \\ &= \frac{(20)^2}{25} + \frac{(-20)^2}{25} \\ &= \frac{400}{25} + \frac{400}{25} = 32 \end{aligned}$$

⇒ Dice ⇒ roll 36 times

1, 2, 3, 4, 5, 6
 ↓
 6

	1	2	3	4	5	6
Expected (E)	6	6	6	6	6	6
Obs (O)	2	4	8	9	3	10

$$\chi^2 = \frac{(2-6)^2}{6} + \frac{(4-6)^2}{6} + \frac{(8-6)^2}{6} + \frac{(9-6)^2}{6} + \frac{(3-6)^2}{6} + \frac{(10-6)^2}{6}$$

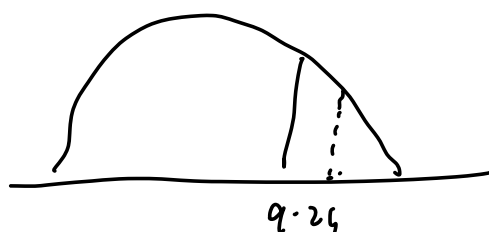
$$= 9.66$$

$$DOF = (2-1) \times (6-1)$$

$$= 5$$

sig lev ⇒ 90%

$$\Rightarrow 9.24$$



⇒ E-commerce company

online purchase	gender
1	Male
1	Male
1	Female
0	Female

	Male	Female	Total
Do not purchase online	527	72	599 (66%)
Purchase online	206	102	308 (34%)
Total	733	174	907

$$\text{not online \%} \Rightarrow \frac{599}{907} = 66\%$$

$$\text{online \%} \Rightarrow 34\%$$

Expected table

	Male	Female	Total
Do. not use gro onl	404	115	599 (66%)
Use gro online	249	59	308 (34%)
Total	733	174	907

$$733 \times \frac{66}{100} = 404$$

$$\chi^2 = \frac{(527 - 404)^2}{404} + \frac{(72 - 115)^2}{115} + \frac{(206 - 249)^2}{249} + \frac{(102 - 59)^2}{59}$$

$$\chi^2 = 59$$

$$DOF = (2-1) \times (2-1) = 1$$

sig level \Rightarrow 90% \Rightarrow 2.71

Gender do matter



2.71

10:20 PM

ANOVA (Analysis of variance)

→ American team: Tall & weight ↑

→ Japan: low height & weight ↓

→ College Prof: mix

A - G ⇒

H - N ⇒

O - Z ⇒

} in-group var ↑

$$F \text{ ratio} = \frac{\text{Variation between groups}}{\text{Variation within group}}$$

⇓

more likely diff means among
grps

A	B	C
25	30	18
25	30	30
27	21	29
30	24	29
23	26	24
20	28	26

Step 1

$$m_1 \rightarrow \bar{y}_1 = \frac{1}{6} \sum y_{1i} = 25.0$$

$$m_2 \rightarrow \bar{y}_2 = \frac{1}{6} \sum y_{2i} = 26.5$$

$$m_3 \rightarrow \bar{y}_3 = \frac{1}{6} \sum y_{3i} = 26$$

Step 2

$$m \rightarrow \bar{y} = \frac{\sum \bar{y}_i}{k} = \frac{m_1 + m_2 + m_3}{3} = 25.83$$

Step 3 "between groups"

$$\begin{aligned}
 SS_B &= 6(25 - 25.83)^2 + 6(26.5 - 25.83)^2 \\
 &\quad + 6(26 - 25.83)^2 \\
 &= 6.9
 \end{aligned}$$

$$\begin{aligned} \text{Mean-square value} &= \frac{SSB}{df} = \frac{6.9}{(3-1)} \\ (MSB) &= 3.49 \end{aligned}$$

Step 4

==>

$$SSW \Rightarrow (x_1 - m_1)^2 + (x_2 - m_1)^2 + \dots + (y_1 - m_2)^2$$

$$\Rightarrow 223.6$$

$$MSW \Rightarrow \frac{SSW}{df} = \frac{223.6}{18-3} = 14.9$$

$$F = \frac{3.49}{14.9} = 0.23$$

$$\Rightarrow \frac{3.68}{\downarrow}$$

Cx

