T-test, Z-test, Binomial, Poisson

Parametrec tes

✱ chi-squre test - non parametrec

It is non-parametric test that is performed on categorical data.

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

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

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

<18	18-35	>35
105	230	165

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

Soln

$\mu = \bar{X} \Rightarrow H_0$

$\mu \neq \bar{X} \Rightarrow H_1$

$\alpha = 0.05$

C.I. = 95%

Degree of freedom = $n - 1$

$= 3 - 1$

$= 2$

chi-square test $= \chi^2$

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

✡ calculations

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

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

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

$$\Rightarrow 0.25 + 42.66 + 28.9$$

$$\Rightarrow 71.81$$

$\chi^2 > 5.991$

$\Rightarrow \quad 71.81 > 5.991$

we reject null hypothesis and accept alternet 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 difference b/w the 3 condition $\alpha = 0.05$.

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

Soln =) $\mu_{0mg} = \mu_{50mg} = \mu_{100mg}$ H_0

$\mu_{0mg} \neq \mu_{50mg} \neq \mu_{100mg.}$ 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	$F = \frac{MS_{B/w}}{MS_{within}} = \frac{49.34}{0.57}$

$SS_{b/w} = \sum \frac{(\sum a_i)^2}{n} - \frac{T^2}{N} \Rightarrow 86.56$

$\Sigma y = (9+8+7+8+8+9+8) = 57$

$\Sigma y = (7+6+6+7+8+7+6) = 47$

$\Sigma y = (4+3+2+3+4+3+2) = 21$

$T = 57+47+21 = 125$

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

$SS_{b/w} = 98.67$

$$SS_{within} = \Sigma y^2 - \sum \frac{(\Sigma y_i)^2}{n} \rightarrow$$

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

$= 853$

$SS_{within} = 10.29$

$F = 86.56$

Table $= 3.5546$

$F > \alpha$ Table

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

Type-I and Type-II

Actual

confusion matrix

Predict

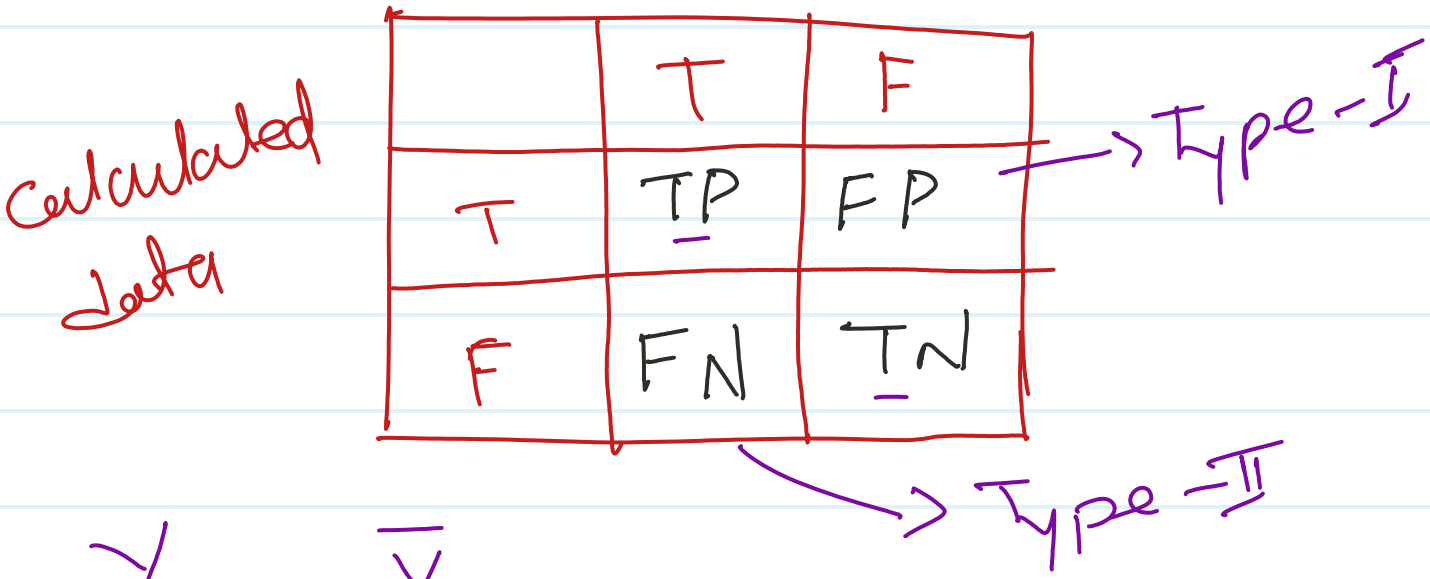
	T	F
T	TP	FP
F	FN	TN

Type-I

Type-II

A.	P		
X	$\overline{X}$		
T	T	-	TP
T	F	-	FN
F	T	-	FP
F	F	-	TN
T	F	-	FN
F	F	-	TN

Real Data



Y	$\bar{Y}$	
R.D.	C.D	
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. Standardization –

ML algo. – $[0, 1]$

1 100 1000

10 100 10000

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

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

$$x' = \frac{x_i - \mu}{\sigma}$$

② Normalization (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