Agenda

- 1 Sample T-test
- 2 Sample T-Test
- Paired T-test

1 Sample T-test

* In most scenarios population std deviation (σ) is unknown, hence we cannot calculate z-score / z-statistic.
* However given the sample Data we can compute sample-standard deviation (s), and use that as **Estimate for (σ)**.

- This results in modified framework: T-test and the statistic is known as T-statistic
- T-statistic follows T-Distribution
- When dealing with smaller samples (<30), t-test is more appropriate as it accounts for increased Uncertainity

Example: 1 Sample Z-test to 1 Sample Test

IQ Increase with pill

data = 110, 105, 98, 102, 99, 104, 115, 95

H_0 : $\mu = 100$ (pill has no effect)

H_a : $\mu \neq 100$ (pill has effect

$\mu = 100$

$\bar{x} = 103.5$

$\sigma = X$

$n = 8$

population SE $= \frac{\sigma}{\sqrt{n}}$ X

(can not calculate)

sample SE $= \frac{s}{\sqrt{n}}$ ✓

s = can be calculated from data, Hence:

Z- statistic $= \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$ $\xrightarrow{\sigma \to s}$ t- statistic $= \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$

s ⟶ Sample S.D (Bessel's correction)

↓

SSE → $\frac{s}{\sqrt{n}}$

2 Sample T-Test

School Example: Two Sample Test

① $H_0 : \mu_1 = \mu_2$

$H_a : \mu_1 \neq \mu_2$ (two tailed)

② $H_0 : \mu_1 = \mu_2$

$H_a : \mu_1 > \mu_2$ (Right tailed)

③ $H_0 : \mu_1 = \mu_2$

$H_a : \mu_1 < \mu_2$ (Left tailed)

For the above test, we can use "ttest_ind" function from scipy.stats

ind stands for independent

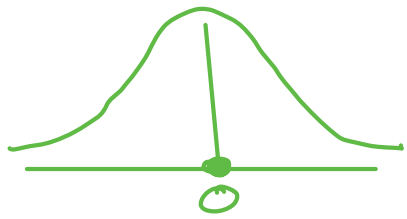
$H_a : \mu_1 > \mu_2$

$H_a : \mu_1 < \mu_2$

For two Sample t-test:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$Z \sim N(0,1)$
$t \Rightarrow t(0,1)$



Paired T-test

* It is used primarily to compare means of two related Groups (Before and after)

- This typically occurs when you're studying the impact of a treatment, intervention, or change within the same subjects over time or in some paired way.
- In your case, you're comparing "Before" and "After" measurements on an individual basis. For each person, you have two measurements:
  - **Person 1: Before and After**
  - **Person 2: Before and After**
- The paired t-test takes into account the paired nature of the data.
- It calculates the mean difference ($\mu_{after} - \mu_{before}$) of the paired measurements and then assesses whether it is statistically significant or not.
- This helps you determine if there's a significant change between the "Before" and "After" measurements within each pair.

Data

id	Before	After
1	Val 1	Val' 1
2	Val 2	Val' 2
3	"	"
"		

$H_0: \mu_{Before} = \mu_{after}$

Null and Alternate hypothesis

- Null Hypothesis (H_0): Problem-solving has no effect on the test scores.
  - In other words, the mean test scores before (test_1) and after (test_2) problem-solving are equal.
- Alternative Hypothesis (H_a): Problem-solving had an effect on the test scores.
  - This implies that the mean test scores before and after problem-solving are not equal.

t - statistic

$$t = \frac{\bar{D}}{\frac{s_D}{\sqrt{n}}}$$

$\bar{D}$: mean of Differences

s_D : SD of Differences

For the above test, we can use 'ttest_rel' function from scipy.stats

id	Before	After	Di
1	Val1	Val'1	$V_1 - V_1'$
2	Val2	Val'2	$V_2 - V_2'$
3	,,	,,	
,,			

$\bar{D}$: mean of Di

s_D : SD of Di's

Conclusion

We covered

1) 1 Sample T-test

2) 2 Sample Independent Test

3) Paired T-test

3) T-test is used for Testing Relationship b/w Categorical → Continous

Note about T-test:

- If sample size $n > 30$, then the sample standard deviation will be very close to the value of standard error, i.e. $\frac{x-\mu}{\frac{s}{\sqrt{n}}} \approx \frac{x-\mu}{\frac{\sigma}{\sqrt{n}}}$
- i.e. T-test and Z-test become essentially the same.
- If the number of samples is low, you go for a T-test
- If it is high, you can use either T test or Z test
- That's why, scipy does not even have an implementation for Z-test.

num vs Cat → Z-test / T-test

Cat vs Cat → Chi Square

num vs num → Correlation

DOF

One week of Movie Nights

7 different Movies on 7 days

One movies on 1st Day →

freedom to pick any of the 7

One movies on 2nd Day →

freedom to pick any of the 6

.

.

.

One movies on 7th Day →

0

Ex: 1st person = 35 lakh

2nd person = 36 lakh

3rd person = ?

Avg salary of 35 lakh

3rd person = ?

Can we find 3rd person Salary?

Ex: 1st person = 35 lakh

2nd person = 36 lakh

3rd person = ?

4th person = 35 lakh

Avg salary 35 lakh

Can we find 3rd person Salary?