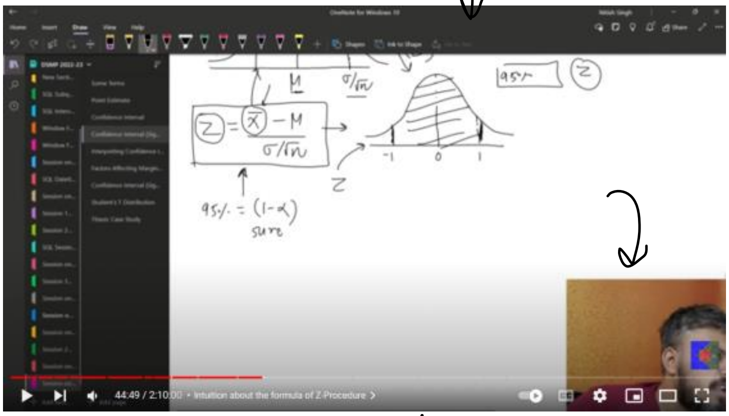
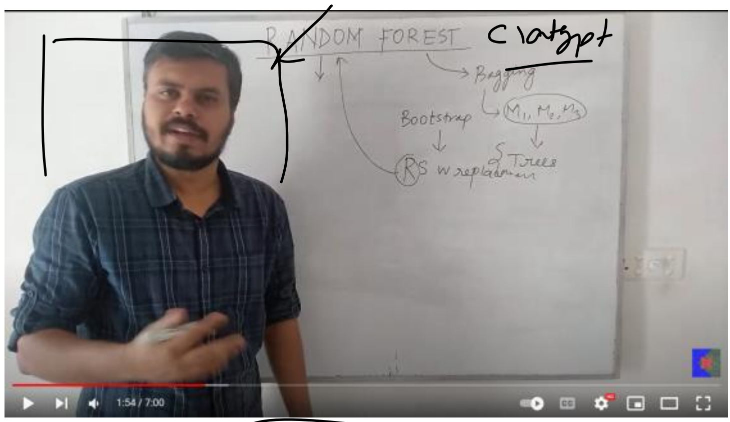
Hypothesis Testing

04 April 2023 07:04

a+b=c

1 video → 13 mins





chatgpt

6mins ← Days

μ

green

purple

5 videos

7 9 5 11 13

A statistical hypothesis test is a method of statistical inference used to decide whether the data at hand sufficiently support a particular hypothesis. Hypothesis testing allows us to make probabilistic statements about population parameters.

why?

Example in the video - nitish sir avg duration of watch time is 6 min , he wants to increase this avg time , after researching he got to know that these famouse youtubers use normal board whereas nitish sir uses Digital tools. So to know that the normal board method actually will have a more acg watch time or not , he made a one video using normal board method , and the avg watch time on that video was 12 min ,. Now the ques is that will that be sufficient to prove the second method is actually better , maybe the reason for increased time for that particular video maybe sir was looking more handsome , or the topic was trending. So this will not be sufficient. So what he did is he made 4 more videos with the famous method . Now duration lets say was 8 , 9 , 7 , 5 , 12. Now the ques is will that be sufficient to prove? the ans is still no. That's where hypothesis testing comes

In Hypothesis testing prove or reject some hypothesis, in our case is that new technique will increase avg watch time , null hypothesis says there will be no effect/significant change . So the null hypothesis in this case will be that even after adopting new technique , the avg watch time is same. And the alternate hypothesis would be that there is a change in avg watch time after adopting new technique.

Null and Alternate Hypothesis

04 April 2023 07:09

new shooting
more avg view
duration

$H_0: \mu = 6min$ ✓
$H_1: \mu > 6 mins$ ✓

Ho

1. **Null hypothesis (H0)**:

In simple terms, the null hypothesis is a statement that assumes there is no significant effect or relationship between the variables being studied. It serves as the starting point for hypothesis testing and represents the status quo or the assumption of no effect until proven otherwise. The purpose of hypothesis testing is to gather evidence (data) to either reject or fail to reject the null hypothesis in favour of the alternative hypothesis, which claims there is a significant effect or relationship.

$H_0: \mu = 100gms$ ✓
$H_a: \mu \neq 100gms$ ✓

the first one means that our null hypothesis says the avg watch time is 6 mins.

Null -
Altenate →

H_1 H_a

2. **Alternative hypothesis (H1 or Ha):**

The alternative hypothesis, is a statement that contradicts the null hypothesis and claims there is a significant effect or relationship between the variables being studied. It represents the research hypothesis or the claim that the researcher wants to support through statistical analysis.

The second means our Alternative hypothesis says that avg watch time is more than 6 mins

Ho: ✓

Important Points

- How to decide what will be Null hypothesis and what will be Alternate Hypothesis(Typically the Null hypothesis says nothing new is happening)
- We try to gather evidence to reject the null hypothesis
- It's important to note that failing to reject the null hypothesis doesn't necessarily mean that the null hypothesis is true; it just means that there isn't enough evidence to support the alternative hypothesis.

Hypothesis tests are similar to jury trials, in a sense. In a jury trial, H0 is similar to the not-guilty verdict, and Ha is the guilty verdict. You assume in a jury trial that the defendant isn't guilty unless the prosecution can show beyond a reasonable doubt that he or she is guilty. If the jury says the evidence is beyond a reasonable doubt, they reject H0, not guilty, in favour of Ha , guilty.

Example

so we try to reject null hypothesis , if we reject it automatically means alternative is correct. rather than proving altenative we try to reject null hypothesis

lets say we fail to reject null hypothesis , that doesn't mean null hypothesis was true , maybe there wasnt enough evidence to support alternative hypothesis

Steps involved in Hypothesis Testing

04 April 2023 13:25

Rejection Region Approach → basic → p value approach

$H_0: \mu = 6 mins$ → reject

$H_a: \mu > 6 mins$ →

1. Formulate a Null and Alternate hypothesis
2. Select a significance level(This is the probability of rejecting the null hypothesis when it is actually true, usually set at 0.05 or 0.01)
3. Check assumptions (example distribution) → σ
4. Decide which test is appropriate(Z-test, T-test, Chi-square test, ANOVA)
5. State the relevant test statistic
6. Conduct the test
7. Reject or not reject the Null Hypothesis.
8. Interpret the result

0.05 or 0.01

→ 5% → 1%

normally distribute

σ given → Z-test

t-test

z-test → Z-score

t-test → t-value

step 2 - significant level generally 5% (0.05) or 1% (0.01) , meaning significant level ka ye hai agr aap 100 bar test perform kr rhe ho , although null hypothesis is true , but still apne use 5 baar , ya 1 bar reject kr dia

step 3 - check data distribution , parameters(population std hai ya nhi), type of data , working with single column or multiple , based on these assumptions we decide test in next step

step 5 , Z - test ka statistic is Z - score , T-test ka statistic is T-score , based on these values we decide whether to reject or not reject null hypothesis

Performing a Z test Example 1

04 April 2023 07:15

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 training program. The average productivity was 50 units per day with a known population standard deviation of 5 units. After implementing the training program, the company measures the productivity of a random sample of 30 employees. The sample has an average productivity of 53 units per day. The company wants to know if the new training program has significantly increased productivity.

$\mu = 50 \quad \sigma = 5$

$n = 30 \quad \bar{X} = 53$

we have population mean = 50 , population std = 5 , sample size = 30 , sample mean(X bar) = 53.

1) $H_0: \mu = 50 \qquad H_a: \mu > 50$

2) $\alpha = 0.05 \rightarrow 5\%$

3) Normality valid / pop std (σ) known → Z test

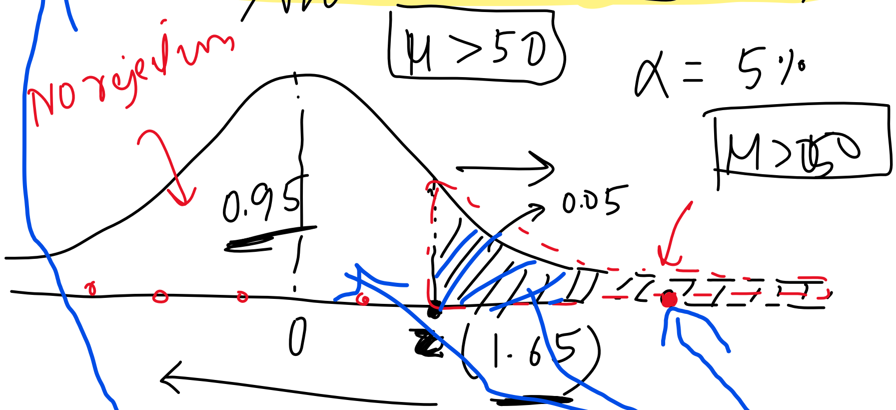
Assumptions - Since sample size >30 , we can say acc to CLT distribution will be normal/pop std known hence , test = Z-test.

4) Z test

5) Z

$$\boxed{Z} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{53-50}{5/\sqrt{30}} = \frac{3}{5/\sqrt{30}} = \boxed{3.28}$$

Rejection



Reject the null hypothesis

$\boxed{>50}$

now since , n = 30 , we can say the sample distribution will be normal , we cam covert into standard normal variate by this formula. Z statistic value is 3.38. Since we convert it into standard normal variate , it will look like this

Our Alternative hypothesis is saying that our mu(mean) should be greater than 50. And the alpha is 5%. You will look for Z jiska right wala area 5% and left wala will be 95% (This area/this value).. Use Z table look for value whose area under the graph is 0.95 , after checking the Z value is 1.65.
Ab , you will look for 3.28. we know that at that curve 3.28 lie at some this point

Now there us concept of rejection region that after the Z value more than 1.65 as alpha = 0.05 , (that falls under 5 percent region) will be rejection region if the sample statistic Z value more than 1.65 . That means the null hypothesis is rejected. or we can say we have strong evidence against null hypothesis.refer to chatgpt explaination in the folder

Example 2

04 April 2023 16:06

Z test

Suppose a snack food company claims that their Lays wafer packets contain an average weight of 50 grams per packet. To verify this claim, a consumer watchdog organization decides to test a random sample of Lays wafer packets. The organization wants to determine whether the actual average weight differs significantly from the claimed 50 grams. The organization collects a random sample of 40 Lays wafer packets and measures their weights. They find that the sample has an average weight of 49 grams, with a known population standard deviation of 4 grams.

$\mu = 50 \quad n = 40 \quad \bar{x} = 49 \quad \sigma = 4$

1) $H_0: \mu = 50 \quad H_a: \mu \neq 50$

2) $\alpha = 0.05$

3) Normality ✓ σ ✓ → Z test

4) Z test

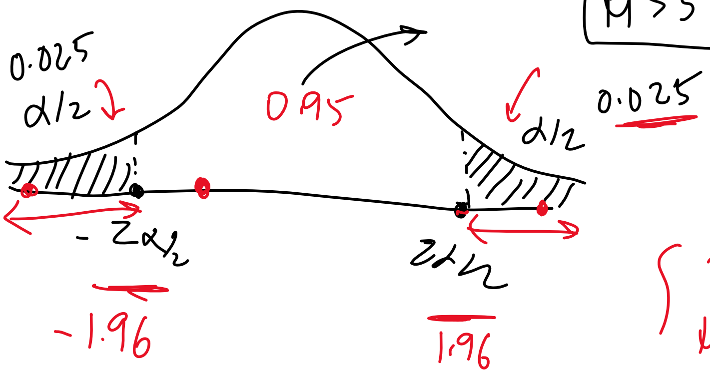
5) Z

6) $Z = \dfrac{49 - 50}{4/\sqrt{40}} = \dfrac{-\sqrt{40}}{4} = \boxed{-1.58}$

$\alpha = 5\%$

$\boxed{\mu > 50}$ $\quad \mu \neq 50$

can't reject the Null hypothesis



$\mu \neq 50$

$\{\mu > 50, \ \mu < 50\}$

IIn the previous ques we had an assumption of mean > 50 hence one tail test , but in this question our assumption is ha: mean not equal 50 , hence there will be two tailed test. Hence now we have to divide the alpha equally. hence there will be two rejection area with alpha/2. 2.5 percent area on left 2.5 area on right . find Z values for both. Agr beech me lie kregea , null hypothesis ko reject nhi kr payenge. Look at the Z table

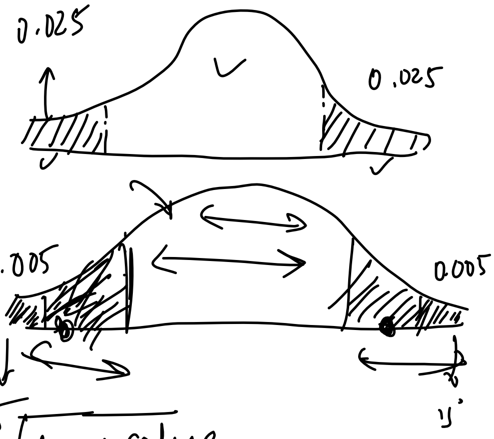
jThis area is .095 + 0.025 = 0.975 now look for Z value whose area is 0.975 Z value is 1.96(Right side)
Since it is a normal distribution both are identical the left Z value will be -1.96.
and our Z value for sample statistic is -1.58 which doesn't come under the rejection region , hence we do nt have strong evidence against null hypothesis.

Rejection Region

04 April 2023 16:21

$\mu \neq 50$ 30 $\boxed{1\%} \leftarrow \boxed{5\%}$ $\alpha \downarrow$ $\alpha \uparrow$

Significance level - denoted as α (alpha), is a predetermined threshold used in hypothesis testing to determine whether the null hypothesis should be rejected or not. It represents the probability of rejecting the null hypothesis when it is actually true, also known as Type 1 error.

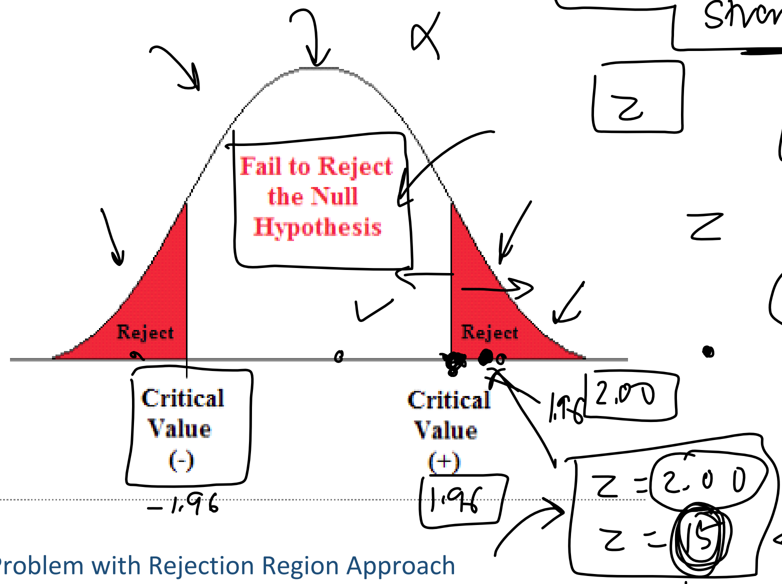


The ~~critical~~ rejection region is the region of values that corresponds to the rejection of the null hypothesis at some chosen probability level.

evidence strength → p-value

Null reject

p-value



z 0.01 1.95 1.97 z = 15 2.00

z = 2.00
z = 15

Problem with Rejection Region Approach

when it is actually true is a bit confusing.lets understand when you increase the alpha the rejection region increases , since the alpha increased to let say 50 , there might be a chance we will be rejecting the null hypothesis even if it is true.(Also Known as Type 1 Error) thats why generally significant level is 5% or 1%.

Important point - Problem with rejection approach.
There are two problems with rejection approach -
1. Lets say our Z value is 1.97 and critical point is 1.96 just because it is just 0.01 more , we rejected the null hypothesis and vice versa.
2. lets say out Z is 15.5 , which is very rare (very far from the critical point) at one side Z is 4 , for rejection region both are same , there is no difference between them.
In short , rejection approach isn't able to tell the strength of evidence. How strong our evidence is to reject the null hypothesis.

Hence introduction of new approach ie P - value approach , where we calculate additional thing that is p-value which is able to tell the strength of an evidence against null hypothesis.

Type 1 vs Type 2 Error

04 April 2023 13:29

hypothesis → H_0 ✗ → ✓

H_0 ✓ → ✗

$\boxed{\alpha =}$

In hypothesis testing, there are two types of errors that can occur when making a decision about the null hypothesis: Type I error and Type II error.

$\boxed{H_0 \rightarrow \text{no crime}}$

significance

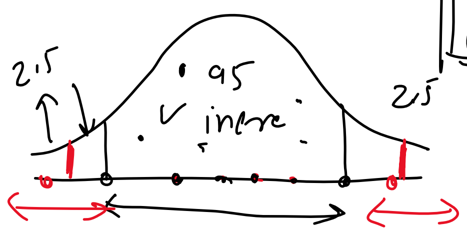
Type-I (False Positive) error occurs when the sample results, lead to the rejection of the null hypothesis when it is in fact true.

In other words, it's the mistake of finding a significant effect or relationship when there is none. The probability of committing a Type I error is denoted by α (alpha), which is also known as the significance level. By choosing a significance level, researchers can control the risk of making a Type I error.

Decision based on sample	Truth about the population: H_0 true	Truth about the population: ~~H_0 true~~ H_0 false
Reject H_0	Type I error	Correct decision
Accept H_0	Correct decision	Type II error

$\alpha = 5$ ↓

$\boxed{\alpha}$ ↓



Type-II (False Negative) error occurs when based on the sample results, the null hypothesis is not rejected when it is in fact false.

This means that the researcher fails to detect a significant effect or relationship when one actually exists. The probability of committing a Type II error is denoted by β (beta).

(β) → power

$\boxed{1-\beta}$

Trade-off between Type 1 and Type 2 errors

$\boxed{\alpha = 0.05}$

Type 1 Error - ek suspect hai we know usne crime mhi kia , Pr court me prove ho gya usne crime kia hai (False Positive)
H0 : No Crime , H1:Crime , H0 is rejected , Hence H1 is True ,. The probability of type 1 error depends on the significance level(alpha) if alpha is 5 , then it is divided both sides 2.5 left , 2.5 right , what if we decrease the significance level the then , since we know null hypothesis true , the chances of being h0 proved true in court will be high.

Type 2-error - ek court me banda hai unhe asal me crime kia hai , which means our alternative hypothesis is actually true , h0:no Crime , h1:Crime , but no strong evidence hence in the eye of court h0 true , but it is actually false , hence False Negative we were not able to reject null hypothesis when it is actually false.
The probability pdf commiting type 2 error is denoted by Beta.

Now the imp thing is that there is tradeoff between type 1 and type 2 , agar hum chahte hai type 1 error na ho , isle lie alpha ka value ghatane lage , ise kya hoga rejection region cohota ho jaega , isse humra current suspect jo ki criminal nhi hai uske chance bad ggye court me na proof hone ke pr ise ek problemhuyi , lets say future me aisa banda hai jisne crime kia , lekin since humne alpha ghataya hua to uske km chances hai rejection region me girne ke , jise vo bhi criminal proof nhi hoega , to dekha type 1 error ko ghatane , se type 2 error bad gya hence the tradeoff

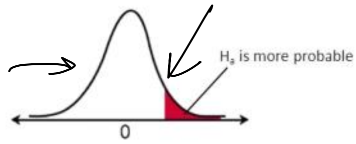
Example , we need more productivity than before , we need less expense than before that's when you use one sided tailed test

One sided vs two sided test

04 April 2023 13:29

one sided

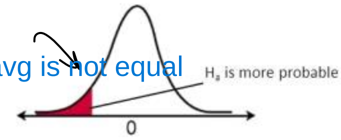
One-sided (one-tailed) test: A one-sided test is used when the researcher is interested in testing the effect in a specific direction (either greater than or less than the value specified in the null hypothesis). The alternative hypothesis in a one-sided test contains an inequality (either ">" or "<").



Right-tail test

H_a: $\mu >$ value

Example: A researcher wants to test whether a new medication increases the average recovery rate compared to the existing medication.

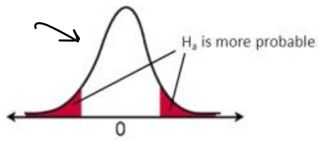


Left-tail test

H_a: $\mu <$ value

lays example , we want to know if lays avg is not equal

Two-sided (two-tailed) test: A two-sided test is used when the researcher is interested in testing the effect in both directions (i.e., whether the value specified in the null hypothesis is different, either greater or lesser). The alternative hypothesis in a two-sided test contains a "not equal to" sign (≠).



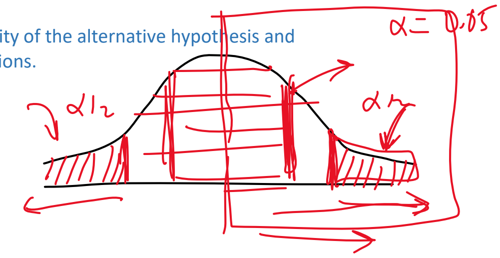
Two-tail test

H_a: $\mu \neq$ value

Example: A researcher wants to test whether a new medication has a different average recovery rate compared to the existing medication.

The main difference between them lies in the directionality of the alternative hypothesis and how the significance level is distributed in the critical regions.

$\alpha = 0.05$



Advantages and Disadvantages?

Two-tailed test (two-sided):

Advantages:

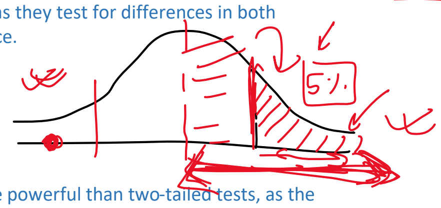
1. Detects effects in both directions: Two-tailed tests can detect effects in both directions, which makes them suitable for situations where the direction of the effect is uncertain or when researchers want to test for any difference between the groups or variables.
2. More conservative: Two-tailed tests are more conservative because the significance level (α) is split between both tails of the distribution. This reduces the risk of Type I errors in cases where the direction of the effect is uncertain.

Disadvantages:

power = $\boxed{1-\beta}$

1. Less powerful: Two-tailed tests are generally less powerful than one-tailed tests because the significance level (α) is divided between both tails of the distribution. This means the test requires a larger effect size to reject the null hypothesis, which could lead to a higher risk of Type II errors (failing to reject the null hypothesis when it is false).
2. Not appropriate for directional hypotheses: Two-tailed tests are not ideal for cases where the research question or hypothesis is directional, as they test for differences in both directions, which may not be of interest or relevance.

$\mu \neq 100$gms



One-tailed test (one-sided):

Advantages:

1. More powerful: One-tailed tests are generally more powerful than two-tailed tests, as the entire significance level (α) is allocated to one tail of the distribution. This means that the test is more likely to detect an effect in the specified direction, assuming the effect exists.
2. Directional hypothesis: One-tailed tests are appropriate when there is a strong theoretical or practical reason to test for an effect in a specific direction.

Disadvantages:

1. Missed effects: One-tailed tests can miss effects in the opposite direction of the specified alternative hypothesis. If an effect exists in the opposite direction, the test will not be able to detect it, which could lead to incorrect conclusions.
2. Increased risk of Type I error: One-tailed tests can be more prone to Type I errors if the effect is actually in the opposite direction than the one specified in the alternative hypothesis.

Where can be Hypothesis Testing Applied?

04 April 2023 07:15

Z test, t test

1. Testing the effectiveness of interventions or treatments: Hypothesis testing can be used to determine whether a new drug, therapy, or educational intervention has a significant effect compared to a control group or an existing treatment.

percentage -1 to 1

pearson

t-test

2. Comparing means or proportions: Hypothesis testing can be used to compare means or proportions between two or more groups to determine if there's a significant difference. This can be applied to compare average customer satisfaction scores, conversion rates, or employee performance across different groups.

correlation

3. Analysing relationships between variables: Hypothesis testing can be used to evaluate the association between variables, such as the correlation between age and income or the relationship between advertising spend and sales.

chi-square

4. Evaluating the goodness of fit: Hypothesis testing can help assess if a particular theoretical distribution (e.g., normal, binomial, or Poisson) is a good fit for the observed data.

titanic

gender	survi-
M	1
F	0

5. Testing the independence of categorical variables: Hypothesis testing can be used to determine if two categorical variables are independent or if there's a significant association between them. For example, it can be used to test if there's a relationship between the type of product and the likelihood of it being returned by a customer.

6. A/B testing: In marketing, product development, and website design, hypothesis testing is often used to compare the performance of two different versions (A and B) to determine which one is more effective in terms of conversion rates, user engagement, or other metrics.

Hypothesis Testing ML Applications

04 April 2023 16:50

$f_1\ f_2\ f_3\ \ldots$

1. $\boxed{\text{Model comparison}}$: Hypothesis testing can be used to compare the performance of different machine learning models or algorithms on a given dataset. For example, you can use a paired t-test to compare the accuracy or error rate of two models on multiple cross-validation folds to determine if one model performs significantly better than the other.

2. Feature selection: Hypothesis testing can help identify which features are significantly related to the target variable or contribute meaningfully to the model's performance. For example, you can use a t-test, chi-square test, or ANOVA to test the relationship between individual features and the target variable. Features with significant relationships can be selected for building the model, while non-significant features may be excluded.

3. Hyperparameter tuning: Hypothesis testing can be used to evaluate the performance of a model trained with different hyperparameter settings. By comparing the performance of models with different hyperparameters, you can determine if one set of hyperparameters leads to significantly better performance.

4. Assessing model assumptions: In some cases, machine learning models rely on certain statistical assumptions, such as linearity or normality of residuals in linear regression. Hypothesis testing can help assess whether these assumptions are met, allowing you to determine if the model is appropriate for the data.