DATE / /

Hypothesis testing

Q what is Hypothesis testing ?
its an statistical Analysis
its formal procedures that statisticans use to test wether a hypothesis can be accepted or not
A hypothesis is an assumption about something

hypothesis testing → null Hypothesis
→ Alternate Hypothesis.

imp note

when standard deviation is known we gonna used Z-test

otherwise t-test

* example of Hypothesis testing
let say we have sample from population and we have to prove if this sample is properly representing population or not.

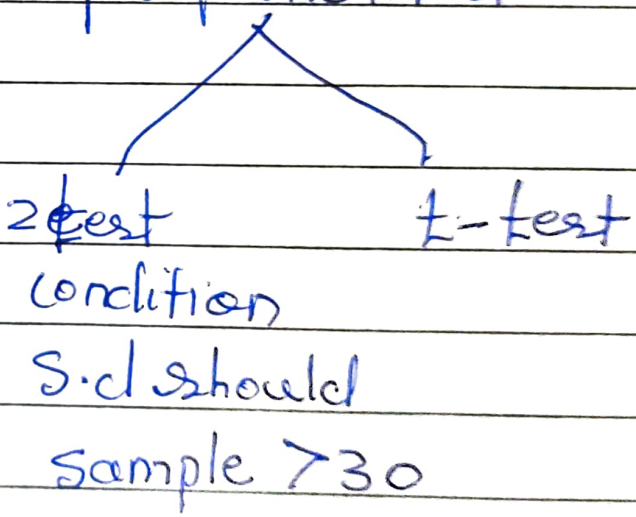
steps in Hypothesis testing

(1) create null and Alternate Hypothesis

H_0 = null (no decision (change))
H_a = Alternate (change)

(2) select test static

So we have proportion test



after selecting the porportion test then we go next is

(3) Determine Significance level ($p = 0.05$) — mostly used
α

(4) Determine the rejection criteria
Significance level represent this hypothesis is correct or incorrect

* $p \geq 0.05$ then Reject the null hypothesis
* $p < 0.05$ Accept the Alternate hypothesis.

5 (6) calculate test static (t-value)
to do this we will use

standard error

$$\frac{\bar{X} - m}{\frac{\sigma}{\sqrt{N}}}$$

for z-test

$$\frac{\bar{X} - M}{\frac{s.d}{\sqrt{N}}}$$

for P-test

= t-values

then put this t-value in python code in order to get p-value.

1 - stats.norm.cdf(1.65 , loc=0, scale=1)

t-value

= p-value

Now ~~calculate~~ check p-value to threshold values.

Determine whether to accept hypothesis based on statistics.

important note

if according to condition if we have one sample we will go for one sample test or 1 tail test (particular for one column)
and
if according to condition if we have two sample we will go for two sample test or 2 tail test. (mostly used in comparison >,<) (particular for two column).

No hypothesis (null hypothesis).

when there is not difference in sample or in condition.

★ So based on (test (statistics) we will determine the null and Alternate hypothesis

① chi square test

② t- student test

③ fisher's z-test

but after doing test I want to determine or level to prove the null and Alternate Hypothesis

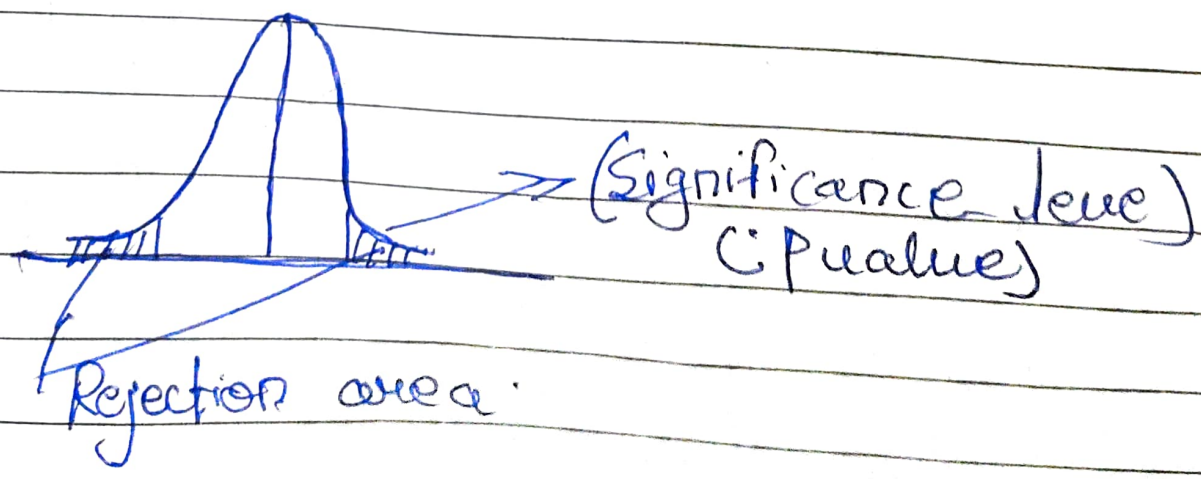
so we will use.

* level of significance (α) (may be 5%.)
  level of confidence (c) (if 5% then (c) is 95%.)

$$\alpha + c = 1$$

$$z = \left| \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right| \quad \text{for} \quad t = \left| \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \right|$$

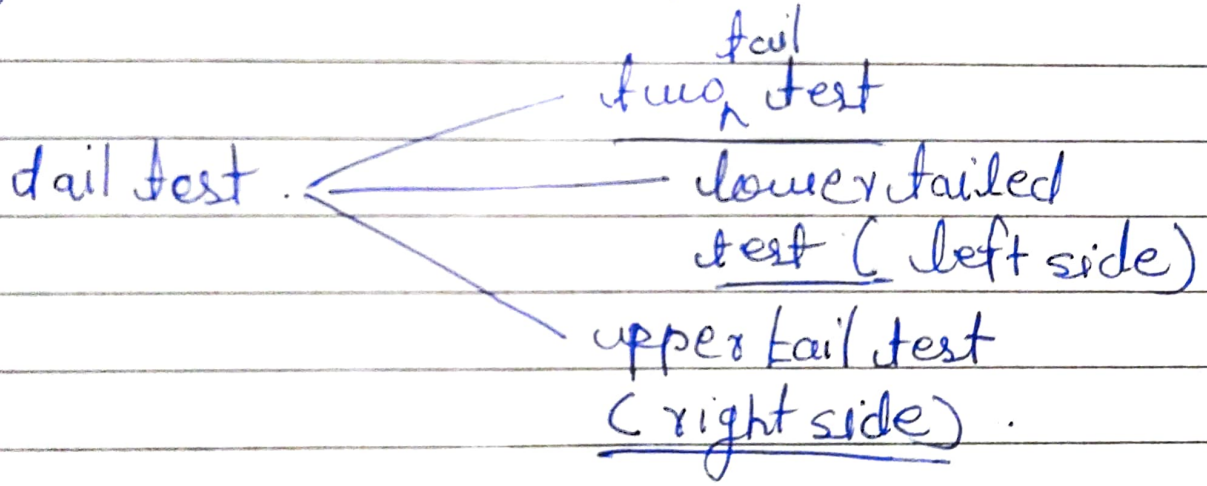
significance level (5% or 0.05)



Critical value methods.

a value in which where you accept the hypothesis or reject the hypothesis.
(same as p-value)

if value come under critical value we Reject otherwise we accept.



$$\star \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$\sigma = \frac{\sigma}{\sqrt{n}}$$

μ = mean.

Critical values formula

$$\mu + (\text{Z-score} \times \sigma_{\bar{x}})$$

↓

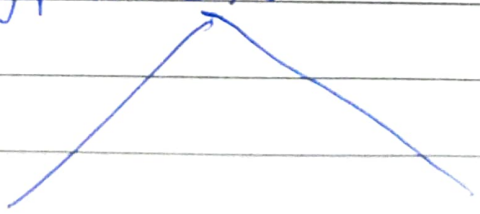
add + for (upper critical value)

or

add − for (lower critical value)

Conclusion

there two method to accept or reject the hypothesis



P-value method

Critical value Method

↓

P-value is probability of null hypothesis being true.

≤ 0.01 → very strong

$0.01 < \text{P-val} \leq 0.05$ strong

$0.05 \leq \text{P-val} \leq 0.1$ mild.

DATE / /

error in hypothesis testing

① type I error
occur when the null hypothesis is true but we reject.

② type 2 error
occur when the null hypothesis is false but we fail to reject it.

★ to determin hypothesis testing we do chi square or Anova or Z but which one to use?

	categorical	continuous
categorical	chi square (for two colum)	two sample t-test (two cat) Anova test (if more than two cat)
continuous.	logistic regression.	correlation

Hence the which the help of these test value we will compare with por critical value and we reject or accept the Hypothesis.

DATE / /

* what is difference between mean and proportion / categorical

mean / continuous.

* regression test will tell you two variable have association or correlation with them or not. It will tell you how correlated or not.

* Chi Square test

you should use this test when there is two categorical variable.

* 1-way Anova.

we will use this test when there is N-sample independent variable.