13th July 2023 :

- Paired T test (Theory + code)
- KS Test Kolmogorov–Smirnov test
- Log Normal.
- Central limit Theorem.
  (Simulation, Theoretical proof)

[Box Cox transformation]

Lin. Reg^n, logistic Reg^n, Naive Bayes, SVM, LDA

Ttest ind.

num. vs Cat.

ID	Income	Gender
1	10	M
2	12	F
3	13.5	M
4	3.5	F
5	4.2	M
6	21	M
1	⋮	F
1	⋮	
100	⋮	

✱ Paired T test.

Revision of Two Sample T-test

μ_m μ_f } avg. income. } ① income & males

H_0 : $\mu_m = \mu_f$

H_A : $\mu_m \neq \mu_f$ } → 2 tailed test

• test = $\left(\frac{\mu_m - \mu_f}{\sqrt{\frac{s_m^2}{n_m} + \frac{s_f^2}{n_f}}} \right)$ ~ t dbn.

μ_m, μ_f

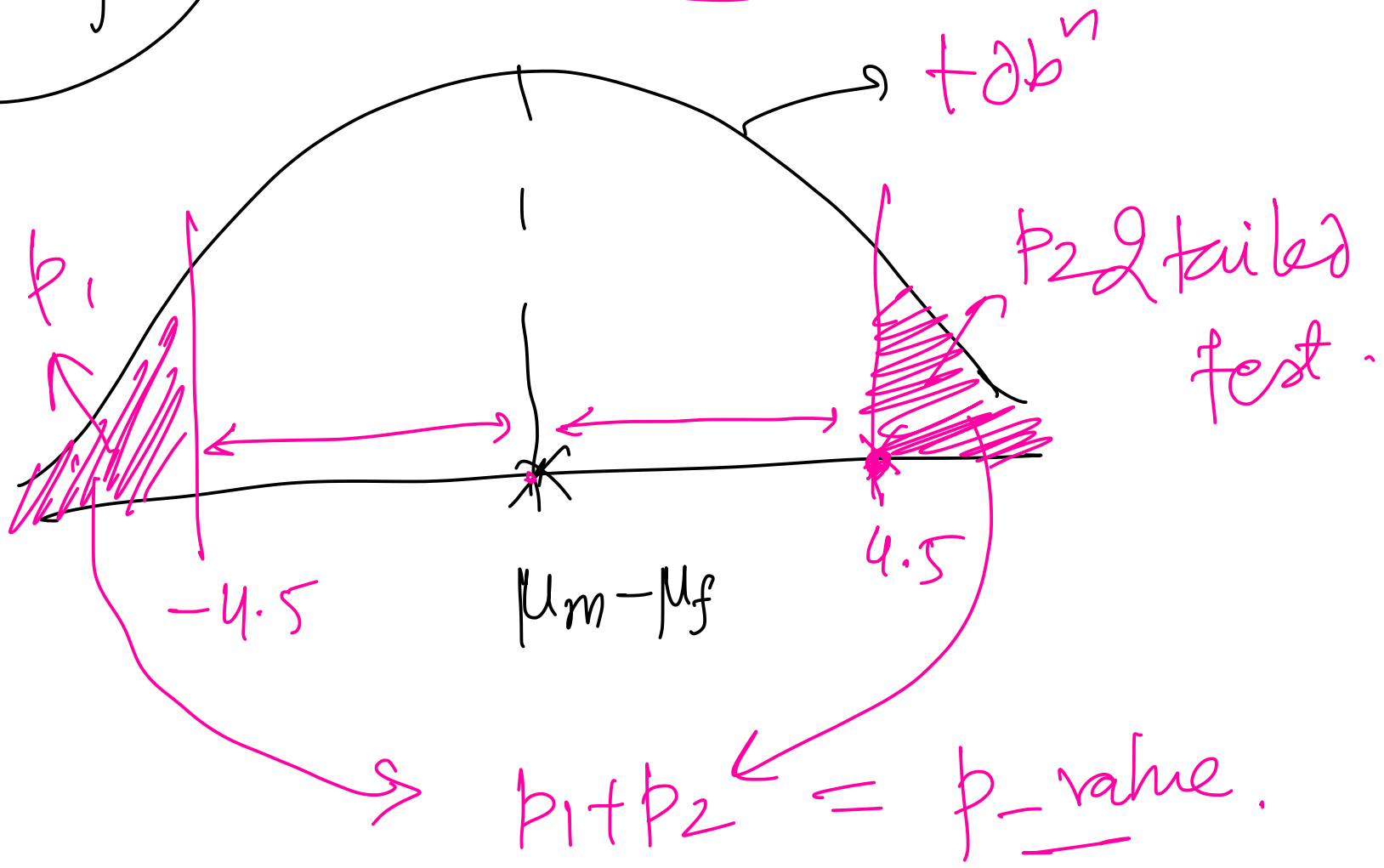
s_m^2, s_f^2

n_m, n_f

✓ t_stat ∩ t dbn

↓

num. #



$p_1 + p_2$ = p_value.

Drug 1 → A → n_1

Drug 2 → B → n_2

Independent of each other

ttest_ind *

$x_1 \rightarrow 65$

$x_2 \rightarrow 66$

$x_3 \rightarrow 64.1$

$x_5 \rightarrow 72.5$

$\frac{x_1 + \cdots + x_5}{5}$

m

μ

Paired test :

— measuring students performance

Student ID	Befor PS session	After PS session
101	30	35
102	35	31
:	41	47
:	:	
n	57	65

PS session

$(t_2 - t_1)$

ID	Test 1	Test 2	diff in marks.
1	40	38	d_1
2	39	45	d_2
3	45	47	d_3
4	53	61	⋮
5	65	72	⋮
⋮	⋮	⋮	⋮
n	⋮	⋮	d_n

$$\frac{d_1 + d_2 + \cdots + d_n}{n} = \bar{d}$$

$$\bar{d} = \frac{\Sigma d_i}{n}$$

↓

mean difference

μ: mean of Population difference.

H_0: PS not helpful, mean diff $= 0$, $\mu = 0$

H_A: PS helpful, mean diff > 0, $\mu > 0$

Test statistic:

↳ under the assumption H_0 is True

$$\text{test stat} = \frac{\bar{d} - \mu}{(s/\sqrt{n})}$$ ($\mu \to 0$)

s: sample std. dev^n

n: # of students.

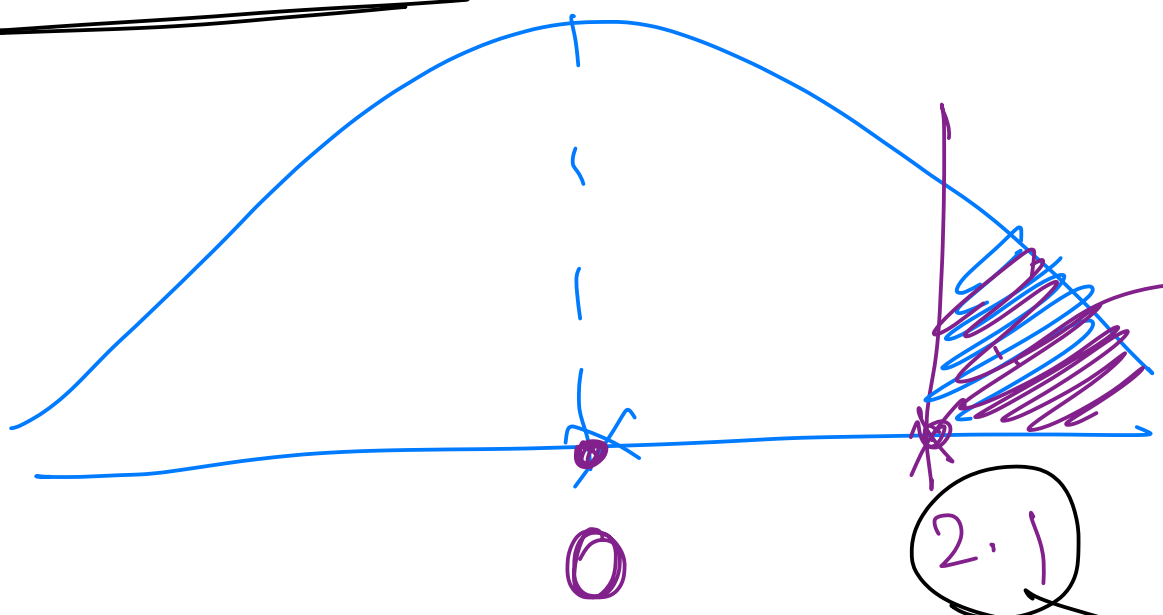
} Z test, t test

$$\sim \left(\frac{m - \mu}{(s/\sqrt{n})} \right)$$

test stat $= \dfrac{\bar{d}}{(s/\sqrt{n})}$ $\sim t_{df}$

$\alpha = 0.05$

Which tail test



under the assumption H_0 is True

$\mu = 0$

p_value ?

mean $[t_2 - t_1]$

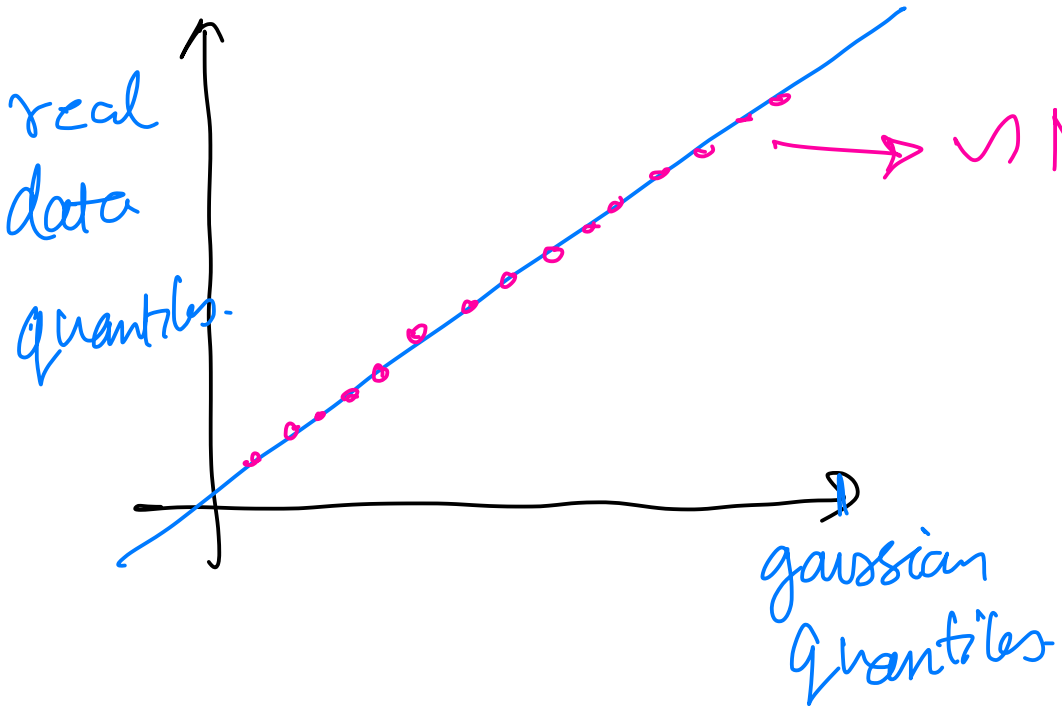
$\dfrac{\bar{d}}{(s/\sqrt{n})}$

p_value $< \alpha$

↓

reject H_0

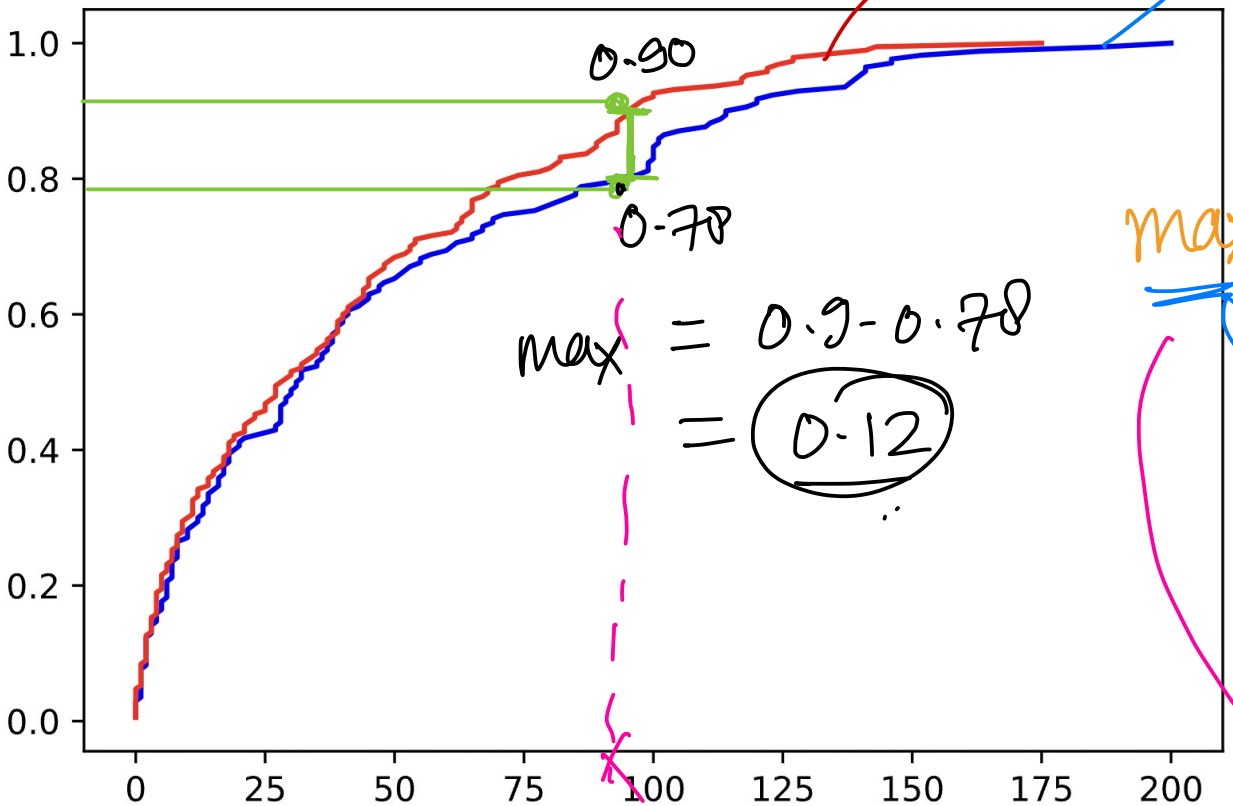
KS test



→ Normal dⁿ

Sachin's Dataset.

KS Test



F(II)

F(I)

at

$|F(II) - F(I)|$

max

$\max = 0.9 - 0.78$

$= 0.12$

statistically significant or not.

Test stat ↳

Kolmogorov

x

Log Normal

right skewed

- age
- income
- waiting time