Applied Statistics for Data Scientists with R

Class 13: Hypothesis Testing Fundamentals

Steps of Hypothesis Testing

Step 1: State null and alternative hypotheses

Step 2: Identify Test Statistic

Step 3: Specify Level of Significance (α, usually 0.05)

Step 4: Formulate decision rule (Reject H_0 if $p < \alpha$=0.05)

Step 5: Compute test statistic

Step 6: Take Decision

What is Hypothesis?

- It is a claim about a population parameter.
- We cannot examine the entire population, and so we take sample.
- The goal is to infer something about the entire population based on this sample.
- Due to sampling, there are chances of error or deviation from the actual scenario of the data.

Why Do We Test Hypothesis?

- To see the difference between the calculated Statistic and the Parameter we test hypothesis.

↓

Estimated sample mean
Estimated sample variance
Estimated regression model coefficients
And so on.

Remarks:

- In the 1920s, Ronald Fisher developed the theory behind the p value and Jerzy Neyman and Egon Pearson developed the theory of hypothesis testing. *Source: https://doi.org/10.1007/s11999-009-1164-4*
- The incident of RA Fisher discovering the framework of hypothesis testing: https://shire.science.uq.edu.au/CONS7008/_book/introduction-to-hypothesis-testing.html#hypothesis-testing

Why Do We Test Hypothesis?

Compare hypothesis testing to a courtroom trial:

H_0: Defendant is innocent.

against H_1: Defendant is guilty.

- Aim is to see whether the person is guilty or not.
- So, find evidence to reject the null hypothesis.
- The null hypothesis can be rejected only if enough evidence is found against the null hypothesis.
- And only open rejecting the null, we can say the alternative hypothesis is true.
- Notice that, not finding enough evidence, does not mean that the person did not commit the crime.

"Not guilty" ≠ "innocent"

So, instead of accepting the null hypothesis, use the word “fail to reject the null”.

Statistical Hypothesis: Intuition

Layman's terms: “If the collected data contains strong evidence supporting a particular statement, we can make conclusions accordingly. However, if the data does not contain such evidence, it does not necessarily mean that the event did not occur. Instead, it simply means that, given the data at hand, we do not have enough statistical evidence to support the statement.”

Statistical Hypothesis Construction: Null vs Alternative

- Identify the research problem. For example, does a new drug reduce blood pressure more than the current drug?
- A statistical hypothesis test involves formulating two competing hypotheses about a population parameter:

1. Null hypothesis (H_0)
2. Alternative hypothesis (H_1 or H_a)

- Null hypothesis is constructed assuming that there is no effect, no difference, or no change in the population parameter.
- Alternative hypothesis is constructed assuming the claim you want to test (e.g., a difference, improvement, or relationship). Also called, research hypothesis.

Hypothesis Construction: One tailed vs Two tailed

- **Two-tailed test**: Tests for *any difference* ($H_a = parameter \neq value$).
- **One-tailed test**: Tests for a difference in a *specific direction* ($H_a = parameter > value$).

Two-tailed Test

- Tests for an effect or difference in any direction
- Alternative Hypothesis (H_1):
  a. $H_a = parameter \neq value$

- Example: To answer the research question "Does Fertilizer X have any effect on crop yield?"

$H_0: \mu_x - \mu_{no} = 0$ against $H_a: \mu_x - \mu_{no} \neq 0$

$\mu_x = \mu_{no}$ $\qquad$ $\mu_x \neq \mu_{no}$

Hypothesis Construction: One tailed vs Two tailed

- **Two-tailed test**: Tests for *any difference* (H_a: $parameter \neq value$).
- **One-tailed test**: Tests for a difference in a *specific direction* (H_a: $parameter > value$).

One-Tailed Test

- Tests for an effect or difference in a specific direction (e.g., "greater than" or "less than").
- Alternative Hypothesis (H_1):
  a. $H_a = parameter > value$ (right-tailed)
  b. $H_a = parameter < value$ (left-tailed)

- Example: To answer the research question "Does Fertilizer X increase crop yield on average by 5 t/h?"
  $H_0: \mu_x + 5 \leq \mu_{no}$ against $H_a: \mu_x + 5 > \mu_{no}$

Hypothesis Construction: Examples

(i) H_0: Population average of phone usage among high school kids in Mirpur = 6 hours

H_0: $\mu = 6\ hours$ against H_a: $\mu \neq 6\ hours$

(ii) H_0: The treatment does not increase SpO2 by 10% compared to no treatment.

H_0: $\mu_{treatment} - \mu_{base} \leq 0.10$ against H_a: $\mu_{treatment} - \mu_{base} > 0.10$

(iii) H_0: The true average yield of BRRI rice BR16 is 6 t/ha under all 4 fertilizer combinations.

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 6$

against H_1: At least one fertilizer combination's average yield ≠ 6 t/ha.

Errors in Hypothesis Testing

		Decision	
		Do not reject H_0	Reject H_0
Actual	H_0 is true	Correct Confidence level $(1-\alpha)$	Type I Error α = Probability of Type-I Error / Level of Significance
	H_0 is false	Type II Error β = Probability of Type-II Error	Correct Statistical power $(1-\beta)$

H_0: Defendant is innocent.
H_1: Defendant is guilty.

Errors in Hypothesis Testing

		Decision: Do not reject H_0	Decision: Reject H_0
Actual	H_0 is true	Correct Confidence level $(1-\alpha)$	Type I Error α = Probability of Type-I Error / Level of Significance
	H_0 is false	Type II Error β = Probability of Type-II Error	Correct Statistical power $(1-\beta)$

H_0: Defendant is innocent.
H_1: Defendant is guilty.

		Decision: Do not reject H_0	Decision: Reject H_0
Actual	H_0 is true	Innocent defendant freed Confidence level $(1-\alpha)$	Innocent defendant jailed α = Probability of Type-I Error / Level of Significance
	H_0 is false	Guilty defendant freed β = Probability of Type-II Error	Guilty defendant jailed Statistical power $(1-\beta)$

Selecting Level of Significance

H_0: Defendant is innocent.
H_1: Defendant is guilty.

		Decision	
		Do not reject H_0	Reject H_0
Actual	H_0 is true	Innocent defendant freed Confidence level $(1-\alpha)$	Innocent defendant jailed α = Probability of Type-I Error / Level of Significance
	H_0 is false	Guilty defendant freed β = Probability of Type-II Error	Guilty defendant jailed Statistical power $(1-\beta)$

- So, level of significance α is selected as minimum as possible to reduce type-I error. Represents the maximum risk you are willing to take of rejecting H_0 when it's true (Type I error).

  Ideally it is taken 0.1 (10%), 0.05 (5%) or 0.01 (1%)

What is p-value?

- Assume that the null hypothesis is true.
- Under this assumption, what is the probability of observing the data?
- Is it high or low?
- If lower than the significance level (α), then it is less likely that the null hypothesis is true.

"The p-value represents the conditional probability that we would arrive at our results (or more extreme results) given that the null hypothesis is true."

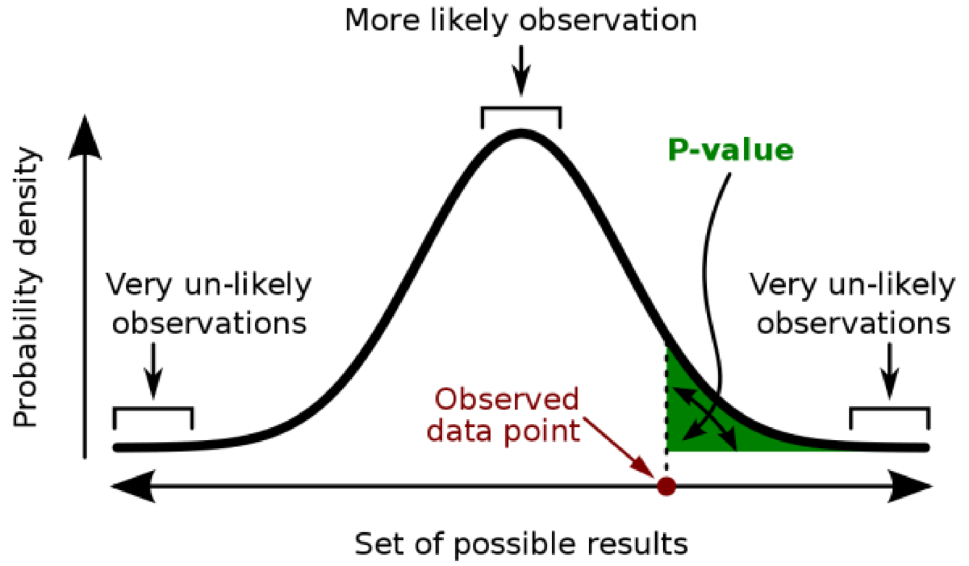
Decision

- If $p \leq \alpha$, the data is **too extreme** to be explained by random chance alone.

Conclusion:

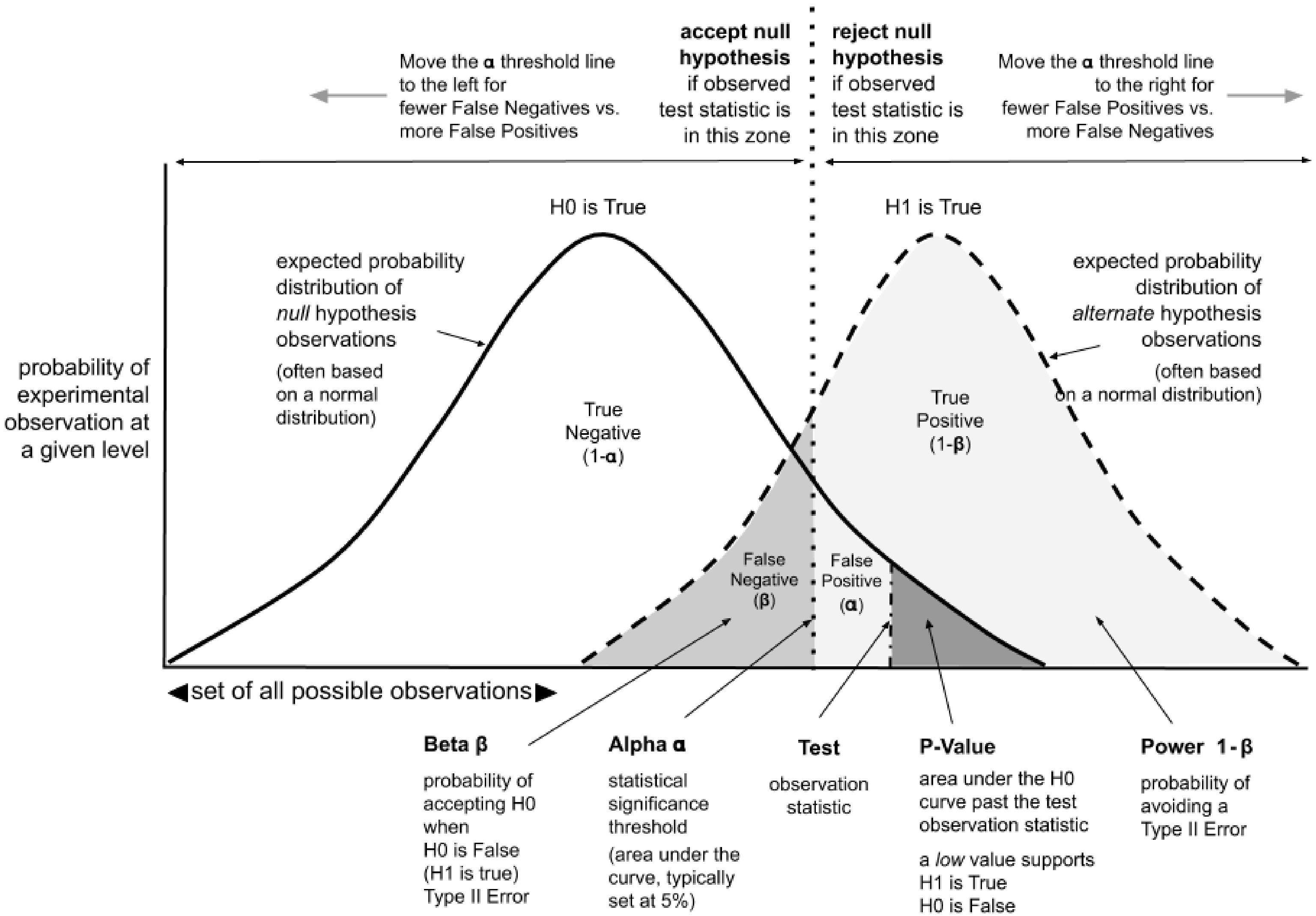
- If $p \leq \alpha$: Reject H_0.
  The data provides strong evidence against the null hypothesis.

- If $p > \alpha$: Fail to reject H_0.
  Insufficient evidence to reject the null hypothesis.

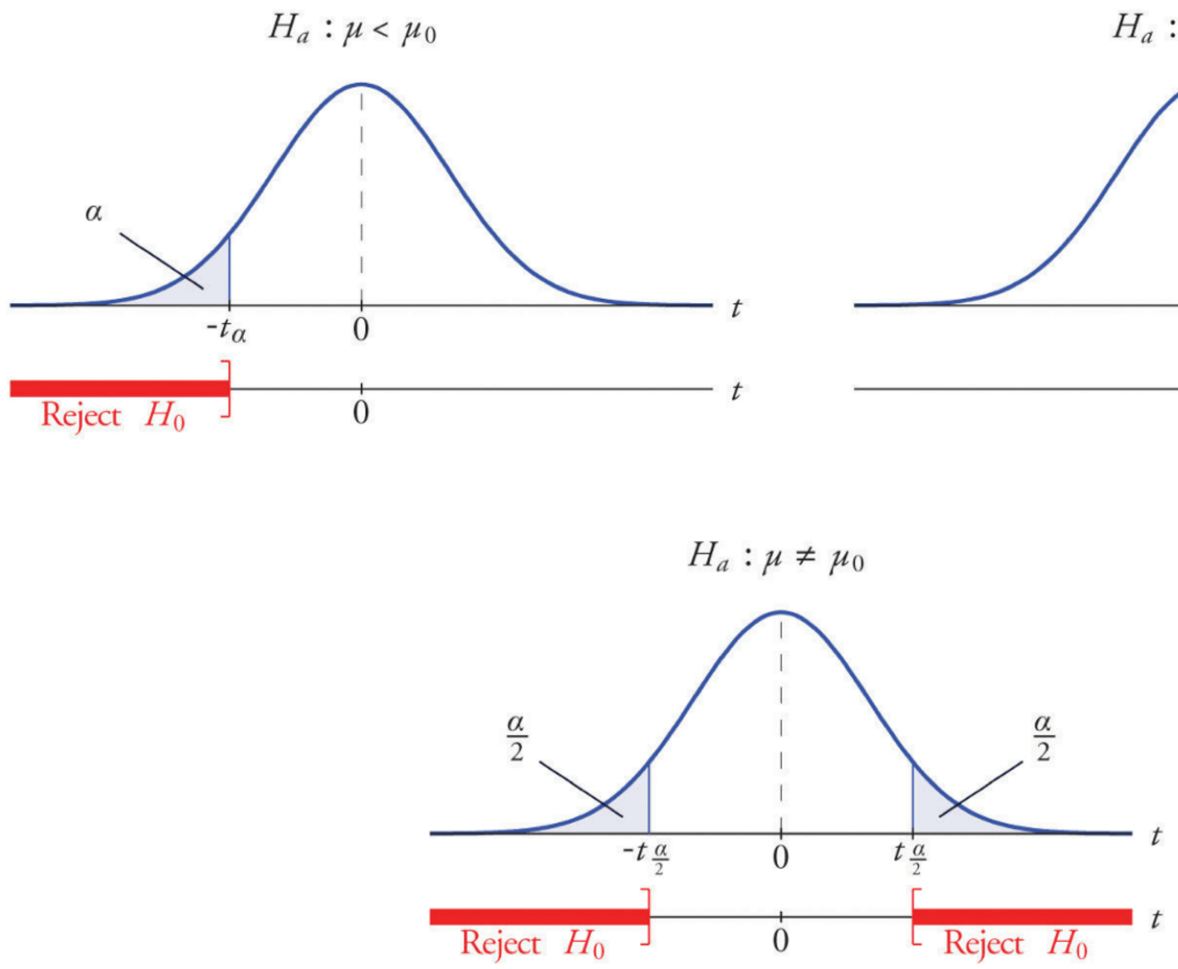


A **p-value** (shaded green area) is the probability of an observed (or more extreme) result assuming that the null hypothesis is true.

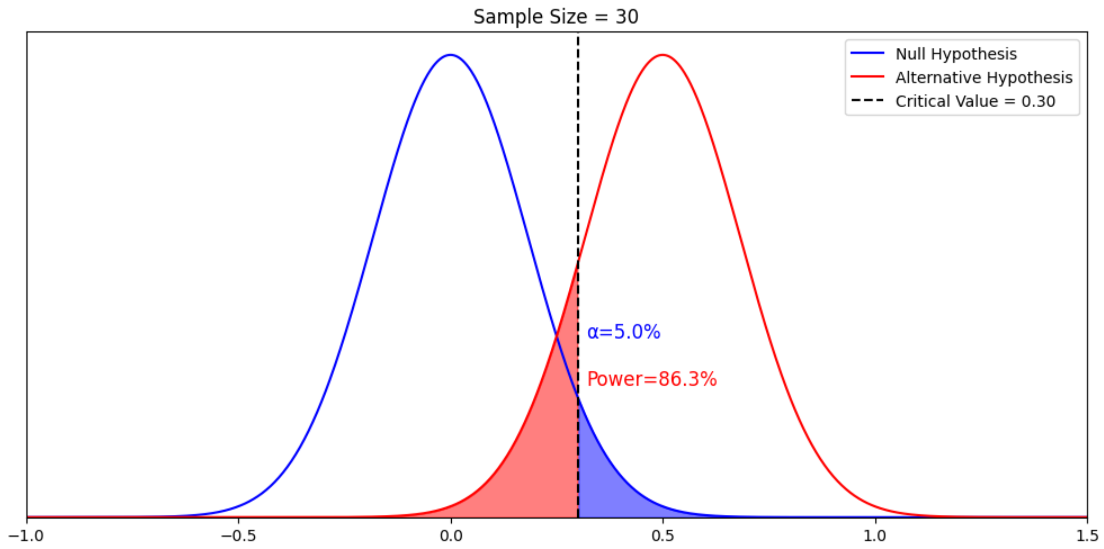
An Illustration

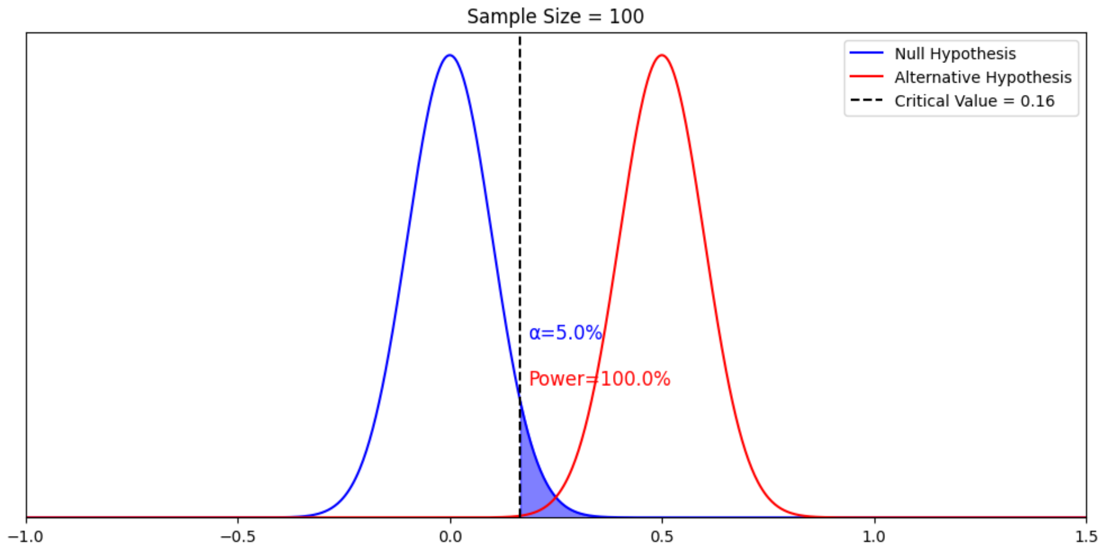


Level of Significance and One-tail, Two-tail tests

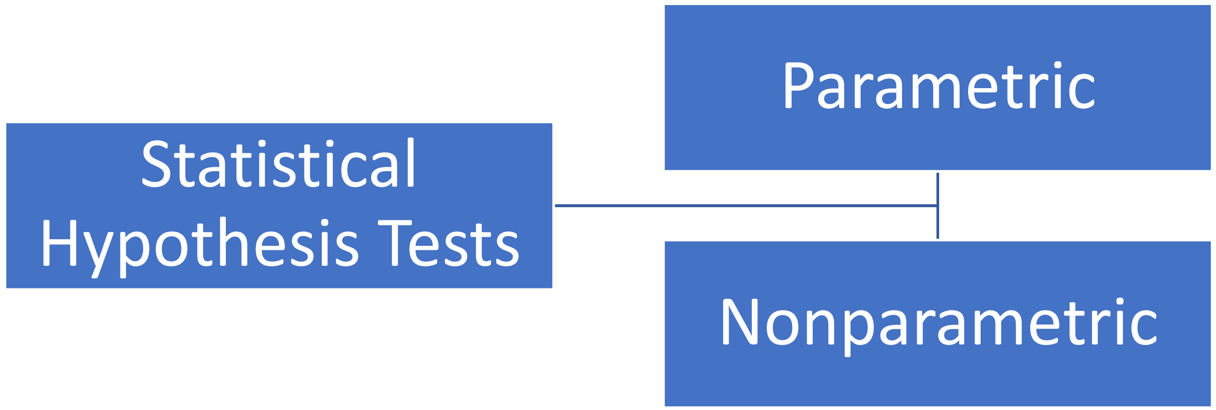


Sample Size and Statistical Hypothesis Test





Different Statistical Tests



Parametric tests

1. One sample t-test
2. Independent samples t-test
3. Paired samples t-test
4. One-way ANOVA
   i. Post-hoc analysis
   ii. Levene's test
5. Two-way ANOVA

Nonparametric tests

1. Kolmogorov-Smirnov test
2. Shapiro-Wilk test
3. One-sample Wilcoxon signed rank test / Median test
4. Mann-Whitney test
5. Wilcoxon-signed rank test
6. Kruskal-Wallis test