Data Science Bootcamp

<> Hypothesis Testing in Statistics

Essential Knowledge

Data Analyst

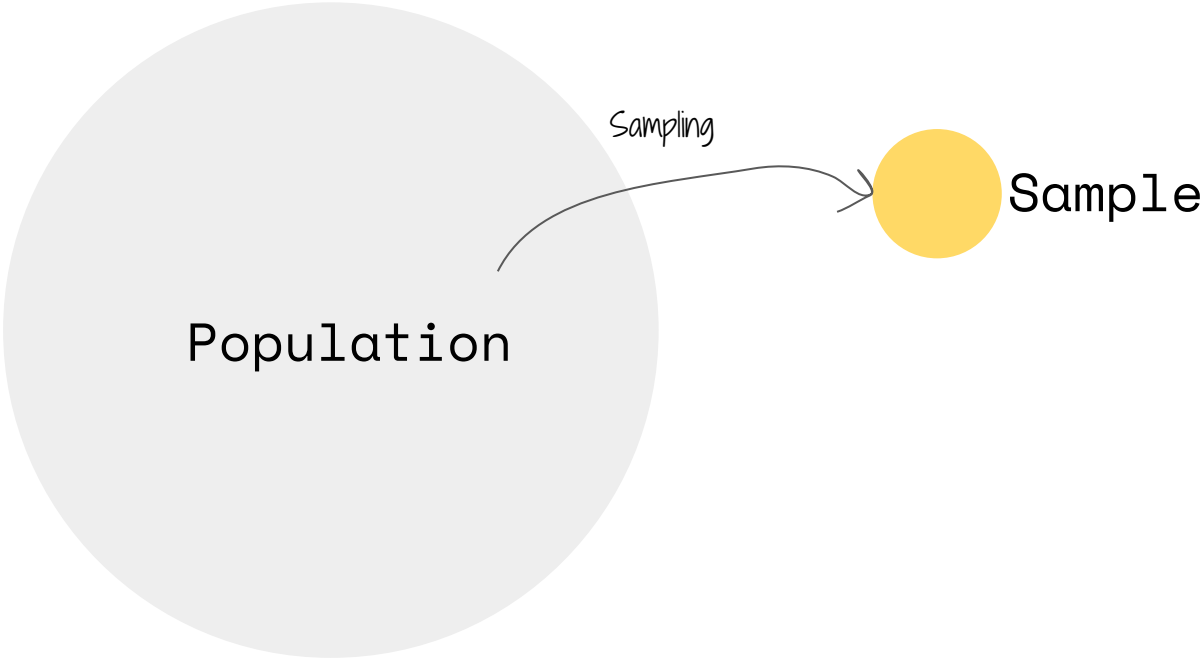
- Descriptive
- Inferential
  - Hypothesis Test/ Basic Model
    - T-Test
    - F-Test
    - Correlation
    - Linear Regression
    - Logistic Regression

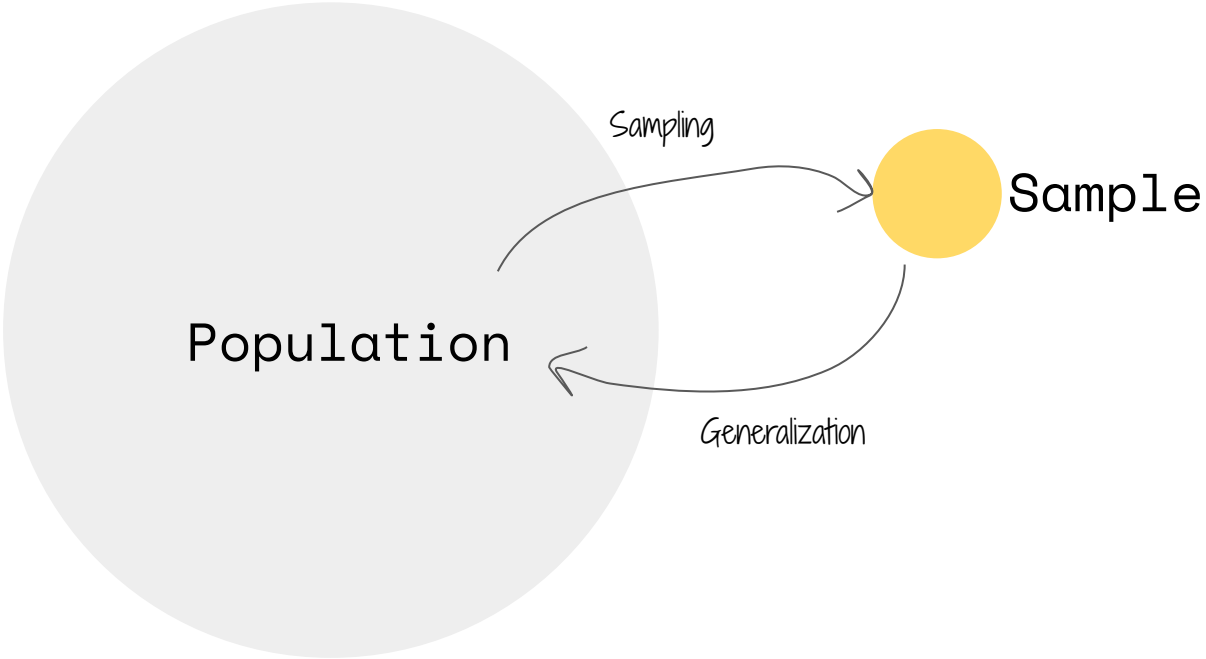
Content 🔥

- Review 3 Ways to do Hypothesis Tests
  - Critical Region
  - p-value
  - Confidence Interval
- Limitation of p-value

The earth is round
(p-value < 0.05)

Recap Core Concepts





https://www.investopedia.com › ... › Tools

Hypothesis Testing Definition - Investopedia

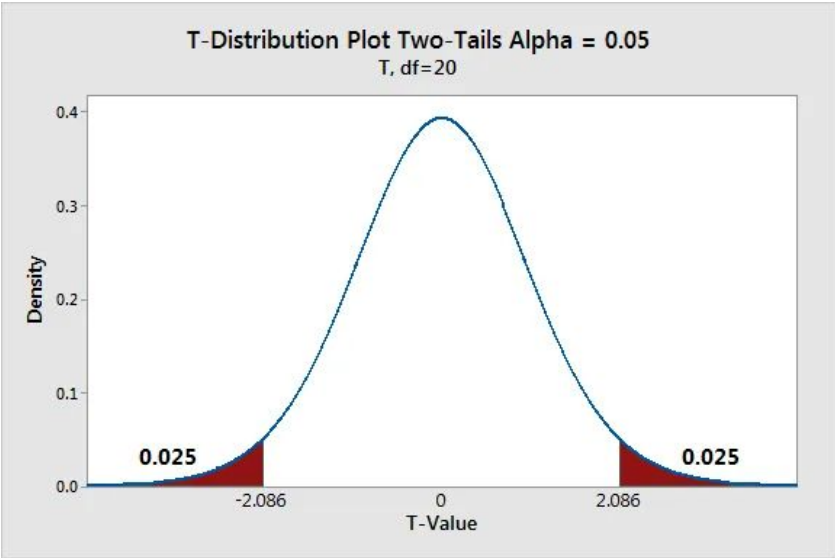
Hypothesis testing is **an act in statistics whereby an analyst tests an assumption regarding a population parameter**. The methodology employed by the analyst ...

Two Types of Testing

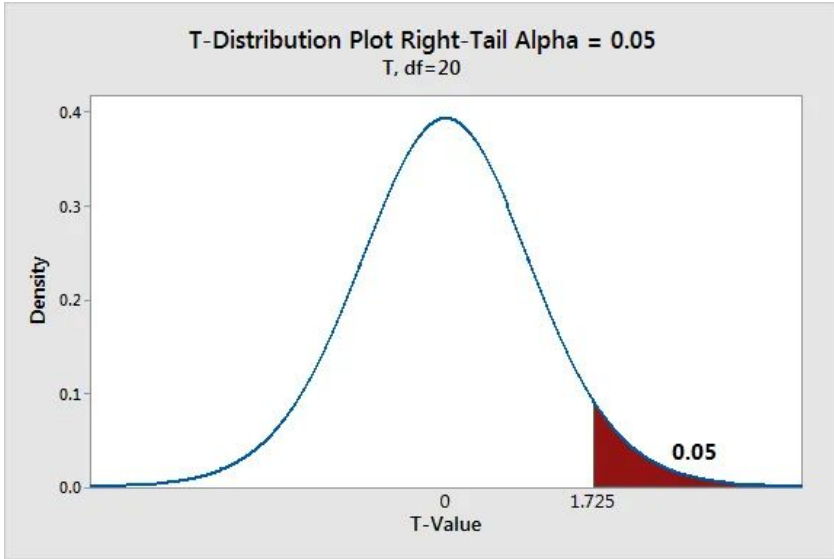
Hypothesis Tests

- One tailed test
- Two tailed test

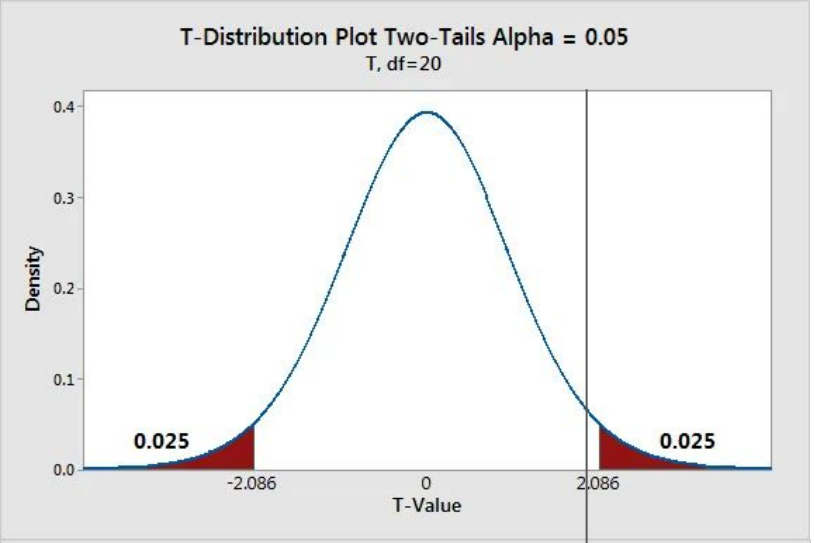
Two-Tailed Test

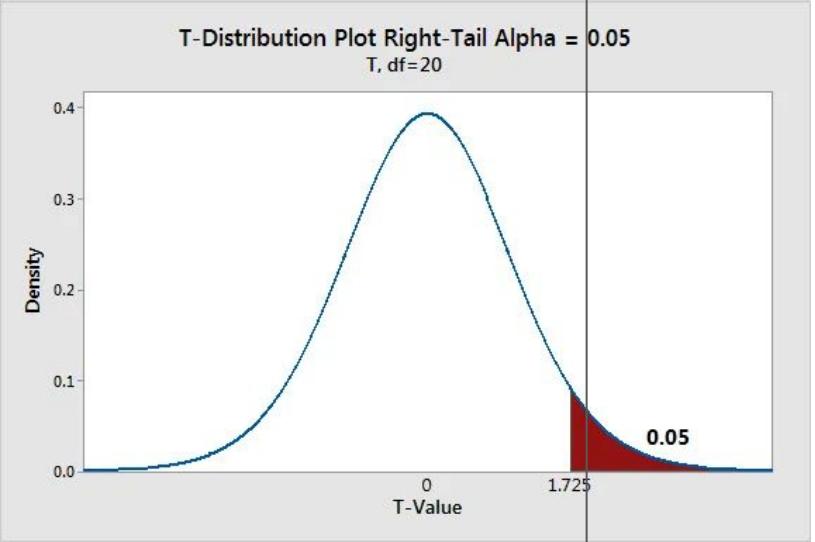


One-Tailed Test



https://statisticsbyjim.com/hypothesis-testing/one-tailed-two-tailed-hypothesis-tests/





Two tailed test

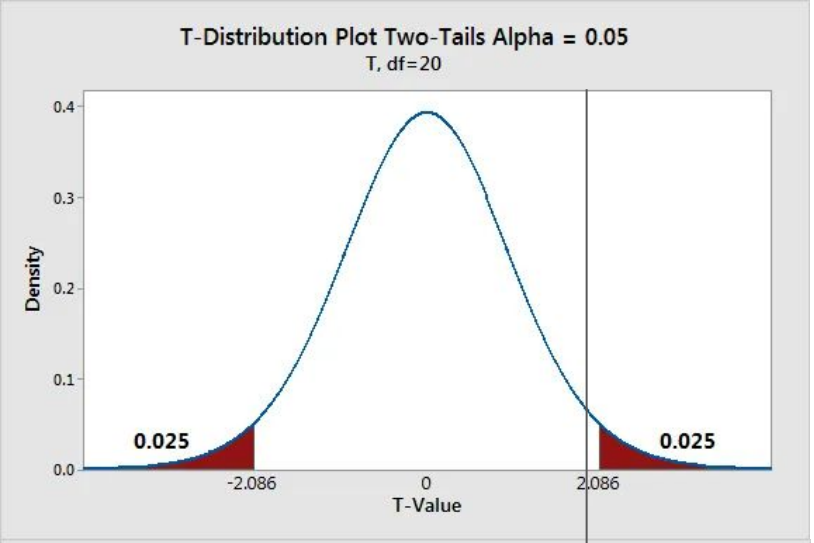
Ho: mean A - mean B = 0

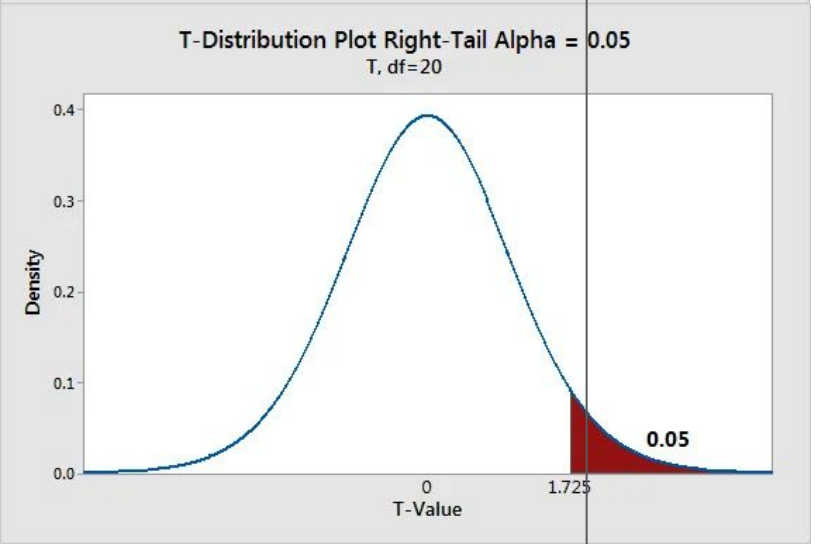
Ha: mean A - mean B != 0

One tailed test

Ho: mean A - mean B <= 0

Ha: mean A - mean B > 0





Two tailed test

Ho: mean A - mean B = 0

Ha: mean A - mean B != 0

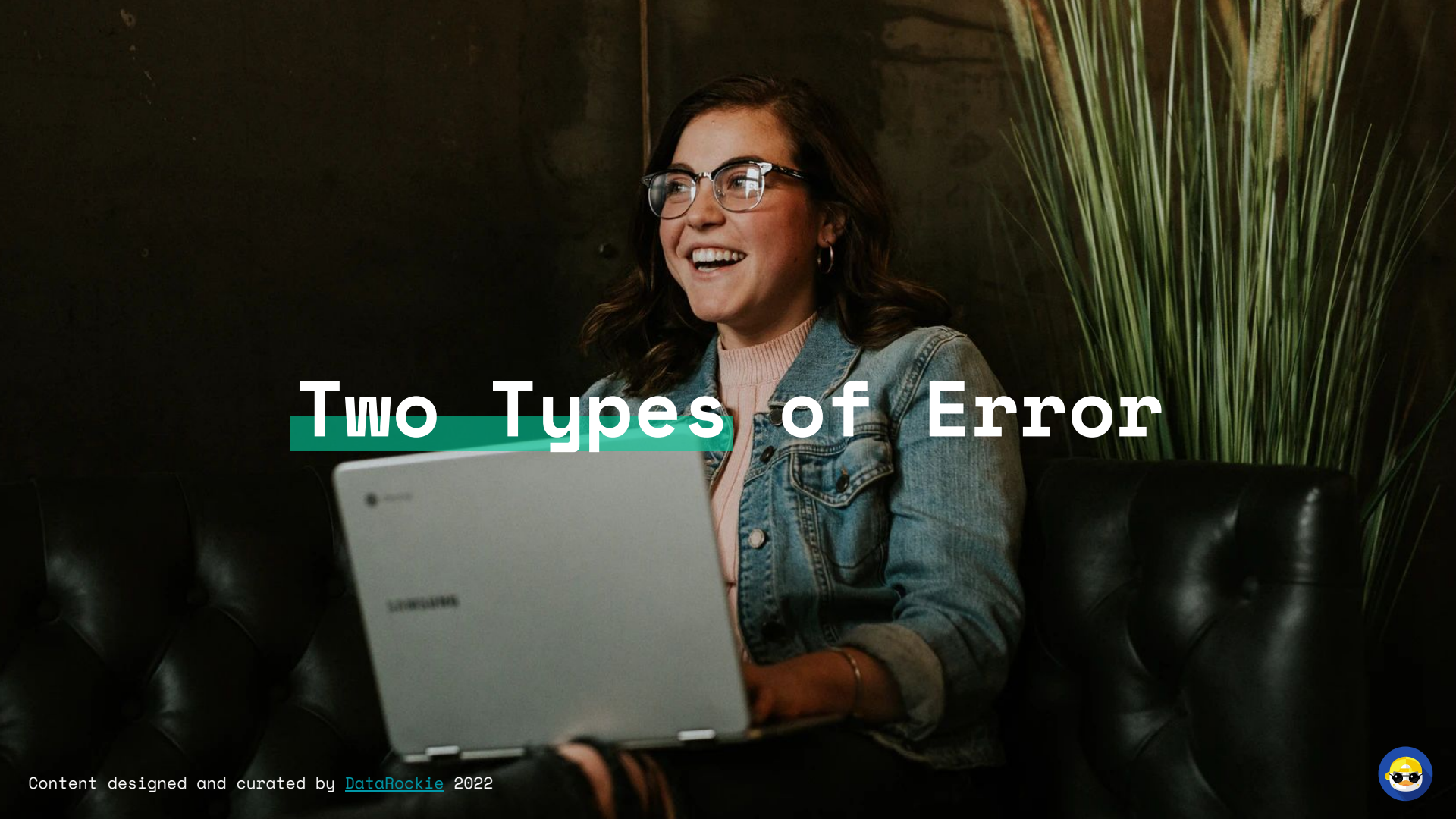
Conclusion: Fail to Reject Ho

One tailed test

Ho: mean A - mean B <= 0

Ha: mean A - mean B > 0

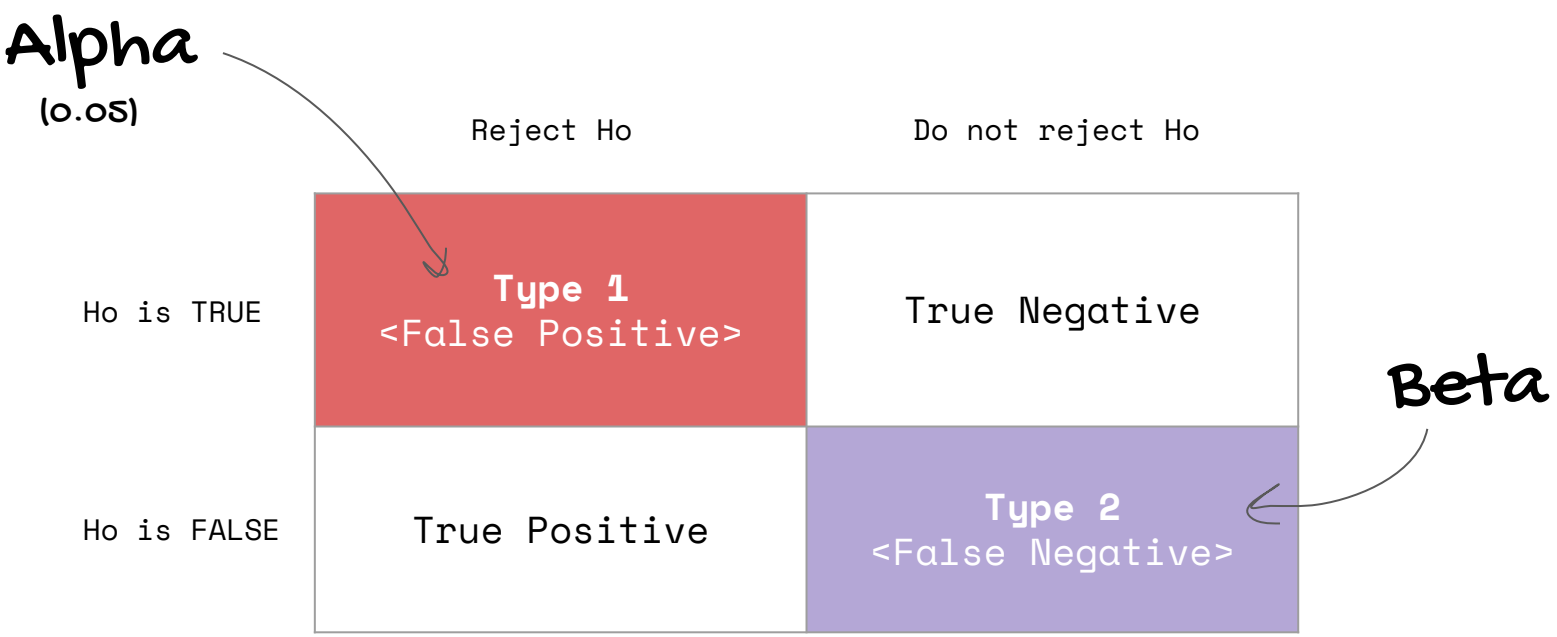
Conclusion: Reject Ho

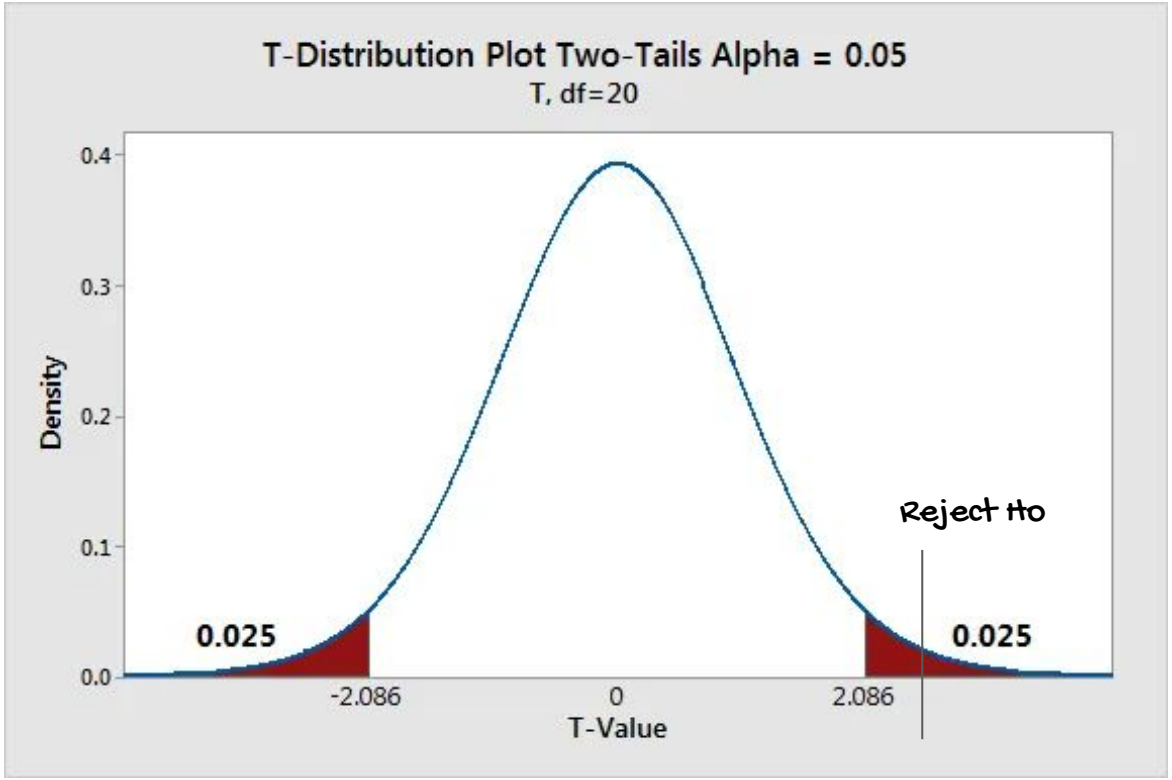


Two Types of Error

Things can go wrong

	Reject Ho	Do not reject Ho
Ho is TRUE	**Type 1** <False Positive>	True Negative
Ho is FALSE	True Positive	**Type 2** <False Negative>

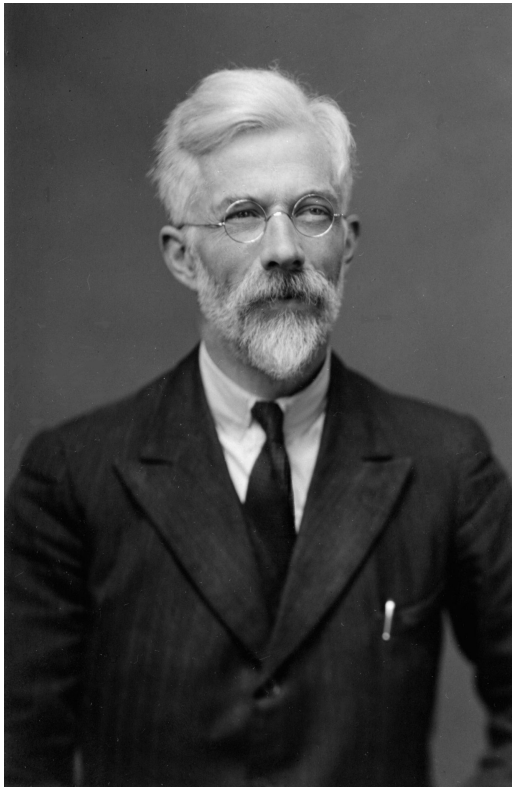




When we reject Ho, we control **the alpha** <= 5%.

1-alpha = confidence level

Hypothesis Testing



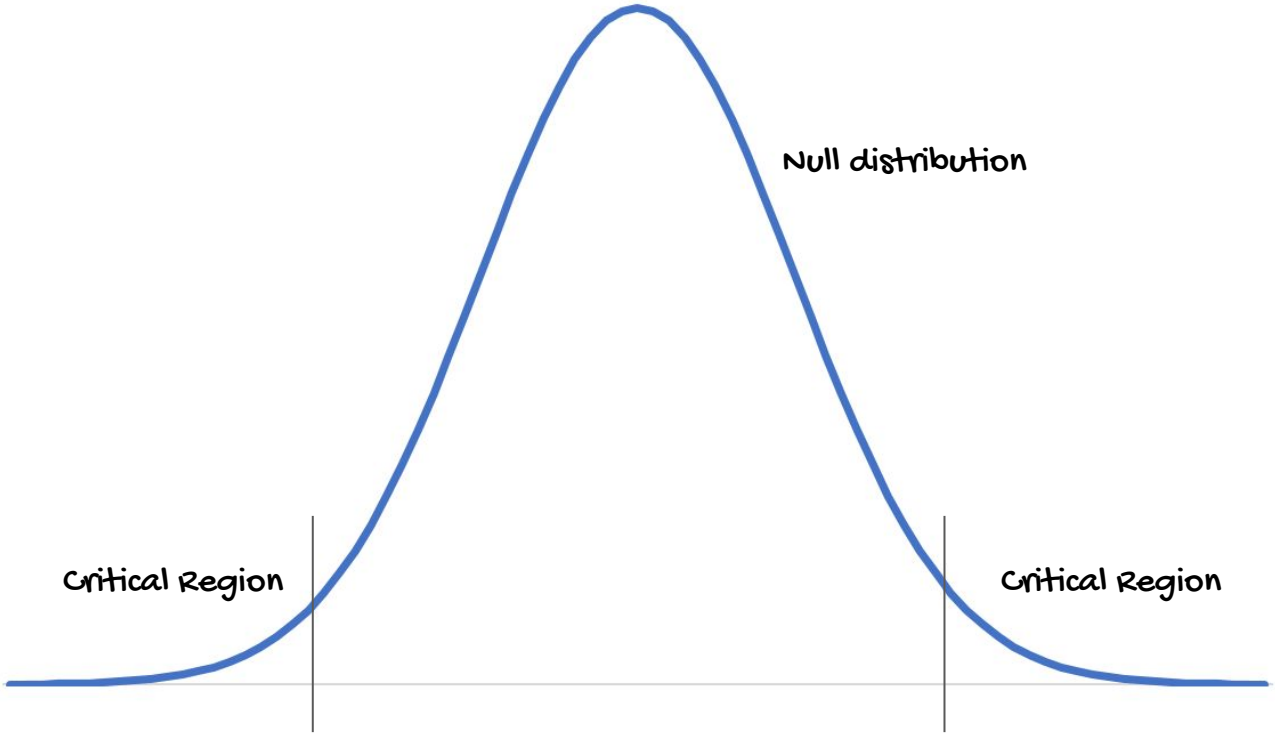
Frequentist School

Ronald Fisher
1890 - 1962

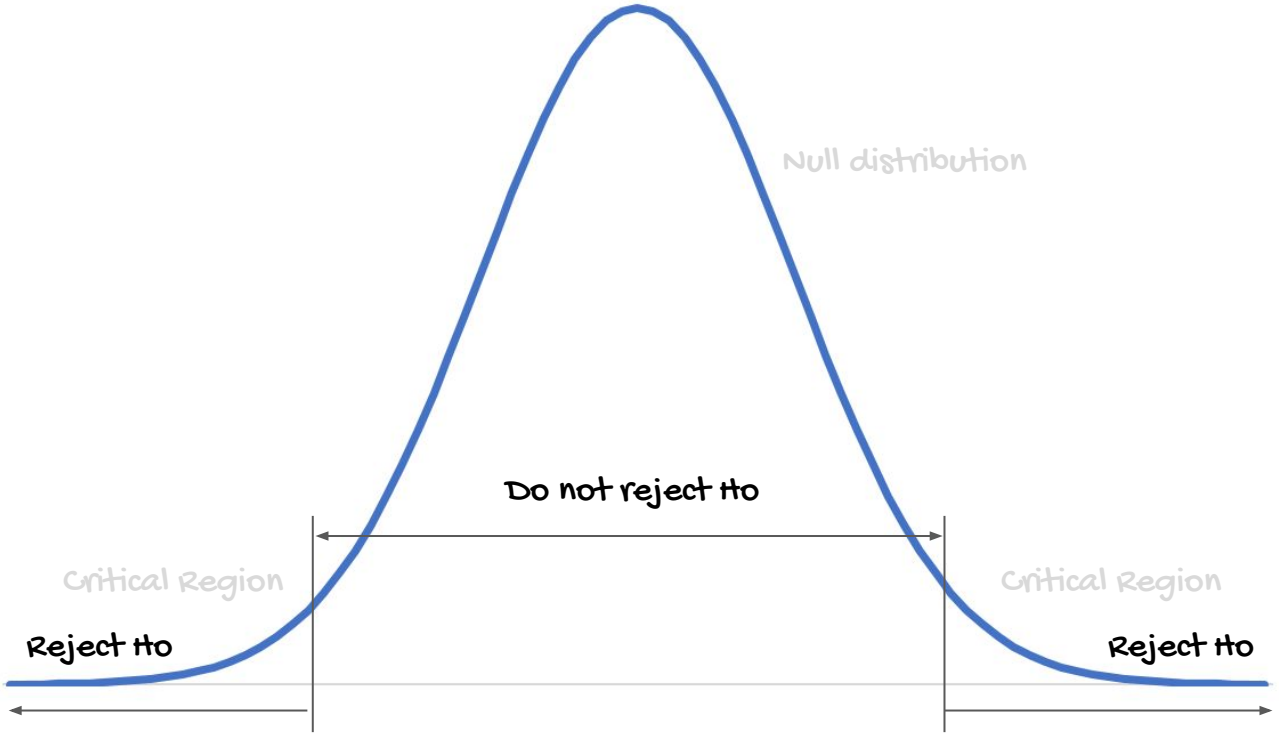
3 Ways to Test

- Critical Region
- p-value
- Confidence Interval

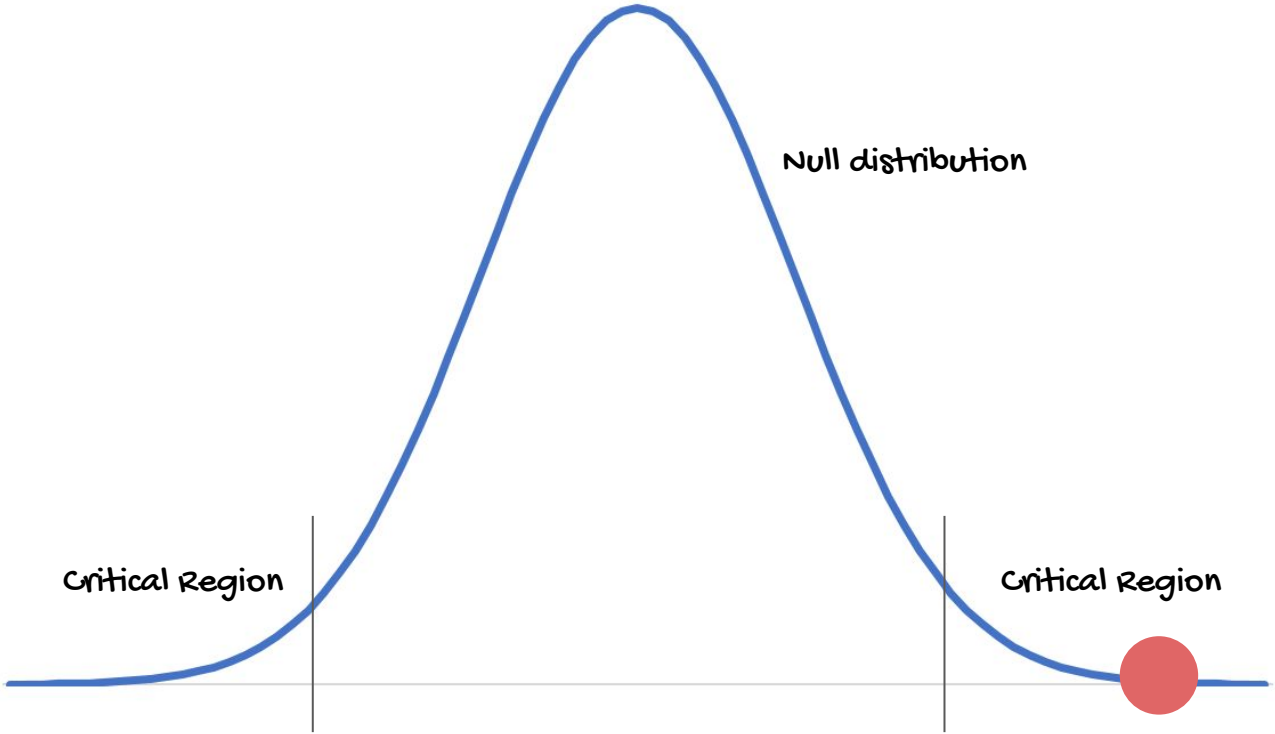
1. Critical Region 🔥



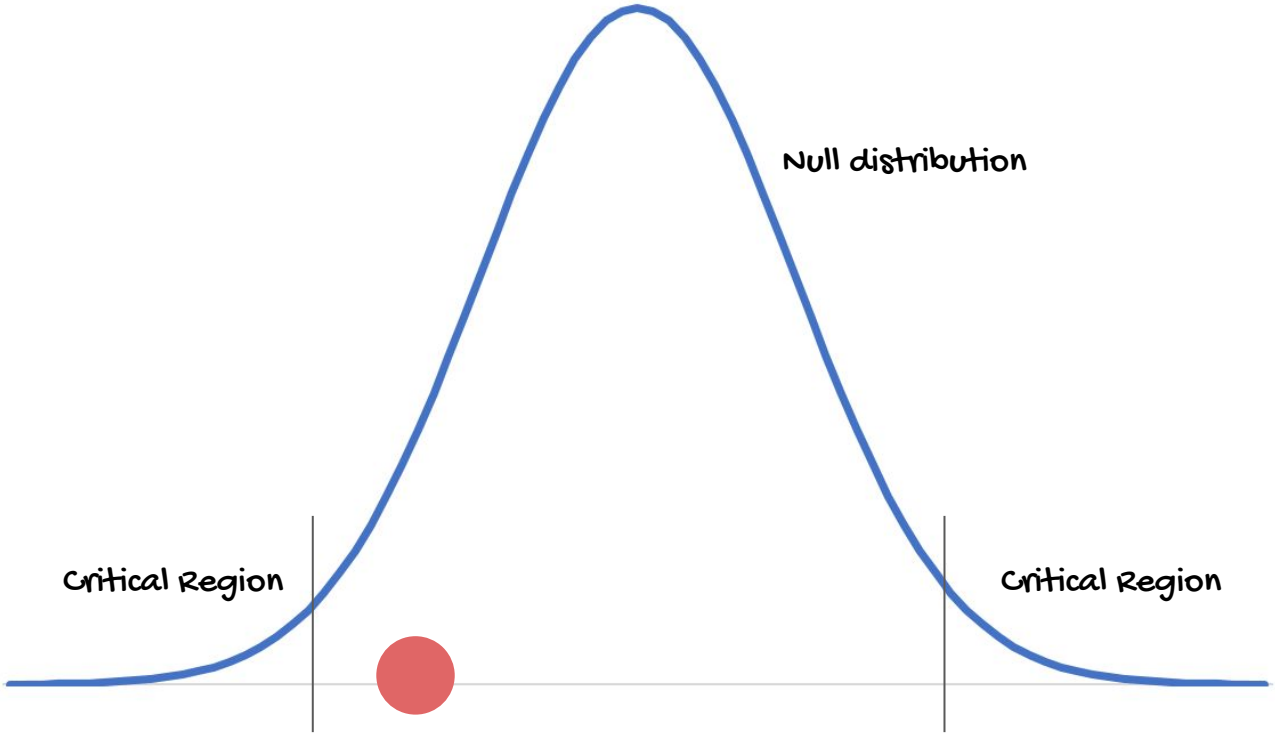
1. Critical Region 🔥



1. Critical Region 🔥



1. Critical Region 🔥



2. p-value 🔥

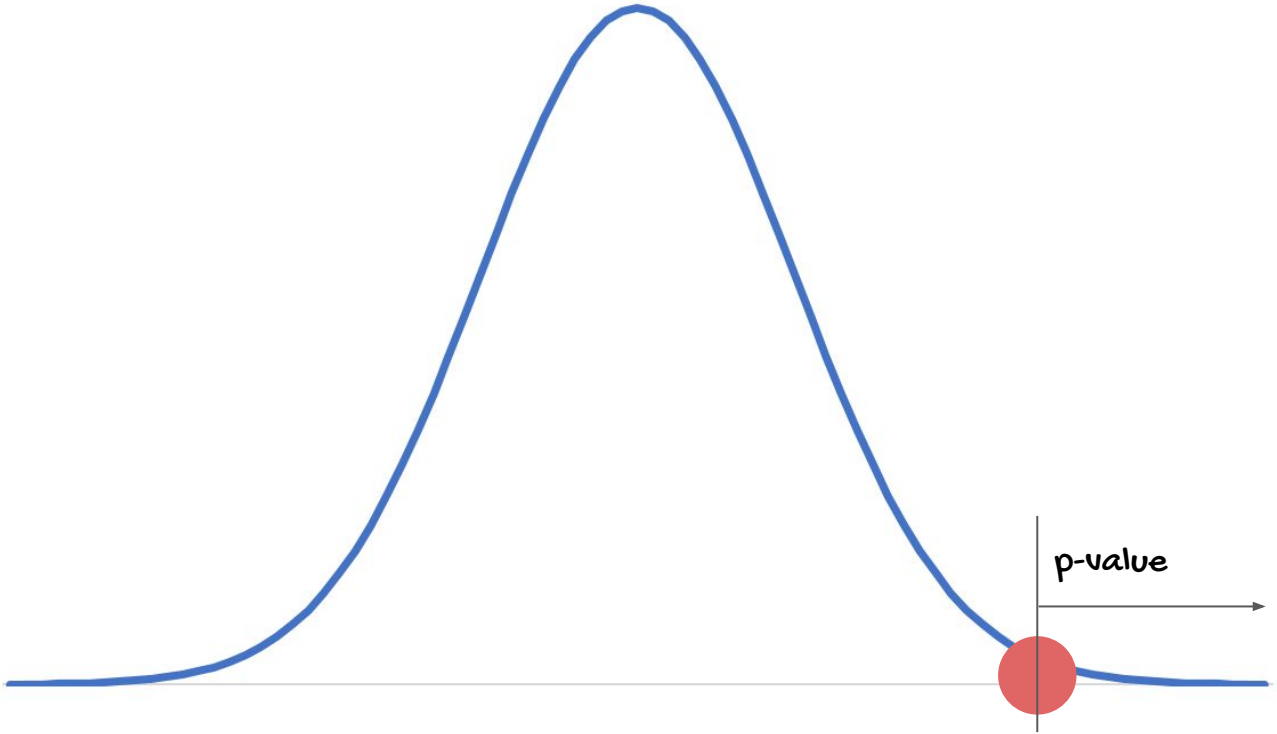
Assuming Ho is true

p(observed data or more extreme | Ho is TRUE)

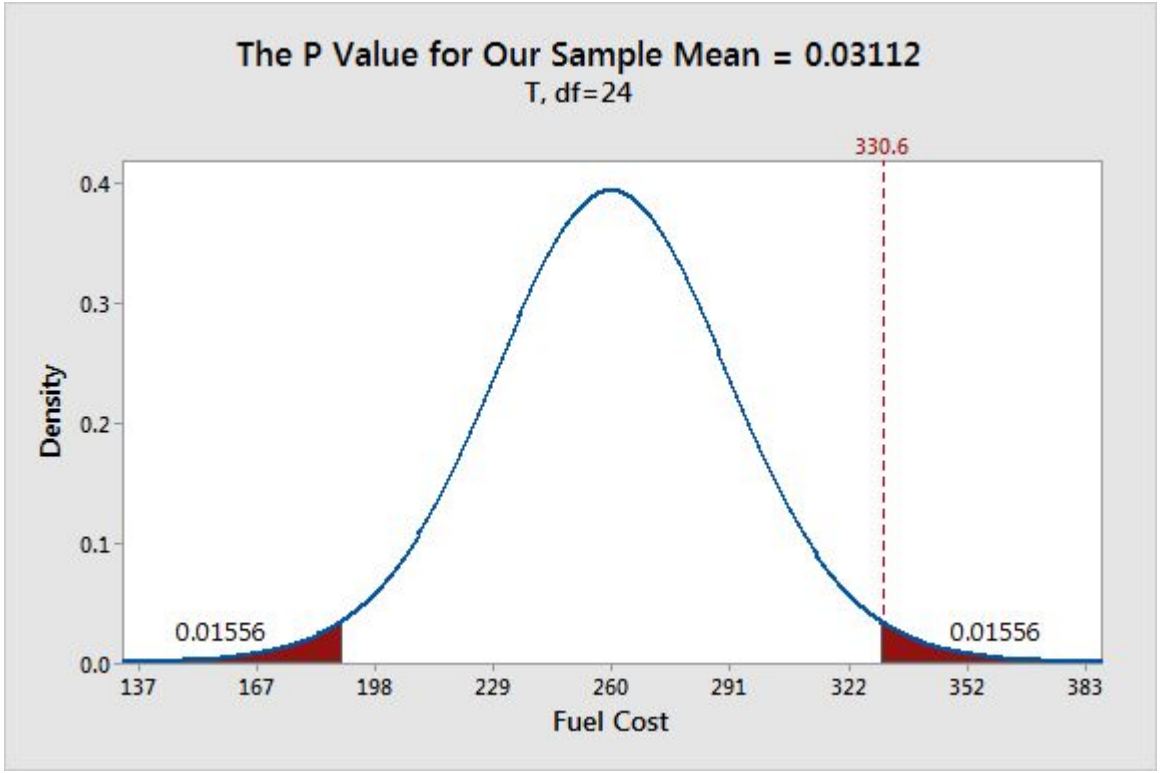
probability

Given

2. p-value 🔥



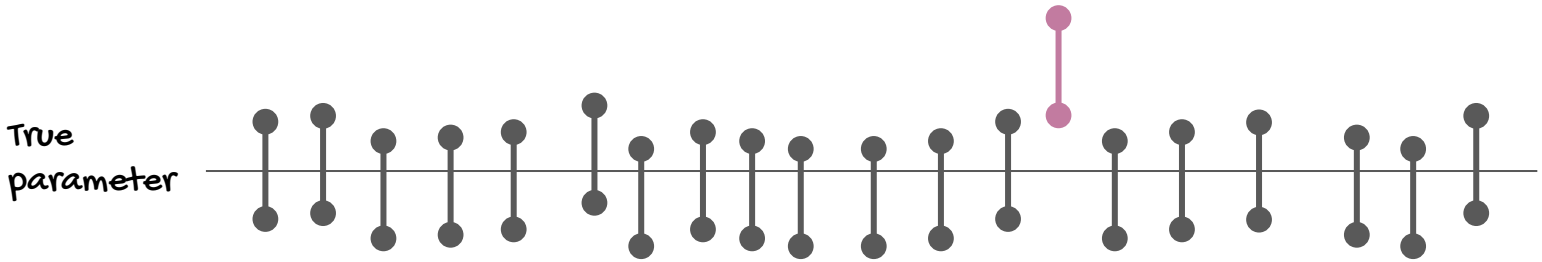
p-value = 0.01556 * 2 = 0.03112



How Hypothesis Tests Work: Significance Levels (Alpha) and P values - Statistics By Jim

3. Confidence Interval 🔥

19/20 = 0.95 = 95% confidence



3. Confidence Interval

3. Confidence Interval 🔥

```
Ho: avg. lays big pack = 260g
Ha: avg. lays big pack != 260g
```

3. Confidence Interval 🔥

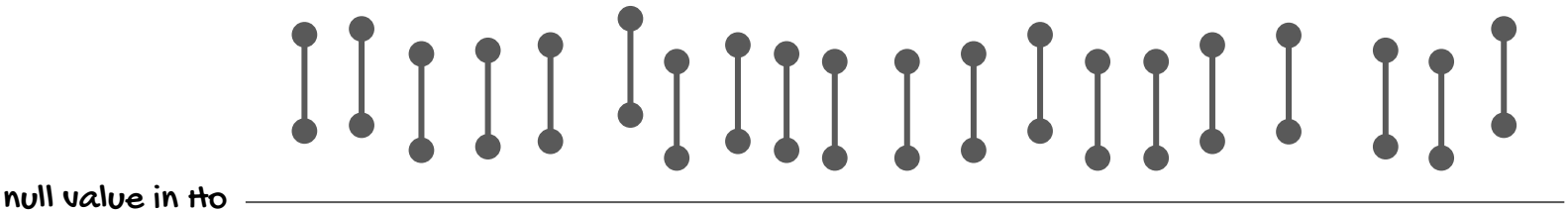
```
Ho: mean diff = 0
Ha: mean diff != 0
```

~~Ho~~ = 0

null value in ~~Ho~~

3. Confidence Interval 🔥

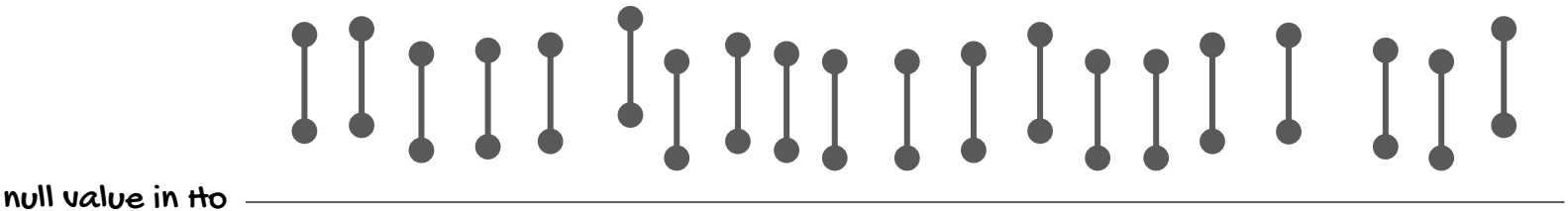
Ho: mean diff = 0
Ha: mean diff != 0



95% CI **[1.2, 3.5]** => ???

3. Confidence Interval 🔥

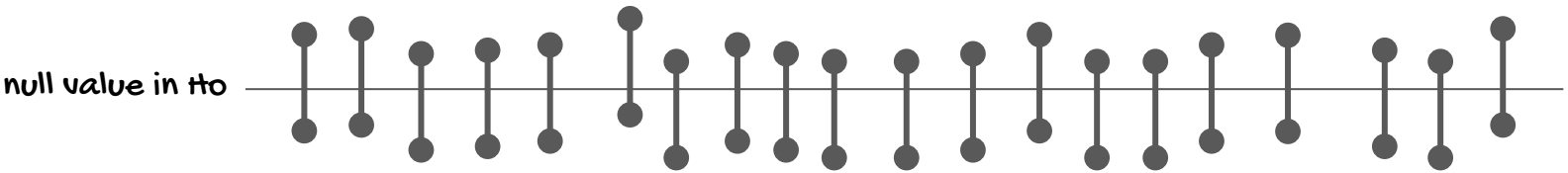
Ho: mean diff = 0
Ha: mean diff != 0



95% CI **[1.2, 3.5]** => Reject Ho

3. Confidence Interval 🔥

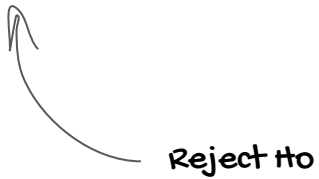
```
Ho: mean diff = 0
Ha: mean diff != 0
```



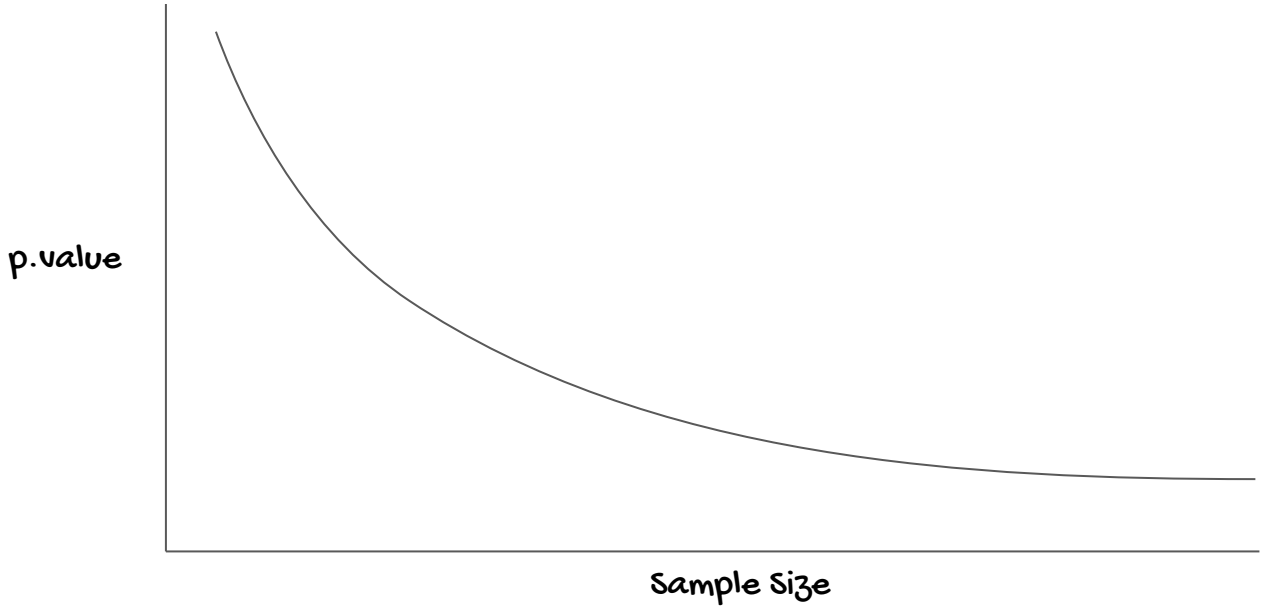
95% CI **[-0.9, 1.5]** => Do Not Reject Ho

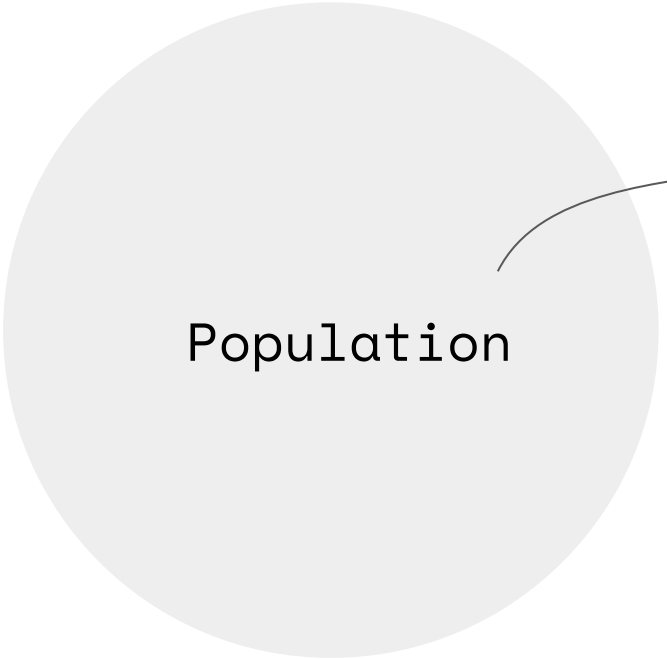
The same conclusion

- If our data falls in critical region
  - p-value <= 0.05
  - 95% CI *will not* contain null value



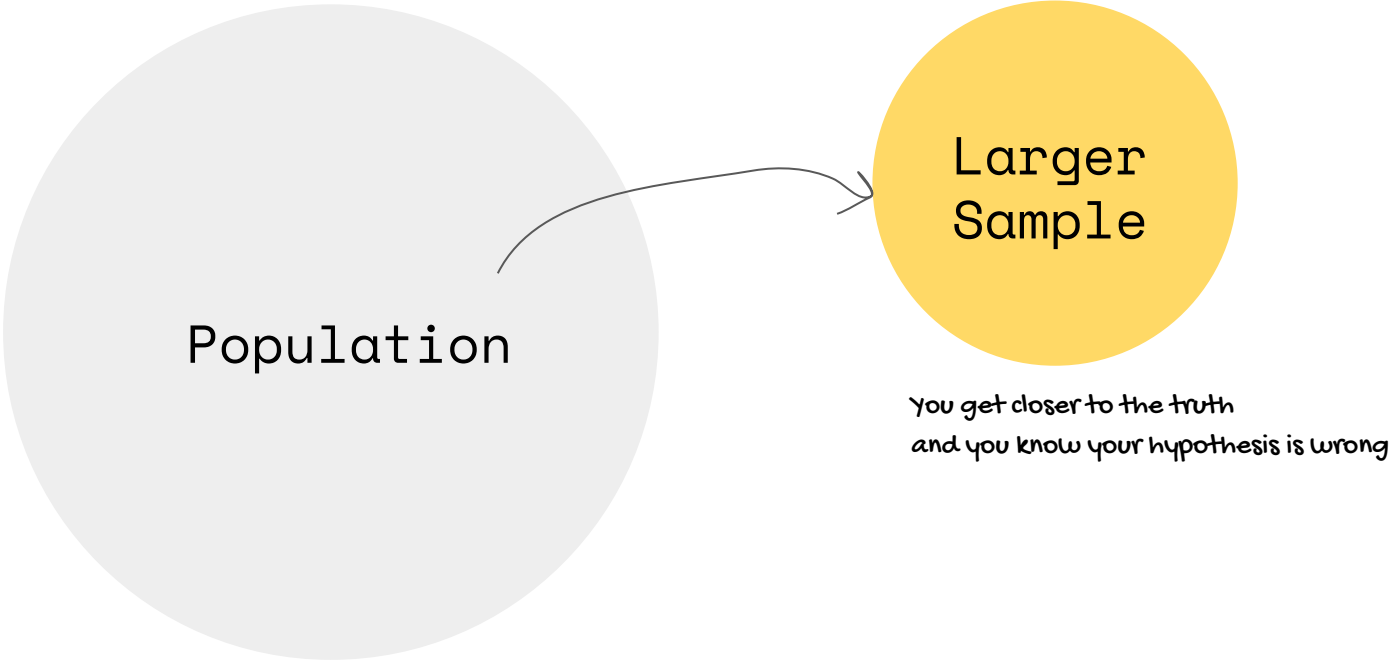
p-value is sensitive to **sample size**

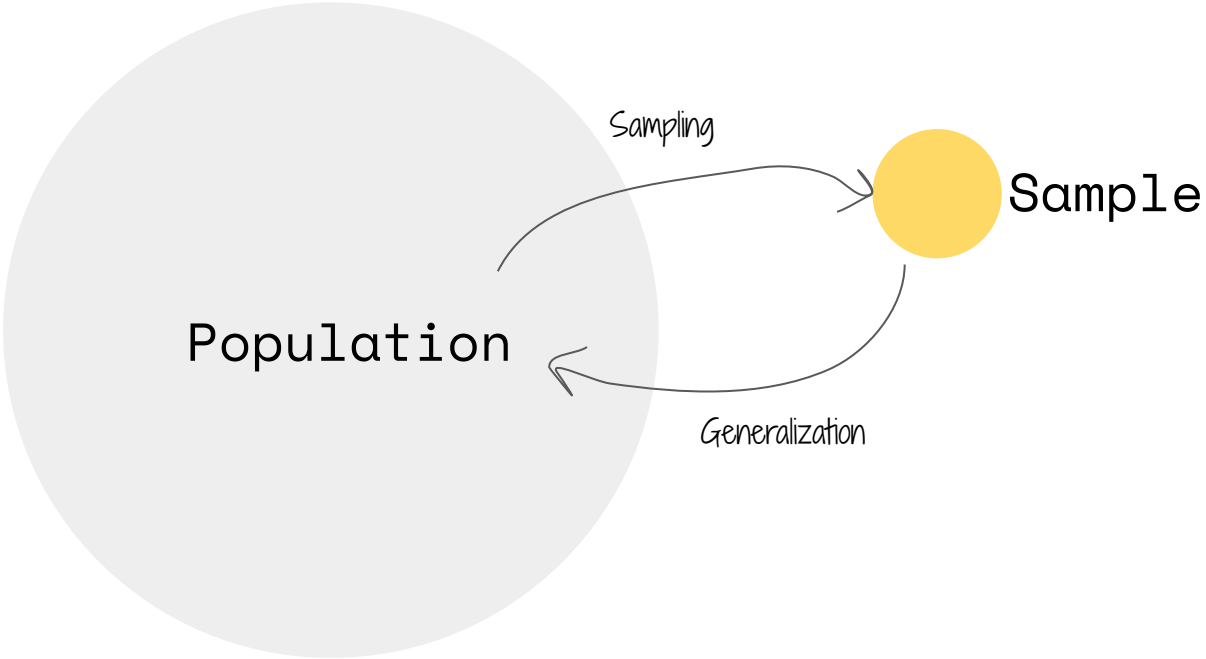




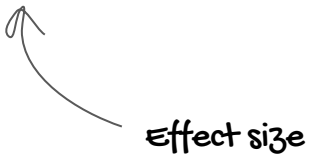
Sample

Ho: Biden popularity = 70%
Ha: Biden popularity != 70%





Hypothesis test tests whether **the difference exists**, but does not test **the impact** of the difference.



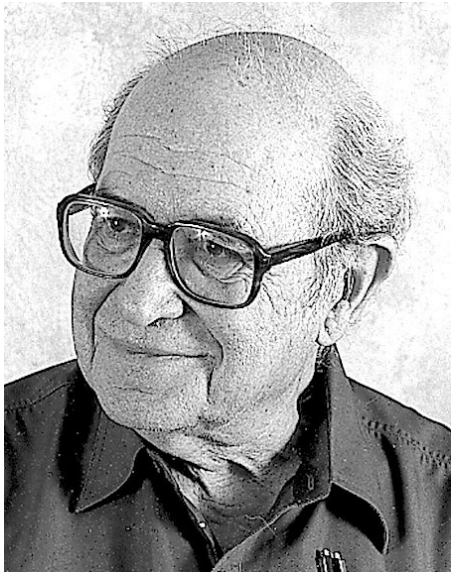
The Earth Is Round ($p < .05$)

Jacob Cohen

After 4 decades of severe criticism, the ritual of null hypothesis significance testing—mechanical dichotomous decisions around a sacred .05 criterion—still persists. This article reviews the problems with this practice, including its near-universal misinterpretation of p *as the probability that* H_0 *is false, the misinterpretation that its complement is the probability of successful replication, and the mistaken assumption that if one rejects* H_0 *one thereby affirms the theory that led to the test. Exploratory data analysis and the use of graphic methods, a steady improvement in and a movement toward standardization in measurement, an emphasis on estimating effect sizes using confidence intervals, and the informed use of available statistical methods is suggested. For generalization, psychologists must finally rely, as has been done in all the older sciences, on replication.*

sure how to test H_0, chi-square with Yates's (1951) correction or the Fisher exact test, and wonders whether he has enough power. Would you believe it? And would you believe that if he tried to publish this result without a significance test, one or more reviewers might complain? It could happen.

Almost a quarter of a century ago, a couple of sociologists, D. E. Morrison and R. E. Henkel (1970), edited a book entitled *The Significance Test Controversy*. Among the contributors were Bill Rozeboom (1960), Paul Meehl (1967), David Bakan (1966), and David Lykken (1968). Without exception, they damned NHST. For example, Meehl described NHST as "a potent but sterile intellectual rake who leaves in his merry path a long train of ravished maidens but no viable scientific offspring" (p. 265). They were, however, by no means the first to do so. Joseph Berkson attacked NHST in 1938, even before it



Jacob Cohen

Ho: earth is round (equal r = 100,000km)
Ha: earth is not round

H0 : โลกเป็นรูปไข่
Ha : โลกไม่ใช่รูปไข่

Key Takeaways

- All methods produce the same result
  - Critical region
  - p-value
  - Confidence interval
- Confidence interval is highly recommended in modern research
- p-value is sensitive to sample size

Data Science Bootcamp

<> Hypothesis Testing in Statistics