T-test, 2-test, Binomial, Poission

Parametrec tes

★

chi-squre test - non parametrec

It is non-parametrz test that is performed on categarical data.

eg:- In the 2000 USA census the age of individuals in a small town were found to be the following.

Less than 18	18-35	>35
20%	30%	50%

In 2010, age of $n = 500$ individual were sample below result.

<18	18-35	>35
105	230	165

using $\alpha = 0.05$, can you conclud distribution of age has been changed in 10 years.

<u>Sol^n</u>

$$\mu = \overline{X} \implies H_0$$
$$\mu \neq \overline{X} \implies H_1$$

$\alpha = 0.05$

C.I. $= 95\%$

Degree of freedom $= n-1$

$$= 3-1$$
$$= 2$$

✓ chi-square test $= x^2$

According to table on $DF = 2$ and $\alpha = 0.05$, $x^2 = 5.991$

✦ calculations

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

f_o = observed value

f_e = expected value.

	<18	18-35	>35
f_o	105	230	165
f_e	$\frac{500 \times 20}{100}$	$\frac{500 \times 30}{100}$	$\frac{500 \times 50}{100}$
	$\Rightarrow$ 100	150	250

$$x^2 = \frac{(105 - 100)^2}{100} + \frac{(230 - 150)^2}{150} + \frac{(165 - 250)^2}{250}$$

$$\Rightarrow 0.25 + 42.66 + 28.9$$

$$\Rightarrow 71.81$$

$$x^2 > 5.991$$

$$\Rightarrow \quad 71.81 > 5.991$$

we reject null hypothesis
and accept altanet hypothesis.

F-test (Anova testing)

E.g. Researcher want to test a medication
They split participant in 3
condition (0 mg, 50 mg, 100 mg)
then anxiety level is check
on scale. 1 - 10 Are there any
diffrence b/w the 3 condition
$\alpha = 0.05$.

0 mg	50 mg	100 mg
9	7	4
8	6	3
7	6	2
8	7	3
8	8	4
9	7	3
8	6	7

7

Sol^n =)

$$\mu_{0mg} = \mu_{50mg} = \mu_{100mg} \quad H_0$$

$$\mu_{0mg} \neq \mu_{50mg} \neq \mu_{100mg} \quad H_1$$

$$\alpha = 0.05$$

Degree of freedom

$$df_{B/w} = \alpha - 1 = 3 - 1 = 2$$

$$df_{within} = N - \alpha = 21 - 3 = 18$$

$$df_{total} = N - 1 = 21 - 1 = 20$$

Decision rule :-

$$(dF_{B/w}, \ df_{within})$$

$$(2, 18)$$

F-table $\alpha = 0.05$

from table $\Rightarrow$ 3.5546

calculate F-test :-

	sum of Square	degree of freedom	mean of Square
B/w	98.67	2	$98.67/2 = 49.34$
within	10.29	18	$10.29/18 = 0.57$
total	108.95	20	$F = \dfrac{MS_{B/w}}{MS_{within}} = \dfrac{49.34}{0.57}$

$$\Rightarrow 86.56$$

$$SS_{b/w} = \sum \frac{\left(\sum ai\right)^2}{n} - \frac{T^2}{N}$$

$$\text{omy} = (9+8+7+8+8+9+8) = 57$$

$$\text{soy} = (7+6+6+7+8+7+6) = 47$$

$$\text{room} = (4+3+2+3+4+3+2) = 21$$

$$T = 57+47+21 = 125$$

$$SS_{b/w} = \frac{(57)^2+(47)^2+(21)^2}{7} - \frac{(125)^2}{21}$$

$$SS_{b/w} = \underline{98.67}$$

$$\underline{SS_{within}} = \Sigma y^2 - \Sigma \frac{(\Sigma c_i)^2}{n} \longrightarrow$$

$$\Sigma y^2 = 9^2+8^2+7^2+8^2+ \cdots \cdots +2^2$$

$$= 853$$

$$SS_{within} \Rightarrow 10.29$$

$$F = 86.56$$

$$Table = 3.5546$$

$$F > \alpha \cdot Table$$

$\Rightarrow$ We reject null hypothesis and accept alternet hypothesis.

⭐⇒ Type-I and Type-II

Real data — confusion matrix

Predict

		T	F
	T	TP	FP
	F	FN	

X	$\overline{X}$	
T	T	-
T	F	-
F	T	-
F	F	-
T	F	-
F	F	-

Real Data

Calculated data

	T	F
T	T<u>P</u>	FP
F	FN	T<u>N</u>

FP → Type-I

FN → Type-II

Y	$\overline{Y}$		
R.D.	CD		
T	F	—	FN
T	T	—	TP
F	F	—	TN
F	T	—	FP
T	F	—	FN
F	T	—	FP

Data transformation

① Standardization
② Normalization

① Standarization —

<u>ML</u> algo. - $[0,1]$

| 1 | 100 | 1000 |
| 10 | 100 | 10000 |

$$\frac{5}{10000} = 0.0005 \qquad \frac{9999}{10000} = 0.99$$

$$\frac{10000}{10000} = 1$$

$$x^1 = \frac{x_i - \mu}{\sigma}$$

② <u>Normalization</u> (min, max)

DL, Algo. — $[-1, 1]$

$$x' = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

X

1

2

4

7

11

12

15

18

20

min = 1

max = 20