Final Part of Inferential Statistics

Z-test ⟶ Numerical Sample

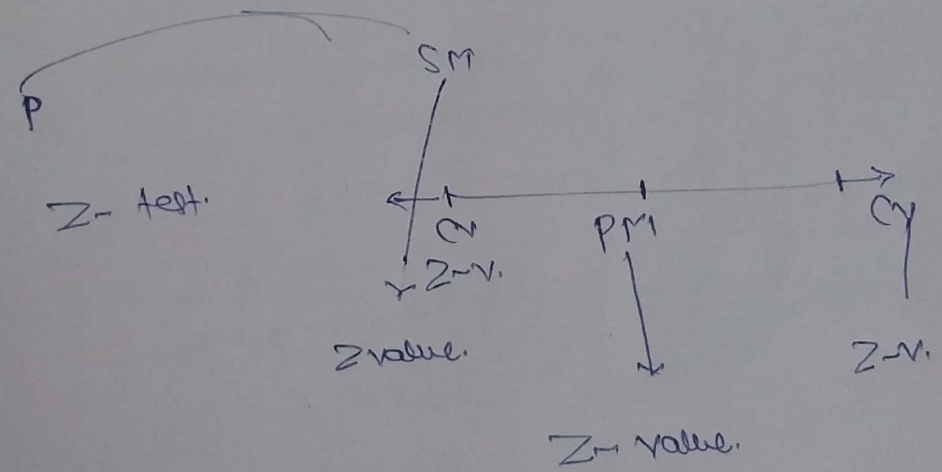
① Sample size ≥ 30

② Data should be normally distributed. → Shapiro will test for normality.

③ population Standard Deviation must be given to us.

④ Sample should be randomly collected

38:min



$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$$Z = \frac{SM - PM}{PSD / \sqrt{n}}$$

Problems

Suppose a company is evaluating the impact of a new training program on the productivity of its employees. The company has data on the average productivity of its employees before implementing the new training program. The average productivity

was 50 units per day with a known pop Standard deviation of 5 units. After implementing the training program, the company measures the productivity of a pram sample of 30 employees. The sample employees have an average productivity 53 units per day. The company wants to know if the new training program has significantly improved the the productivity of the employee.

PM = 50 units.
PSD = 5 units
SS = 30 (n)
SM = 53 units.

Note :- Z-test work with numpy library.

→ Import numpy as np
from statsmodels.stats.weightstats import Ztest
Sample_data = list(np.random.randomint(50,59
size = 30))

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Note the Z value of population mean is always zero (0)

$$t = \frac{SM - PM}{SSD\sqrt{n}}$$

6. [illegible] Necessary.

$$Z\text{-Score} = \frac{\bar{x} - \mu}{PSD\sqrt{n}}$$

$$= \frac{53-50}{5/\sqrt{30}}$$

$$= \frac{53}{5\sqrt{30}}$$

$$= \frac{3}{5/5.4772255751} =$$

$$= \frac{3}{0.9128709292}$$

$$= 3.286335449$$

Notes :- $p < 0.05$ → Not normally distributed.
→ according to Shapiro wilk test.

$p \geq 0.05$ → normally distributed.

Since we are here working with two dataset
so

$S_1 = S_2$

Variance Variance

So
there is Levene test for checking the variance

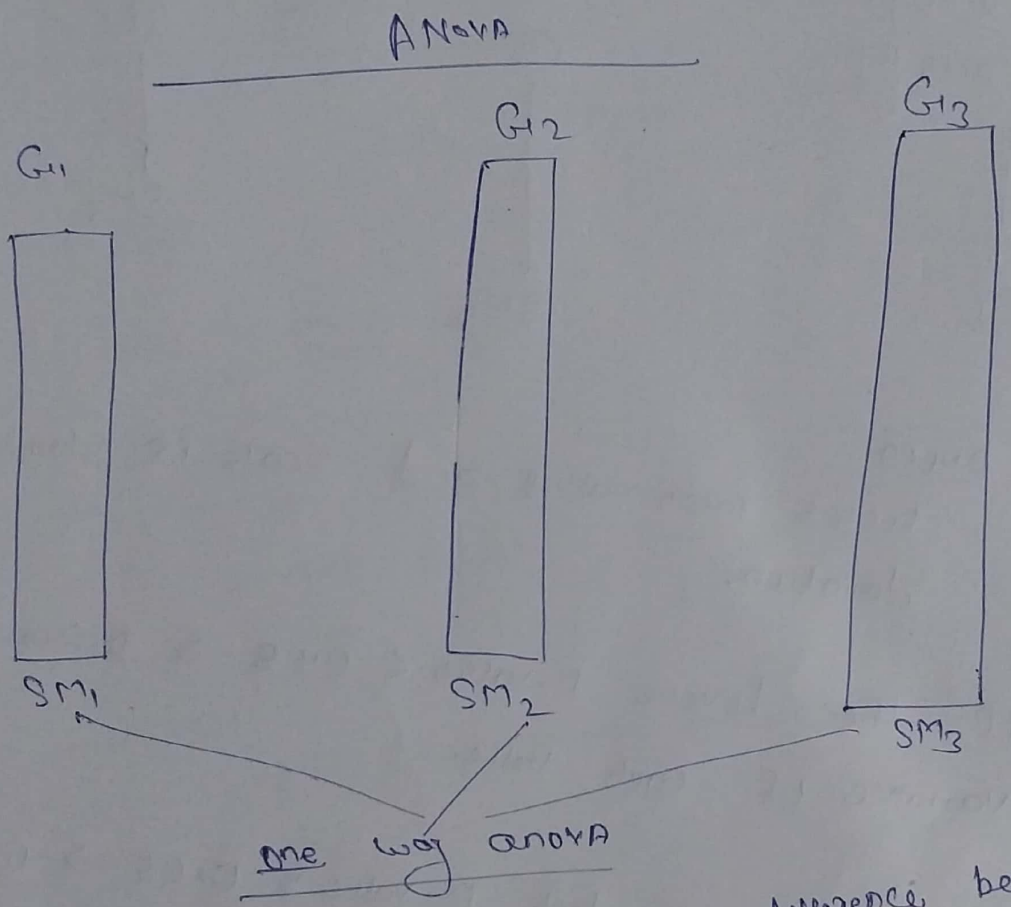
Levene Test

$P < 0.05$ → variance not same

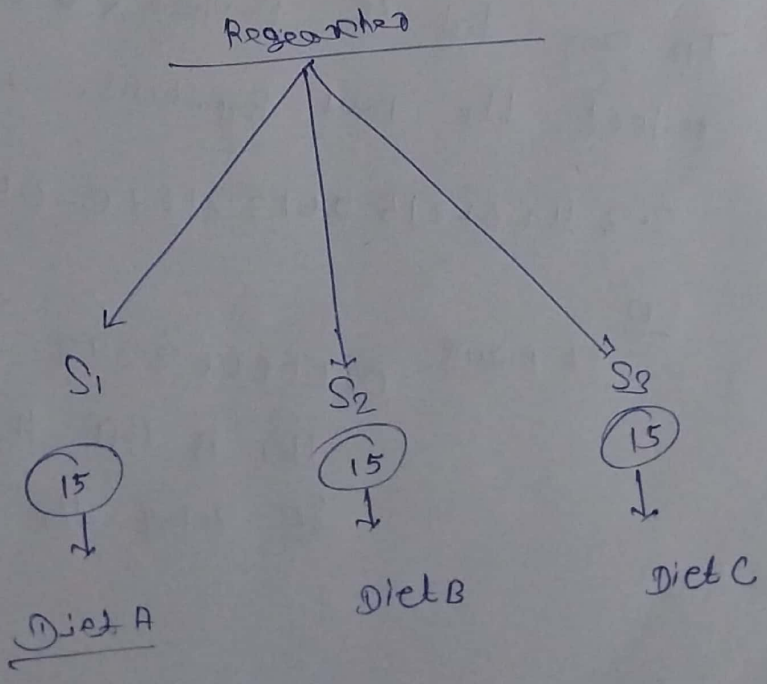
$P \geq 0.05$ → variance are same → Checked the normality

normally distributed

→ ANOVA uses F-Test to check the null hypothesis.



Is there a significant difference between these three sample means or not.



In one month weight Loss

(1) 3.8 kg
(2) 4.' kg
(3) 2.15 kg

↓

SM1

In one month weight loss

(1)
(2)
(3)

↓

SM2

In one month weight Loss

(1)
(2)
(3)

↓

SM3

Notes: in ANOVA

Loc = 5 mean values and scale is standard deviation.

Notes: if the Levene p-value < 0.05 → Different variance of each value.

if the Levene test p-value > 0.05 → we have same variances in each test.

⇒ In any test if p-value < 0.05 we will reject the null hypothesis.

Impt 2.2426215901321 67 e-05

so

It means 0.000002242

this is less than < 0.05

so reject the null hypothesis

Perform Levene's test for equal variance.

levene_stat, levene_p_value = levene(diet_A_loss, diet_B_loss, diet_C_loss)

print(levene_p_value)

Test → must package

→ PRT

Que Chi2 test

SQL

→ Time series

→ random forest

→ OOPS

→ CUT CT, common table expression, view