Applied Statistics for Data Analytics

Module 4: Hypothesis testing

DeepLearning.AI

Hypothesis testing

Module 4 introduction

DeepLearning.AI

Module 4 outline

- **Hypothesis testing for means**
  - Formulate hypotheses
  - Calculate test statistic
  - Determine significance level
  - Interpret p-values
- **Other hypothesis tests**
  - Tests for proportions
  - Two-sample tests
  - Other types of tests
- **LLMs for hypothesis testing**
  - Formulate hypotheses
  - Interpret results
  - Run tests for you

Hypothesis testing

Demo: hypothesis testing in action

DeepLearning.AI

Hypothesis testing

Did you observe an outcome because of:

- Random chance — “You got lucky”
- True pattern

Information

Size of difference:	2 seconds
Cubes solved:	20

Hypothesis testing

Who’s the better cube solver?

Hypothesis testing

Hypothesis testing: means

DeepLearning.AI

Scenario

30 day free trial

Problem: Investigate whether users with a free trial stay subscribed for longer

Users without free trial subscribe for **10 months**

Sample: 100 users who received a free trial

$\bar{x} = 10.4$ months

$s = 2$ months

Close!

Is **0.4 months** large enough to convince you that free trial is effective?

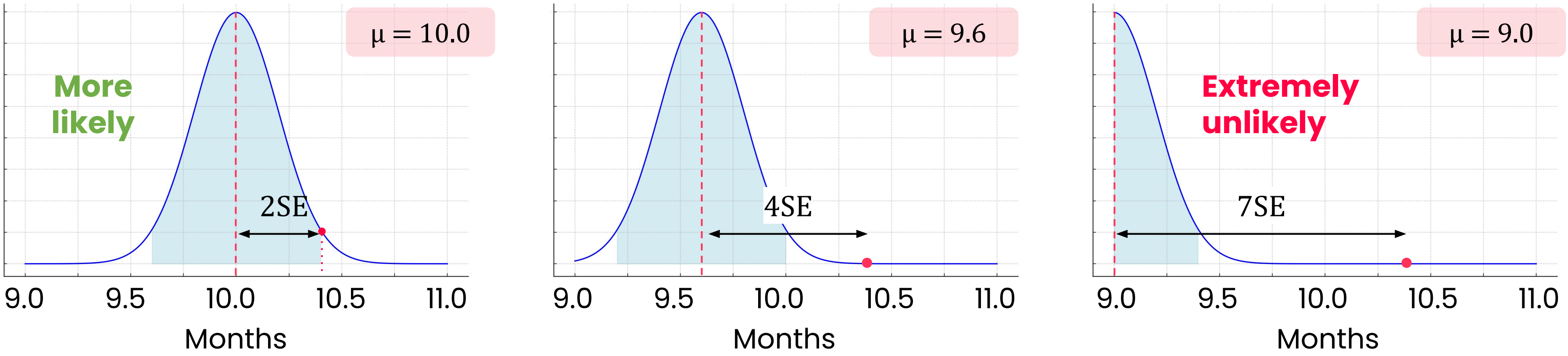
Scenario

$\bar{x} = 10.4$ months

$s = 2$ months

- **Unknown**: Where mean of 10.4 falls in sampling distribution
- **Known**: Sampling distribution for means is normally distributed

$$\mathrm{SE} = \frac{s}{\sqrt{n}} = \frac{2}{\sqrt{100}} = 0.2$$



Scenario

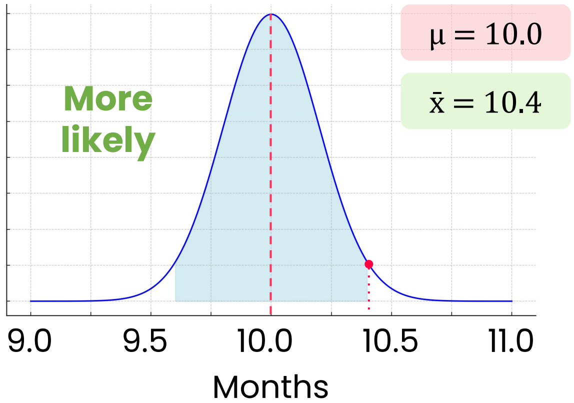
🚫 **You can never:** know true population mean to compare with

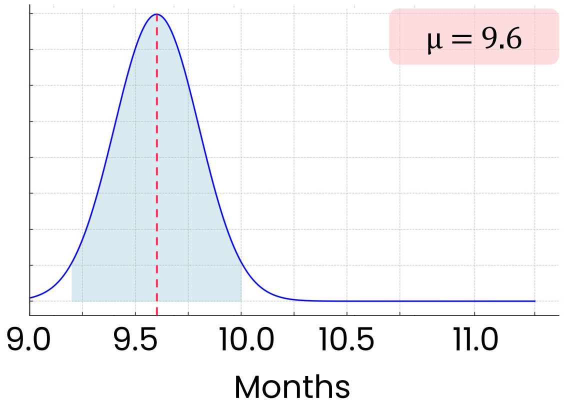
✅ **You can:** calculate likelihood of observing $\bar{x}$ if μ is the value you suspect

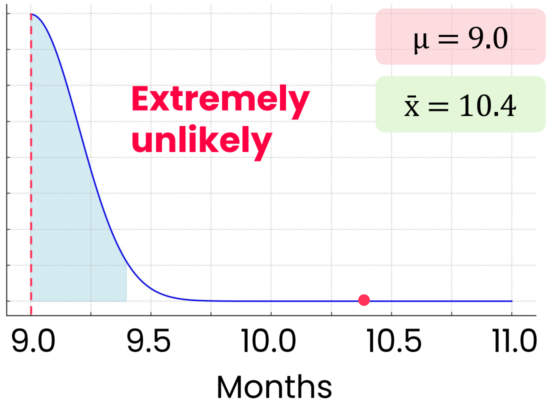
Tempting:

$10.4 > 10$

Close!







👎 Cost money, giving no benefit

👎 Slightly worse for retention

Hypothesis testing

- Evaluate whether $\bar{x}$ is significantly different from known μ, given **variability** and **sample size**

Sample mean
for free trial users

$\bar{x} = 10.4$ months

Population mean
for existing users

$\mu = 10.0$ months

- Distinguishes between two possible scenarios:
    - ⬜ Observed difference is **due to random chance** →
    - ⬜ Observed difference **reflects a genuine difference** →

Statistical significance

Not statistically significant
Doesn’t help you draw meaningful conclusion

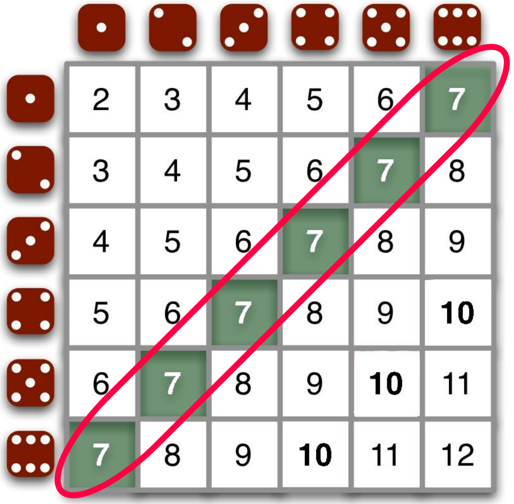
Statistically significant
Difference is likely real,
can provide evidence for hypothesis

Hypothesis testing

Very common to:

- 📝 Collect a sample
- 📊 Calculate x̄ different from μ
- ⛔ Two values are different: **not enough** to conclude that difference is meaningful

Example: Rolling dice



Population mean:

$\mu = 7$

Mean of n=10 rolls:

$\bar{x} = \sim 7$

Erickson, T. (2011, January 8). An Empirical Approach to Dice Probability. Best Case [Blog post]. Retrieved from https://bestcase.wordpress.com/2011/01/08/an-empirical-approach-to-dice-probability/

Hypothesis testing

Only work effectively under certain conditions:

- ✅ Data is a representative sample, ideally a random sample
- ✅ Observations in your data must be independent
- ✅ Must meet one of two conditions:
  - ⬜ Normally distributed
  - ⬜ Sample size must be large

Sample isn't random: no way of knowing what biases sampling method introduced

"Large": Typically means 30, but 50 or more is ideal

Central Limit Theorem: sampling distribution of the mean approaches normal distribution as n increases

You will learn

How to perform a hypothesis test

- Defining your hypotheses
- Calculating the test statistic
- Defining the significance level
- Calculating the p value
- Interpreting the results

How to define errors

How to work with small sample sizes

Hypothesis testing

The hypothesis

DeepLearning.AI

Scenario

30 day free trial

- **Problem**: Mean retention time (μ) of people with free trial
- **Outcomes**:

Hypotheses

Null hypothesis	Alternative hypothesis
$\mu = 10$ months	$\mu > 10$ months
H_0	H_1

H_0

- Aren't able to find evidence that $\mu > 10$
- Finding no effect or no difference

H_1

- Evidence that $\mu > 10$
- Free trial was effective

Defining your hypotheses

1. Start with the **null hypothesis**
    - Identify value you expect for no effect
2. For the **alternative hypothesis**
    - Compare population parameter with value in the null hypothesis
    - Define the comparison you're interested in

No effect of free trial → $H_0 : \mu = 10 \text{ months}$

$H_1 : \mu > 10 \text{ months}$ — Is the mean significantly **greater than** 10?

$H_1 : \mu < 10 \text{ months}$ — Is the mean significantly **less than** 10?

$H_1 : \mu \neq 10 \text{ months}$ — Is the mean significantly **different than** 10?

Explaining your hypotheses

Test outcome	Statement	Conclusion
Evidence for H_1	"We **reject** the null hypothesis"	Null hypothesis is likely not true
No evidence for H_1	"We **fail to reject** the null hypothesis"	Not enough evidence to reject null hypothesis

You should **avoid** phrases like:

- 🚫 "Prove the alternative hypothesis"
- 🚫 "Accept the null hypothesis"

This terminology:

- ✅ Helps avoid overstating conclusions
- ✅ Reminds stakeholders that tests can never prove with certainty

Selecting your hypotheses

✅ Hypotheses should be based on:

- Theory
- Observable evidence

⛔ Don't just select them at random

Example:

$H_0 : \mu = 10$ months

✅ Known average for existing users

✅ Plausible that the behavior of users who got a free trial is similar

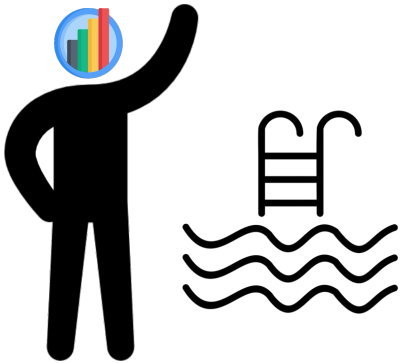
Hypothesis testing

Identifying the hypothesis and test type

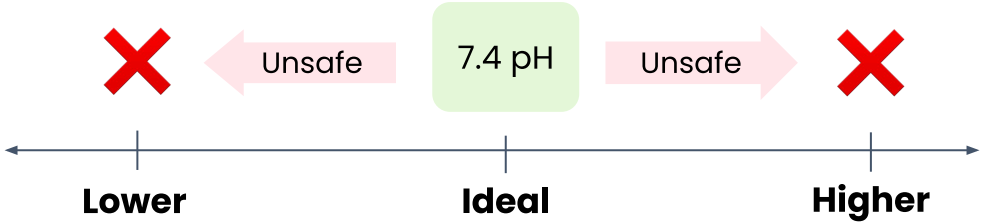
DeepLearning.AI

Scenario

Problem: Testing the pH of a swimming pool, for which the ideal pH is 7.4.



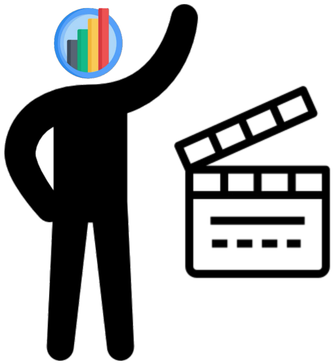
You
Data Analyst



- **Null hypothesis:** $H_0 : \mu = 7.4 \text{ pH}$
- **Alternative hypothesis:** $H_1 : \mu \neq 7.4 \text{ pH}$

Scenario

Problem: Are movie lengths in 2013 greater than 120 mins?



You
Data Analyst

- **Null hypothesis:**

$$H_0 : \mu = 120 \text{ mins}$$

No difference between movies in 2013 & expected value

- **Alternative hypothesis:**
  - ⬜ Find evidence and reject the null hypothesis
  - ⬜ Don't find evidence and fail to reject the null hypothesis

$$H_1 : \mu > 120 \text{ mins}$$

Movies are longer than two hours

Scenario

Problem: Whether average delivery time is less than 45 mins

You
Data Analyst

- **Null hypothesis:**

$H_0 : \mu = 45 \text{ mins}$

Delivery times not different from 45 mins

- **Alternative hypothesis:**
  - ⬜ Reject the null hypothesis if you do find that evidence
  - ⬜ Fail to reject it otherwise

$H_1 : \mu < 45 \text{ mins}$

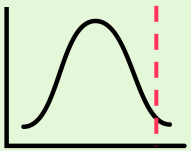
Delivery times are less than 45 mins

Test types

One-tailed tests

$\mu > \_\_$

Reject the H_0 if:

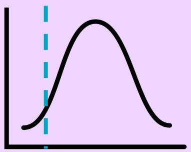


$\bar{x}$ is rare on the upper end

Perform a **right-tailed test**

$\mu < \_\_$

Reject the H_0 if:

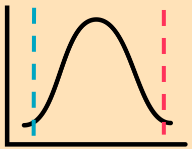


$\bar{x}$ is rare on the lower end

Perform a **left-tailed test**

$\mu \neq \_\_$

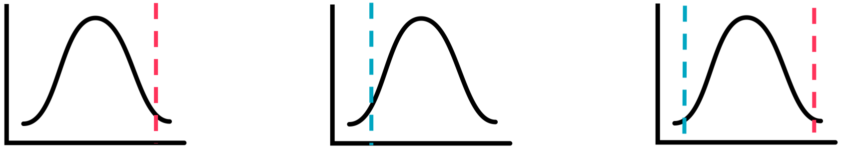
Reject the H_0 if:



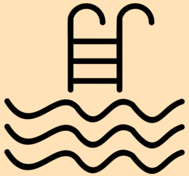
$\bar{x}$ is rare on the either end

Perform a **two-tailed test**

Test types

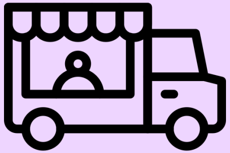


Right-tailed | Left-tailed | Two-tailed



$H_1 : \mu \neq 7.4 \text{ pH}$

$H_1 : \mu > 120 \text{ mins}$



$H_1 : \mu < 45 \text{ mins}$

Perform a **two-tailed test**

Perform a **right-tailed test**

Perform a **left-tailed test**

Hypothesis testing

Calculating the test statistic

DeepLearning.AI

Scenario

- $n = 50$ movies
- $\bar{x} = 123$ minutes
- $s = 12$ minutes
- $SE = \frac{s}{\sqrt{n}} = \frac{12}{\sqrt{50}} \approx 1.7$

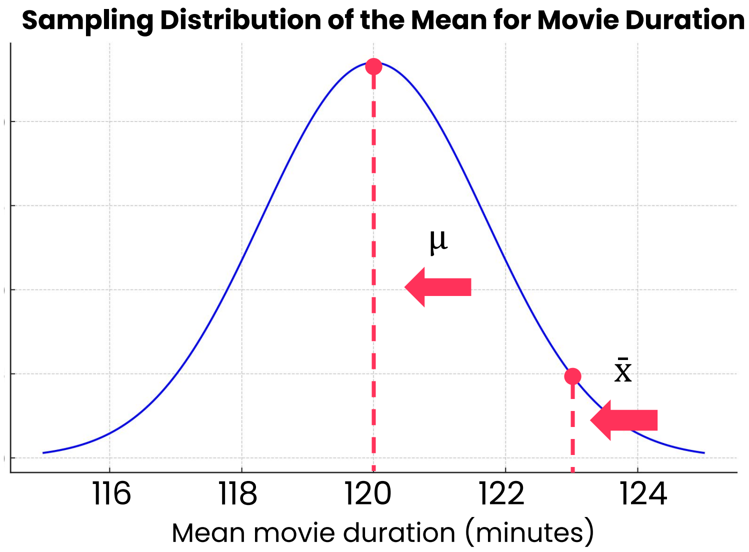
Test statistic

$z = ?$

You're performing a **right-tailed** test:

$H_0 : \mu = 120$ mins

$H_1 : \mu > 120$ mins



- ⛔ Highly variable → can't be as confident that means are different
- ⛔ n is lower → not certain your results really reflect the population

Scenario

- $n = 50$ movies
- $\bar{x} = 123$ minutes
- $s = 12$ minutes
- $\mathrm{SE} = \frac{s}{\sqrt{n}} = \frac{12}{\sqrt{50}} \approx 1.7$

z -score

Test statistic

$z = 1.76$

You're performing a **right-tailed** test:

$H_0 : \mu = 120$ mins

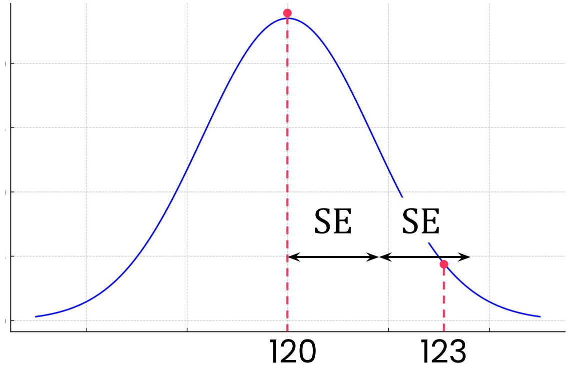
$H_1 : \mu > 120$ mins

1. Calculate difference between $\bar{x}$ and μ

$$\bar{x} - \mu = 123 - 120 = 3$$

2. Divide difference by the standard error

$$= \frac{3}{1.7} = 1.76$$



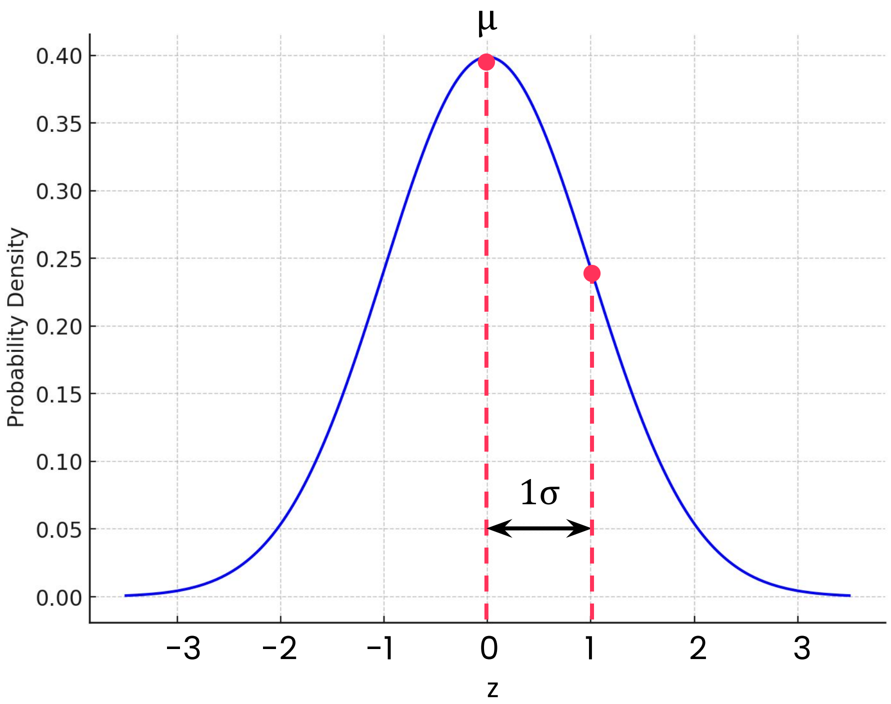
The test statistic (z)

- Number of σ from μ on the standard normal distribution
- Translating x̄ into a value on a standardized scale

$\bar{x} = 123$ → $z = 1.76$

Where:

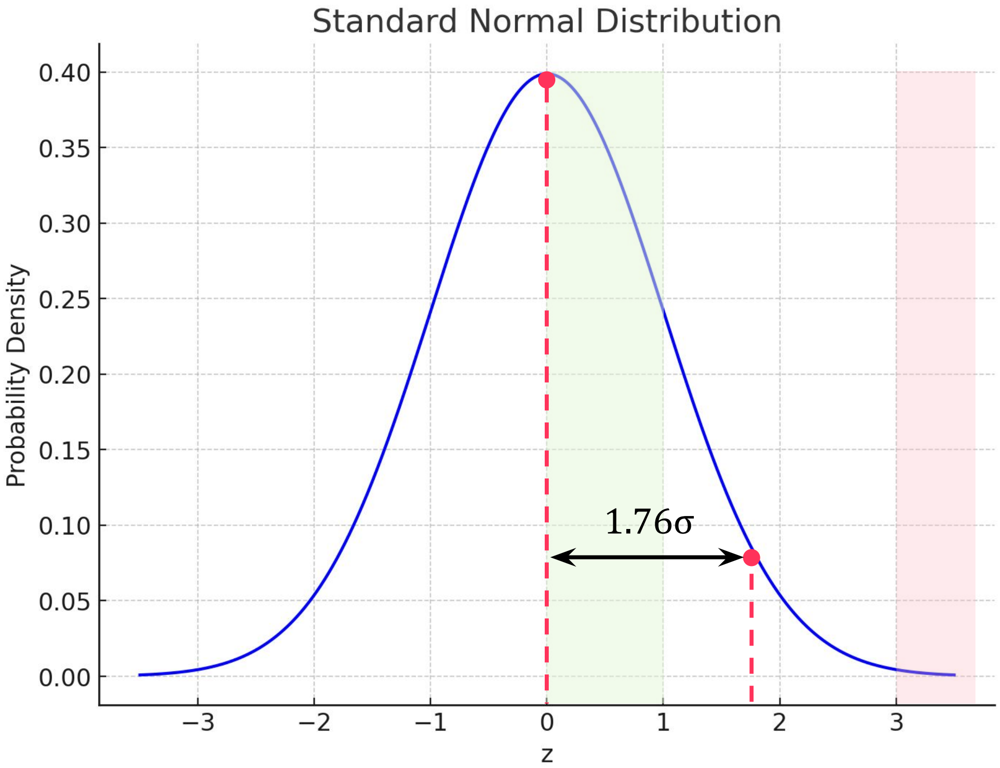
- $\mu = 0$
- Each step is 1σ



Visualizing your test statistic

$z = 1.76$

- Not in the **tail** above $z = 3$
- Not in the **middle** set of common values $0 < z < 1$



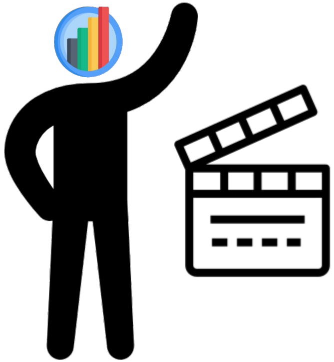
Hypothesis testing

Determining the significance level and rejection region

DeepLearning.AI

Scenario

Problem: Test whether movies in 2013 were longer than 120 minutes



You
Data Analyst

Confidence

The level of certainty you have in your conclusions

90% confidence level

95% confidence level

99% confidence level

- Choice depends on how precise your estimate needs to be
- You will have to make this call
- You are trying to **manage** uncertainty

$\mathbf{n}$ = 50 movies

$\bar{\mathbf{x}}$ = 123 minutes

$\mathbf{s}$ = 12 minutes

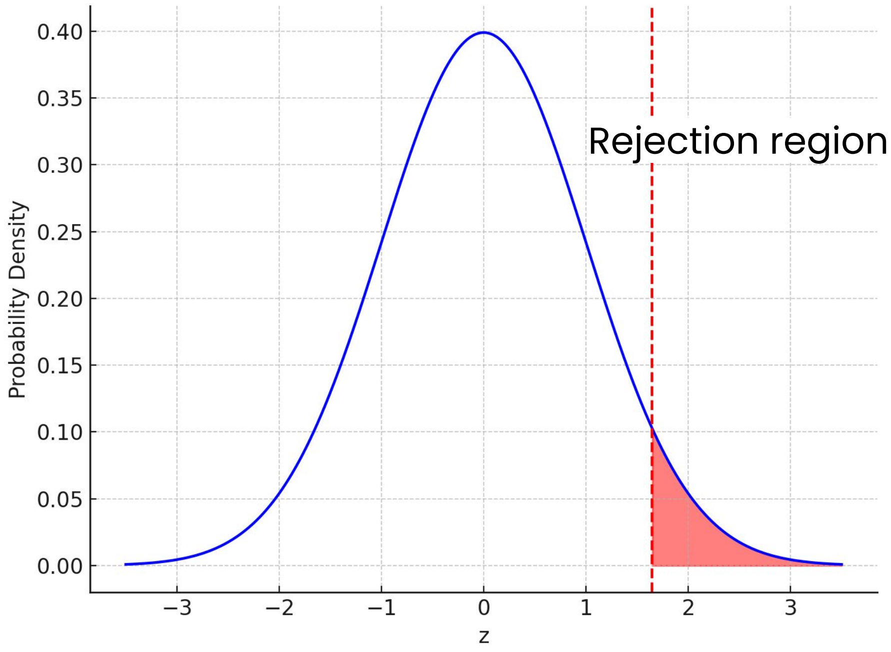
$H_0 : \mu = 120$ mins

$H_1 : \mu > 120$ mins

$z = 1.76$

95% confident that you **correctly** reject H_0:

- Look for results above the mean that occur **5% of the time or less**
- Any z in this region would lead you to reject H_0



$\mathbf{n}$ = 50 movies

$\bar{\mathbf{x}}$ = 123 minutes

$\mathbf{s}$ = 12 minutes

$H_0 : \mu = 120$ mins

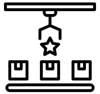
$H_1 : \mu > 120$ mins

$z = 1.76$

For **one-sided** test with a **95%** confidence:

- Ability to reject H_0 depends on size of rejection region
- α = 0.05 is often used in:

Research

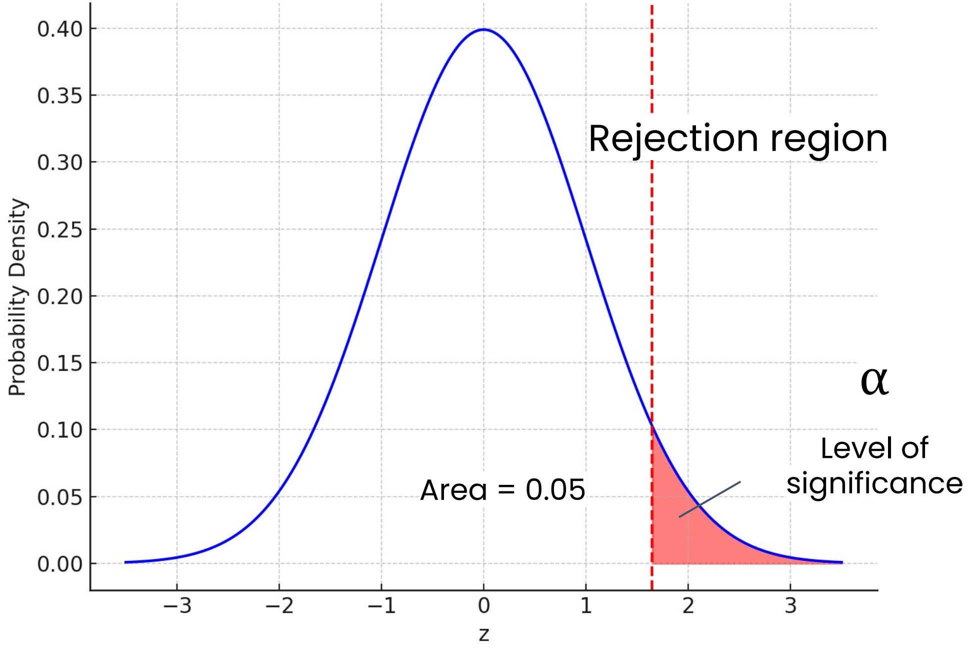


Manufacturing

Sciences

Confidence

$1 - \underbrace{0.05}_{\alpha} = 0.95$



Problem: Be absolutely certain that movies are above 120 minutes before adjusting the schedule

- For a more precise test:

$\alpha = 0.01$

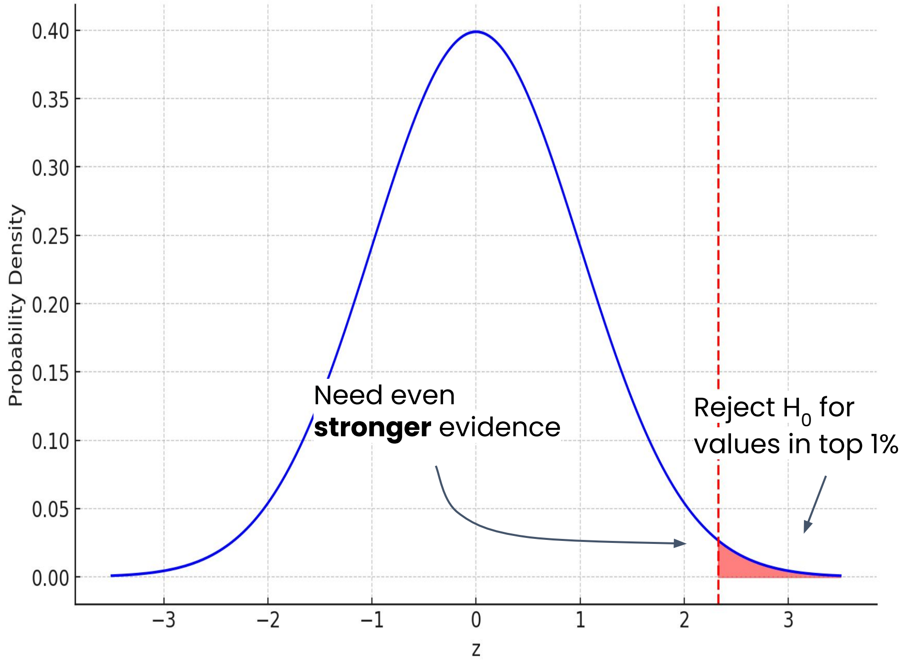
99% confidence level

- Reject H_0 if test statistic is in top 1%

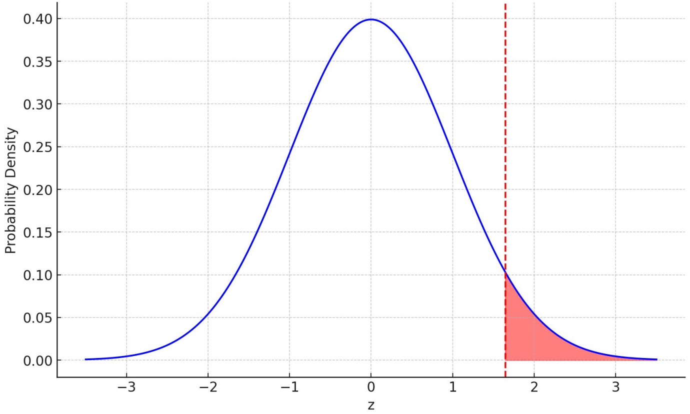
Clinical trials

Impact studies

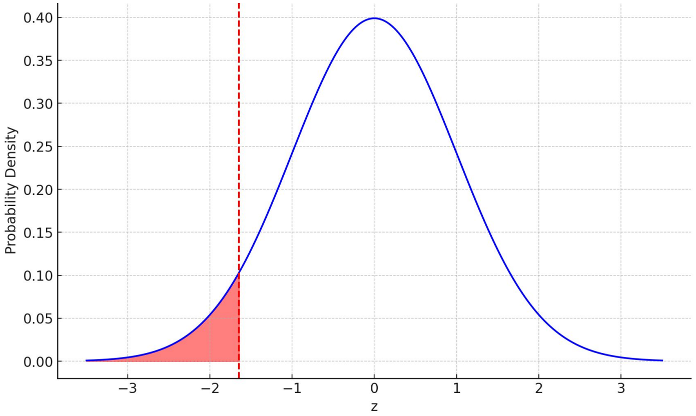
Audits



Right-tailed test



Left-tailed test



Two-tailed test

Values above
and below the mean

?

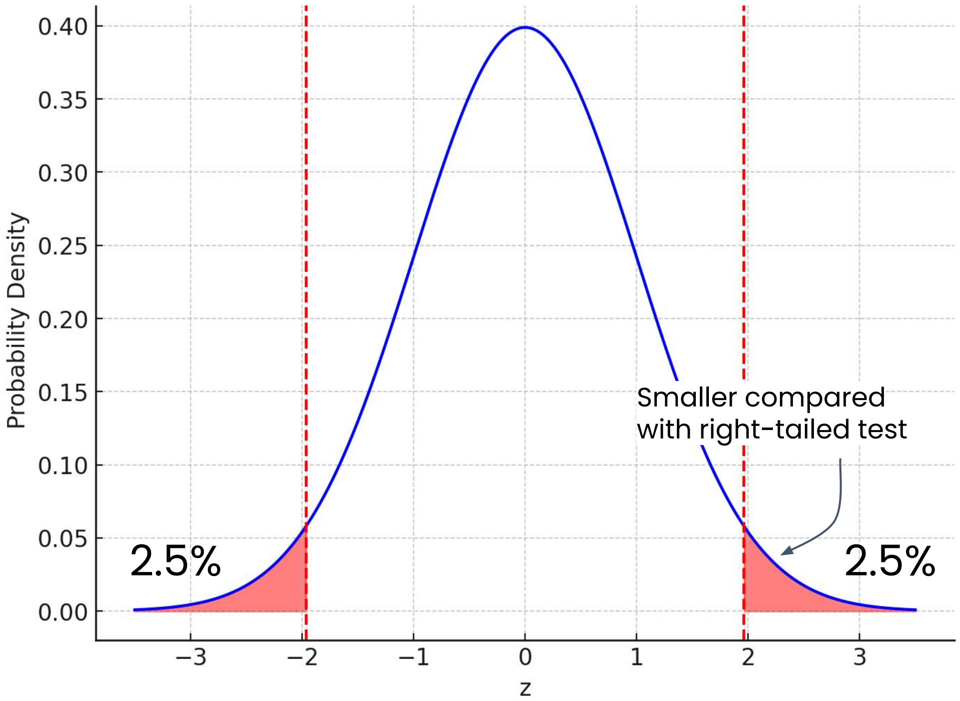
Two-tailed test

$H_0 : \mu = 120$ mins

$H_1 : \mu \neq 120$ mins

- Rejection region contains 2.5% of the data on either side
- Maintains precision, with errors only 5% of the time

❌ If both contained 5% of the values → error rate of 10%



Terminology

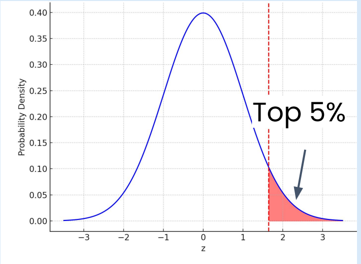
α

significance level

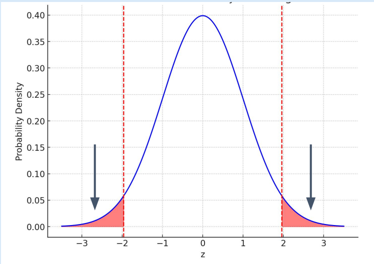
- Defines strength of evidence required to reject H_0
- Larger α lowers bar for strength of evidence
- Smaller α requires stronger evidence against H_0
- Common values for α: **0.10**, **0.05**, and **0.01**

Rejection region

- Area of sampling distribution that contains values that would lead you to reject the H_0



Movie lengths that are surprisingly long



Testing if pool pH is different from 7.4

Hypothesis testing

Calculating the p value

Rarity of the test statistic

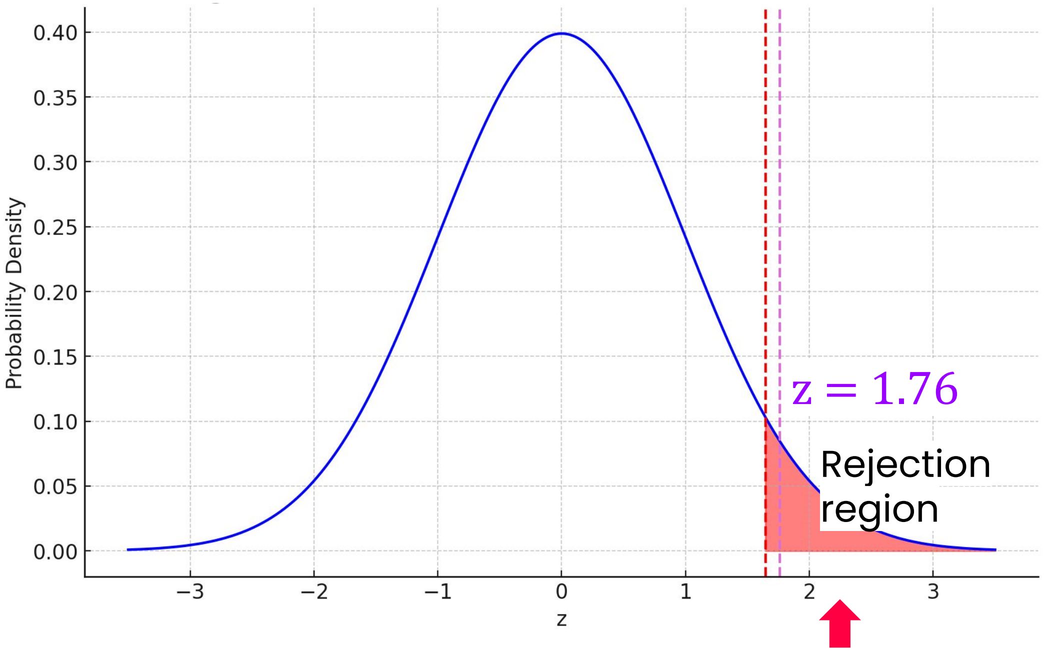
$\alpha = 0.05$

$z = 1.76$

How rare is this value?

$\bar{x}$ = 123 minutes would happen
less than 5% of time

if μ = 120 minutes



Calculating the p-value

$\alpha = 0.05$

$z = 1.76$

Calculate exactly how rare using:

p-value

- Probability of $\bar{x}$ as rare or rarer than z in direction of H_1
- Probability of $z = 1.76$ or higher

You'll need one of two things:

- ⬜ Lookup table
- ⬜ Spreadsheet or programming language

DeepLearning.AI

Sean Barnes

Calculating p-values with the CDF

$\alpha = 0.05$

$z = 1.76$

Calculate: probability of z = 1.76 or higher

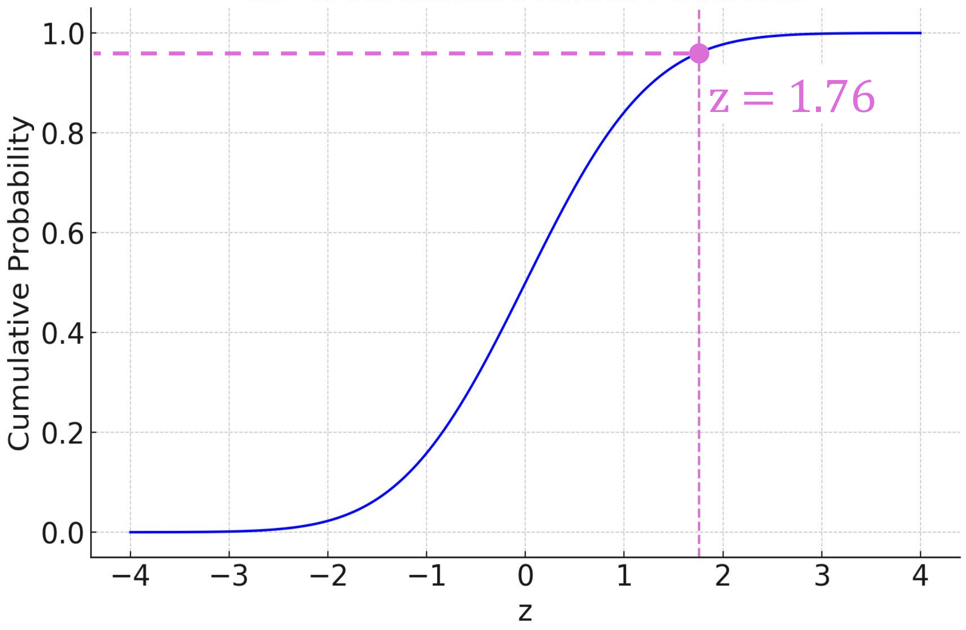
$$P(z > 1.76) = 1 - P(z <= 1.76)$$

$$= 1 - 0.9608$$

$$\textbf{p-value} = 0.0392$$

Sample data → **Z.TEST**

Standard normal CDF $P(Z <= z)$



Sean Barnes

Interpreting the p-value

$\alpha = 0.05$

$z = 1.76$

$\text{p-value} = 0.0392$

If μ = 120,

you would observe $\bar{x}$ = 123 minutes or longer

about **3.92%** of the time.

Compare p-value with α:

$$0.0392 < 0.05$$

Since p-value is below α, you **reject** H_0.

Is this event expected to happen **less than 5% of the time**?

Yes!

Interpreting the p-value

$\alpha = 0.01$

$z = 1.76$

$\text{p-value} = 0.0392$

If μ = 120,

you would observe $\bar{x}$ = 123 minutes or longer

about **3.92%** of the time.

Compare p-value with α:

$$0.0392 > 0.01$$

You **fail to reject** H_0, since p-value is above α.

Determine α before testing!

DeepLearning.AI

Sean Barnes

Scenario 🎬

Are movie durations longer than 120 mins?

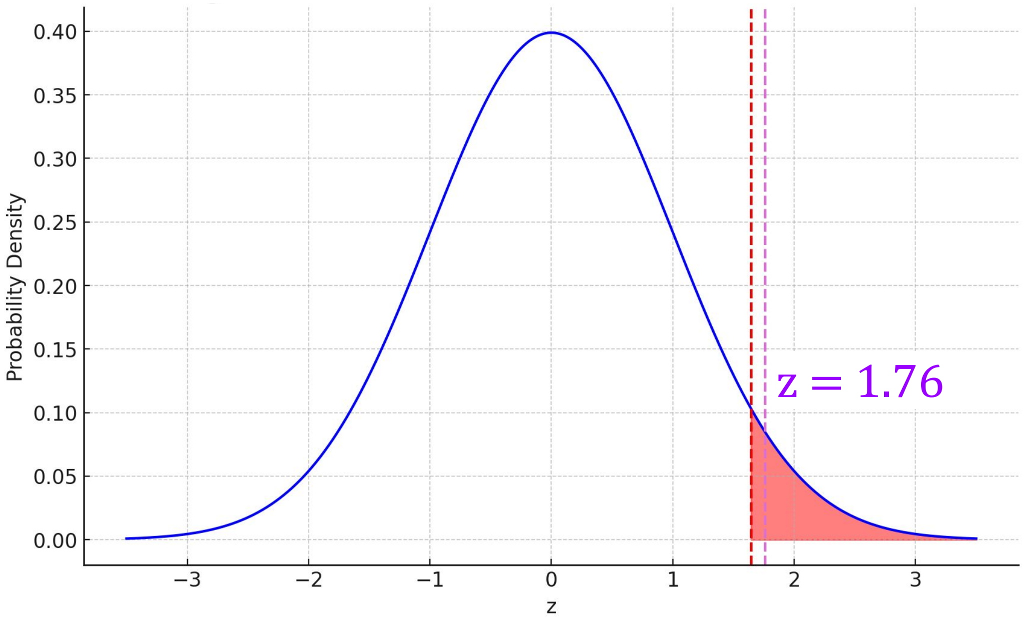
- $\mathbf{n}$ = 50 movies
- $\bar{\mathbf{x}}$ = 123 minutes
- $\mathbf{s}$ = 12 minutes

With 95% confidence in your conclusion:

- Rejected H_0
- Sufficient evidence that μ is > 120 mins

Chance of observing $\bar{x}$ $>= 123$:

$$\text{p-value} = 0.0392$$



Hypothesis testing

Demo: hypothesis testing for means

DeepLearning.AI

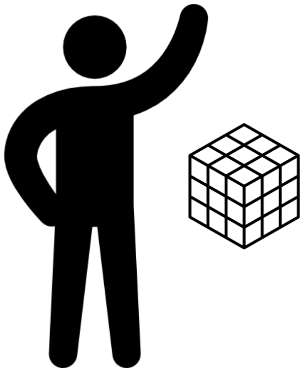
Hypothesis testing

Hypothesis testing errors

DeepLearning.AI

Scenario

Problem: Who's the better solver?



You
Data Analyst

Real life

Only two possible realities:

1. You're equally matched
2. One person is better

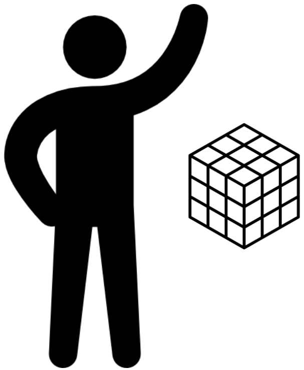
Hypothesis test

Two conclusions:

1. No significant difference between you
2. One person is significantly better

True effects

True negative



You
Data Analyst

No difference to find

Real life

Only two possible realities:

1. You're equally matched ✅
2. One person is better

Didn't find significant difference

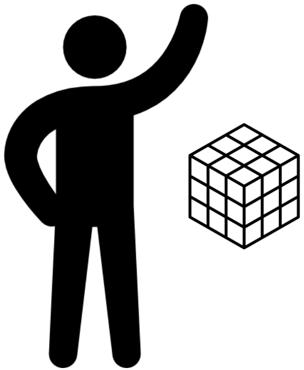
Hypothesis test

Two conclusions:

1. No significant difference between you
2. One person is significantly better

True effects

True positive



You
Data Analyst

Difference in the real world

Real life

Only two possible realities:

1. You're equally matched
2. One person is better

Test correctly identified it

Hypothesis test

Two conclusions:

1. No significant difference between you
2. One person is significantly better

Errors

False positive

You
Data Analyst

No difference in reality

Real life

Only two possible realities:

1. You're equally matched
2. One person is better

Found a difference

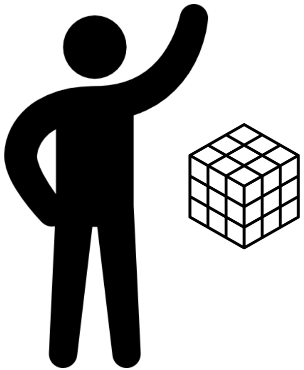
Hypothesis test

Two conclusions:

1. No significant difference between you
2. One person is significantly better

Errors

False negative



You
Data Analyst

Difference in reality

Real life

Only two possible realities:

1. You're equally matched
2. One person is better

Failed to find a difference

Hypothesis test

Two conclusions:

1. No significant difference between you
2. One person is significantly better

True effects and errors

	True state of the world	
Conclusion	H_0 is true	H_0 is false
Reject H_0	⛔ False positive	✅ True positive
Fail to reject H_0	✅ True negative	⛔ False negative

Sean Barnes

Controlling error rates

1️⃣ Significance level α

Chance of incorrectly rejecting the H_0

- ✅ Same as the false positive rate
- ✅ Can minimize false positives by setting α to a very low value

- ⛔ Too low: increase false negatives
- ⛔ Too low: very strict about what counts as significant
- ⛔ Too low: harder to find true effects

2️⃣ Sample size

Increasing or planning to have large sample

- ✅ Provides a more accurate estimate
- ✅ Helps reduce both types of errors
- ✅ Helps detect very subtle effects

Choosing which error to minimize

Sector	False positive	False negative	Preferred error to minimize
Manufacturing	Identifying product as defective when it is fine	Not detecting defective product	**False positives** if customers can return
Medical testing	Over-diagnose patient	Patient has illness, but test fails to detect it	**False negatives** to ensure lifesaving treatment
Loan approval	Approving loan for someone who defaults	Denying loan to someone who would pay back	Balance between both

Hypothesis testing

The t distribution

DeepLearning.AI

$$\mathrm{SE} = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$$

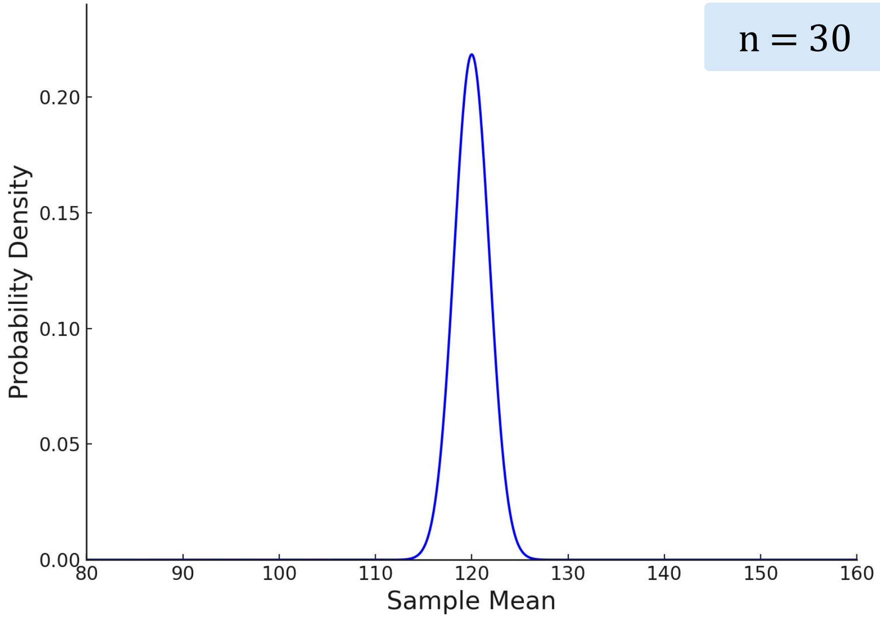
Introduce more uncertainty

$n < 30$ is common

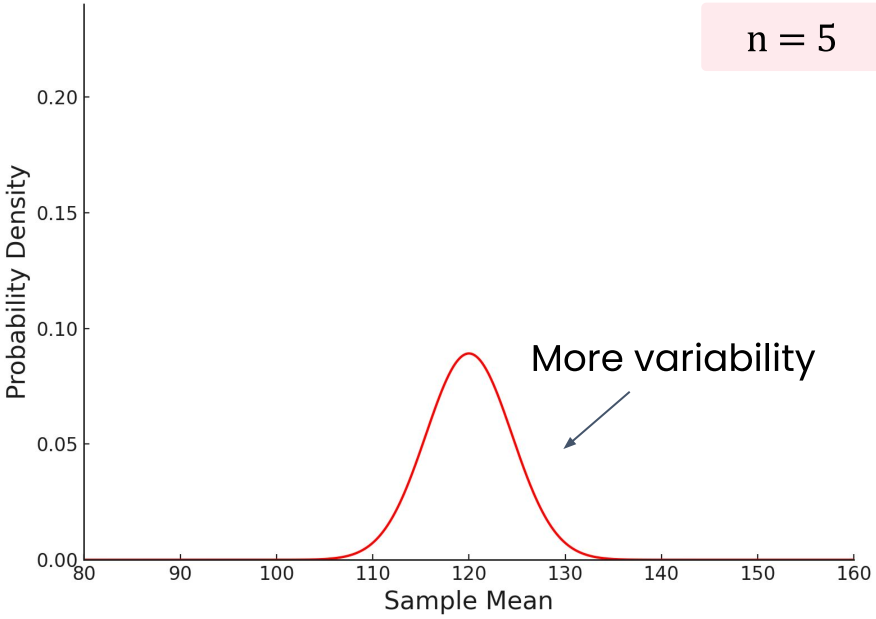
σ is rarely known

Use t distribution

$n = 30$



$n = 5$



t distribution

The normal and t distributions are similar:

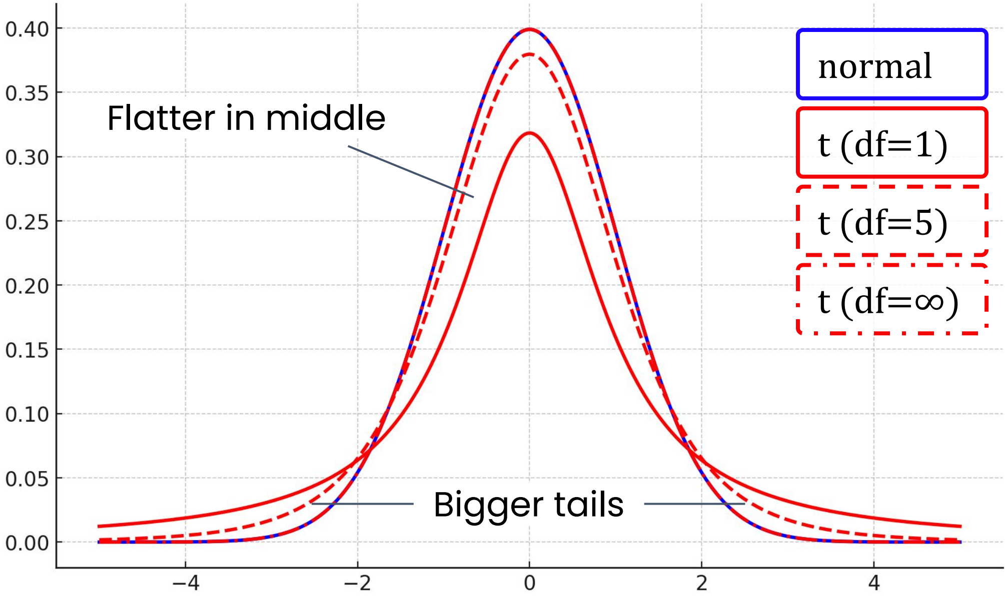
- Smooth shape
- Symmetric about the mean of 0

The t distribution is:

- Flatter in the middle
- Has bigger tails
- Defined by **degrees of freedom** (df)

$\mathbf{df} = n - 1$ ← Using s instead of σ to calculate SE

- As df increases, t distribution is more similar to normal distribution
- Choice of distribution becomes less important
- As n increases, you get better information about variability of population



DeepLearning.AI

Sean Barnes

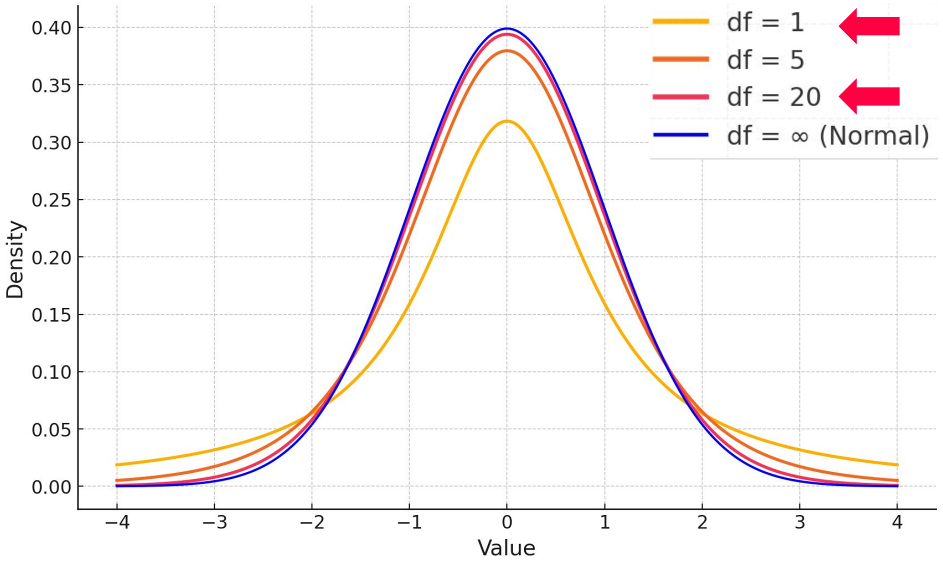
t distributions

The fundamental idea is:

- In order to calculate a confidence interval or perform a hypothesis test, you generally don't know the population standard deviation

- As your sample size increases, you get better information about variability of the population

- With smaller samples, you're working with a less precise estimate



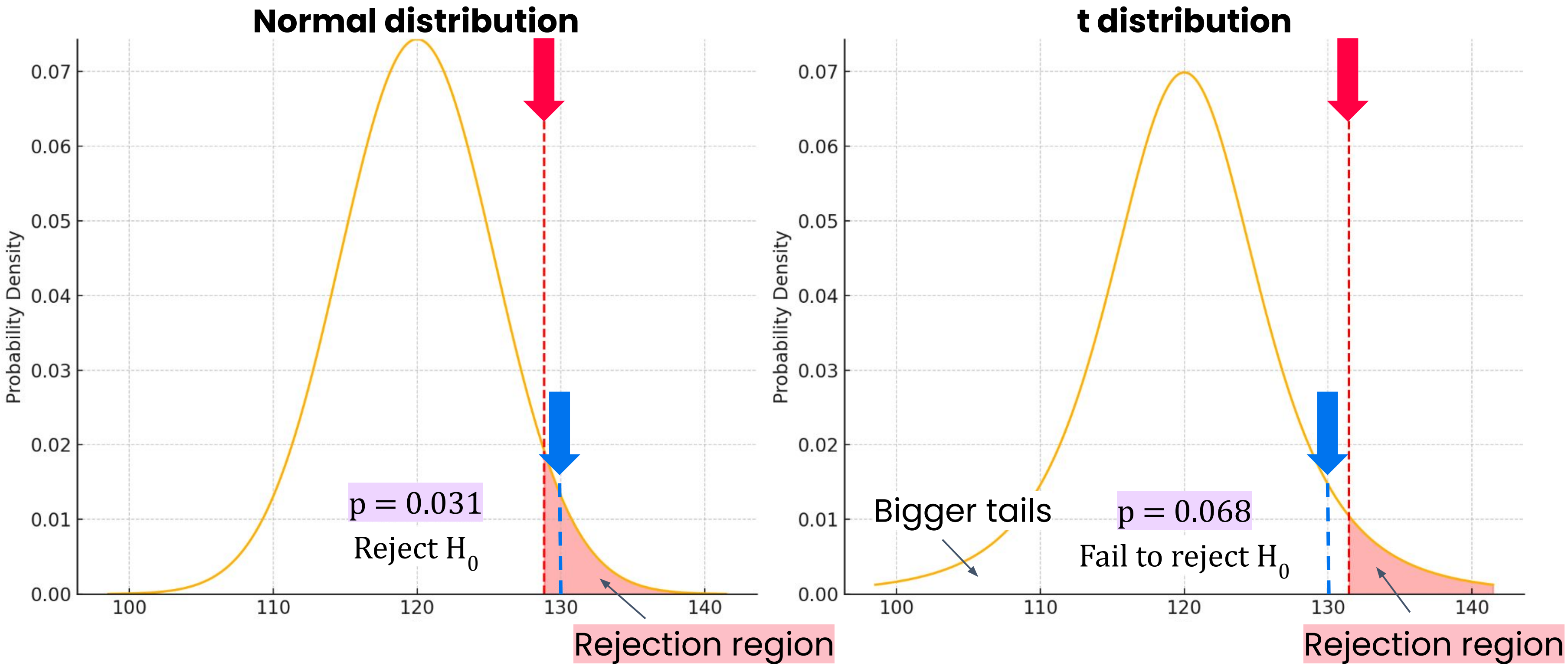
Hypothesis tests with the t distribution

	Normal distribution	t distribution
Defining hypotheses	Same!	
Calculate sample statistics	Same!	
Calculate test statistic	Normal distribution (z)	t-distribution (t)
Spreadsheet function	`z.test`	`t.test`

Hypothesis testing

$\bar{x} = 130$ mins

$\mu = 120$ $s = 12$ $n = 5$



DeepLearning.AI

Sean Barnes

Application of the t distribution

t distribution

$n < 30$

⭐ Still relevant!

Researching rare illnesses

Studying endangered species

Normal distribution

$n >= 30$

Big data

Tech

Manufacturing

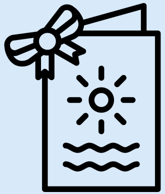
Healthcare

Hypothesis testing

Hypothesis testing for proportions

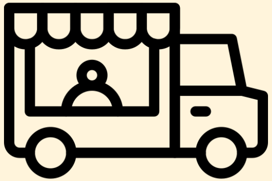
DeepLearning.AI

Proportions



What proportion of coworkers is in favor of birthday card?

Is proportion of valid canine DNA test kits 0.7?

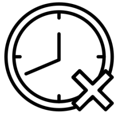


What proportion of deliveries to zoo are on time?

Scenario

You
Data Analyst

Problem: Whether true proportion of late deliveries is **greater than 5%**



More than 5% of deliveries are late

→

Risk jeopardizing contract with the zoo

You would want to:

⬜ Move up delivery driver's start time

Hypothesis testing for proportions

- $\mathbf{n}$ = 250 deliveries

Late **Not late**

$\hat{p}$ — Estimate for true proportion p

Hypothesis testing

- Define H_0 and H_1
- Determine α
- Calculate z and p-value
- Interpret results and make a decision

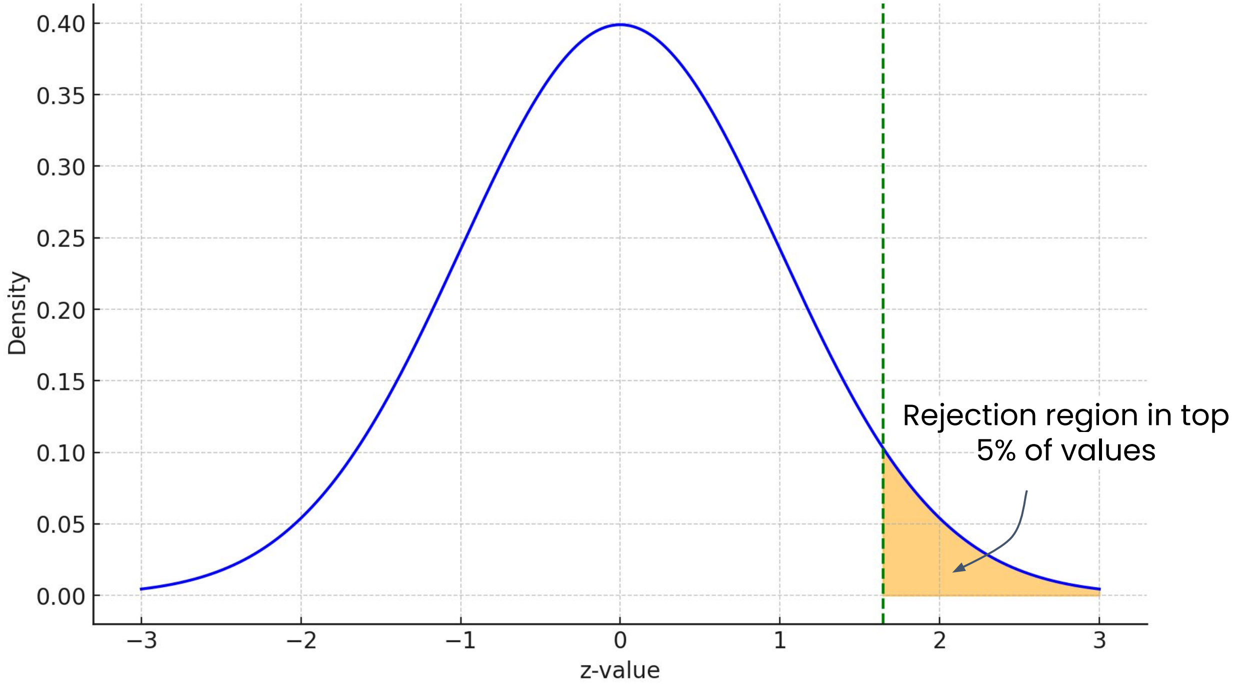
5% false positive rate

$H_0: p = 0.05$

$H_1: p > 0.05$

$\alpha = 0.05$

- $\mathbf{n} = 250$



$H_0 : p = 0.05$

$H_1 : p > 0.05$

$\alpha = 0.05$

Calculate test statistic

1. Measured sample proportion ($\hat{p}$) of late deliveries

Late → $\hat{p} = 0.06$

2. Calculate test statistic z

1. **Subtract hypothesized proportion from $\hat{p}$**
2. **Divide by the SE for the distribution**

$$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

SE → $\sqrt{\frac{p_0(1-p_0)}{n}}$

Centers z at 0

How many SEs $\hat{p}$ is from p_0

Why not use $\hat{p}$ for standard error?

Means

- μ is hypothesized
- σ often unknown
- Use s to approximate

Proportions

- Derive σ from p_0 and n
- Don't need to estimate

Calculate test statistic

1. Measured sample proportion ($\hat{p}$) of late deliveries

Late → $\hat{p} = 0.06$

2. Calculate test statistic z

1. **Subtract hypothesized proportion from $\hat{p}$**
2. **Divide by the SE for the distribution**

$$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Centers z at 0

SE → $\sqrt{\frac{p_0(1-p_0)}{n}}$

How many SEs $\hat{p}$ is from p_0

$$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.06 - 0.05}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.01}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.01}{\sqrt{\frac{0.05\times0.95}{n}}} = \frac{0.01}{\sqrt{\frac{0.0475}{n}}} = \frac{0.01}{\sqrt{\frac{0.0475}{250}}} = \frac{0.01}{\sqrt{0.00019}} = \frac{0.01}{0.01378} = 0.725$$

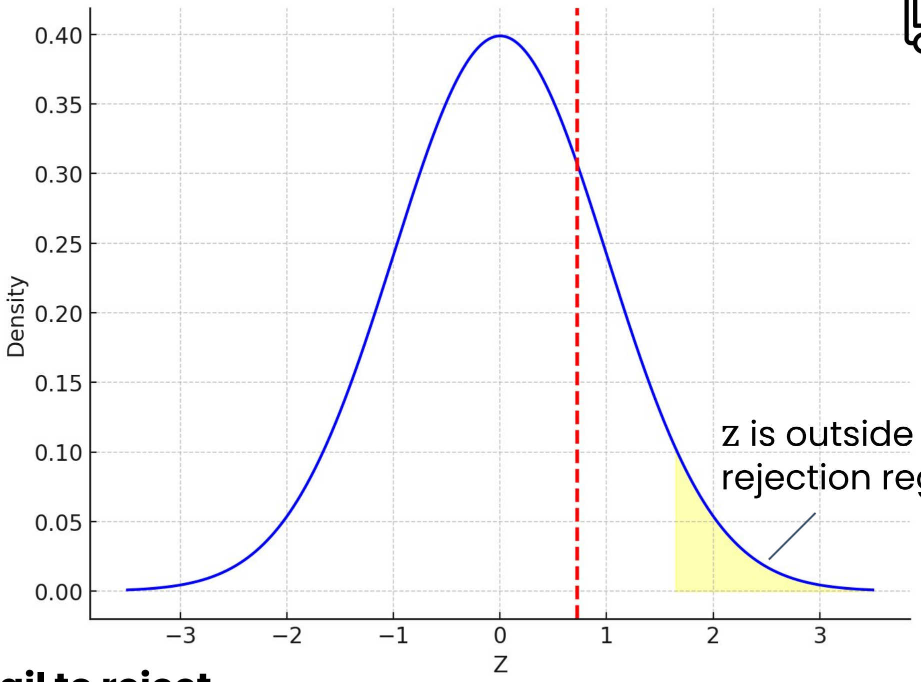
z is above the mean by less than 1 SE

Calculate test statistic

How rare would a value of 0.725 or higher be?

$$z = 0.0725$$

Do you think you would reject or fail to reject H_0?



Fail to reject

$H_0 : p = 0.05$

$H_1 : p > 0.05$

$\alpha = 0.05$

Calculate p-value and compare with α

3️⃣ Calculate p-value

You can use:

⬜ A lookup table

⬜ Software

$$P(z >= 0.725) = 0.2342$$

$$\text{If } p = 0.05, \;\; 23.42\% \text{ of } \hat{p}\text{s will be} >= 0.05$$

4️⃣ Compare p-value with α

$\alpha = 0.05$

- 0.373 is greater than α
- Fail to reject H_0

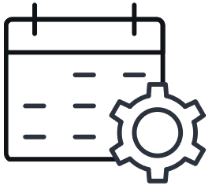
$H_0 : p = 0.05$

Scenario

You
Data Analyst

Your interpretation of the results:

- **Not sufficient** evidence to conclude that late deliveries happen > 5% of the time.



✅ Keep start time the way it is

✅ Contract with zoo should remain in good standing

Hypothesis testing

Demo: hypothesis testing for proportions

DeepLearning.AI

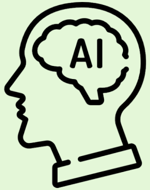
Hypothesis testing

Two sample tests

DeepLearning.AI

Two sample tests

Do movies from 2013 have the same duration as movies from 1934?



Is proportion of people with positive opinion of AI the same across age groups?

Do users with basic and premium stay subscribed for different lengths?

Two sample tests

Do users with **basic** and **premium** stay subscribed for **different** lengths?

$$\alpha = 0.05$$

Hypothesis testing

1. Define H_0 and H_1
2. Determine α
3. Calculate z or t and p-value
4. Interpret results and make a decision

- $H_0 : \mu_{\text{basic}} = \mu_{\text{premium}}$ → no difference
- $H_1 : \mu_{\text{basic}} \neq \mu_{\text{premium}}$ → means are different

Two sample tests

Do users with **basic** and **premium** stay subscribed for **different** lengths?

$\alpha = 0.05$

Basic subscribers:

$n_{basic} = 30$

$\bar{x}_{basic} = 9.9$ months

$s_{basic} = 3.3$ months

Premium subscribers:

$n_{premium} = 30$

$\bar{x}_{premium} = 10.4$ months

$s_{premium} = 1.9$ months

Independent

Next step: calculate test statistic (t)

- ⬜ Spreadsheet
- ⬜ Programming language

Two sample tests

Do users with **basic** and **premium** stay subscribed for **different** lengths?

$\alpha = 0.05$

Basic subscribers:

$n_{basic} = 30$

$\bar{x}_{basic} = 9.9$ months

$s_{basic} = 3.3$ months

Premium subscribers:

$n_{premium} = 30$

$\bar{x}_{premium} = 10.4$ months

$s_{premium} = 1.9$ months

Independent

Subscription for **basic** not influenced by **premium**

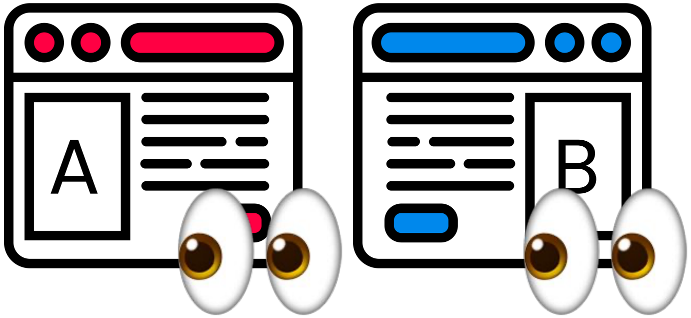
Promotional rate ↓ Basic ↑ Premium

Next step: calculate test statistic (t)

- [ ] Spreadsheet
- [ ] Programming language

Two sample tests

A/B testing

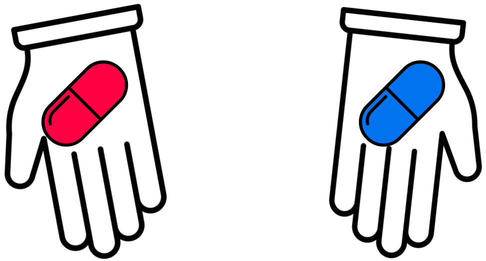


Do groups respond differently?

Clinical research

Experimental

Control



Compare whether experimental group showed more benefits

Hypothesis testing

Other hypothesis tests

DeepLearning.AI

Keep in mind...

🚫 Don’t need to memorize each of the tests

✅ As you need them, you can look up specifics

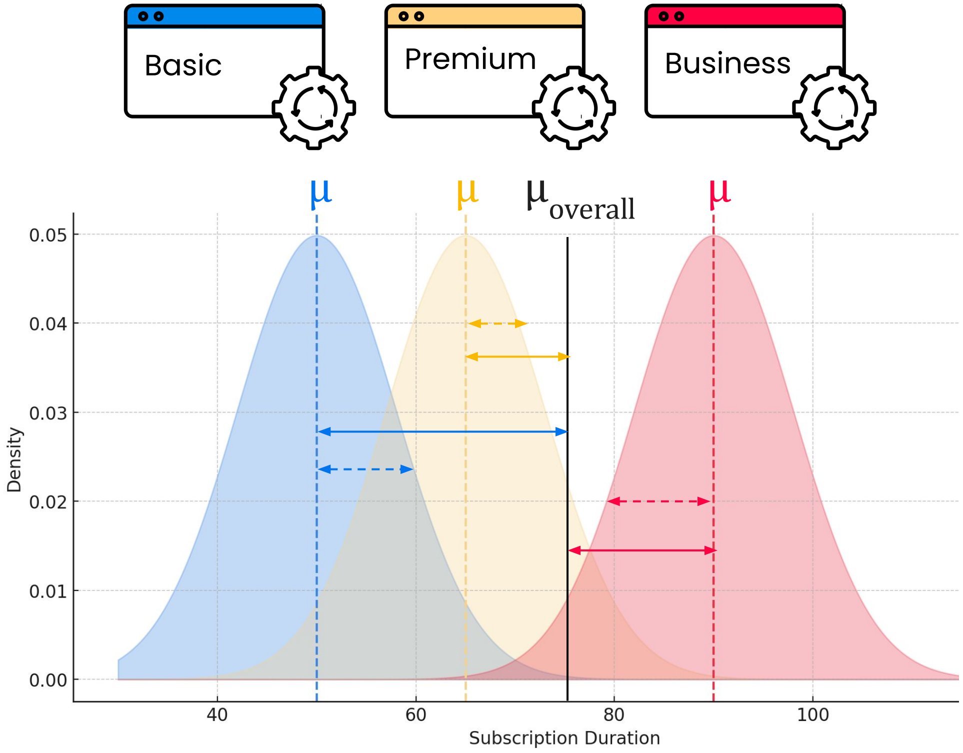
✅ You already have foundational knowledge to conduct and interpret them

ANOVA

Scenario: Comparing three or more groups

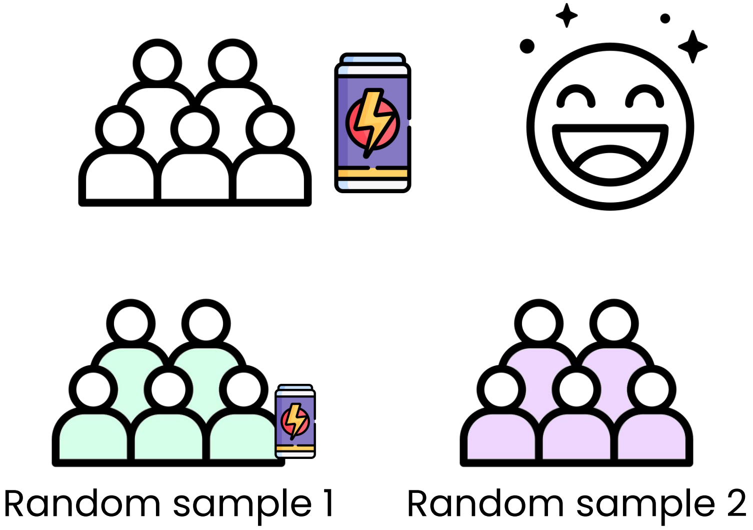
Test: Analysis of variance (ANOVA)

1. Calculate the means for each group
2. Calculate the overall mean
3. Compare:
    - How much group means differ from overall mean
    - How much scores differ from group means
4. **A small p-value (typically < 0.05)**: suggests differences between groups are significant



Paired t-test

Scenario: Working with data that represents a before and after condition

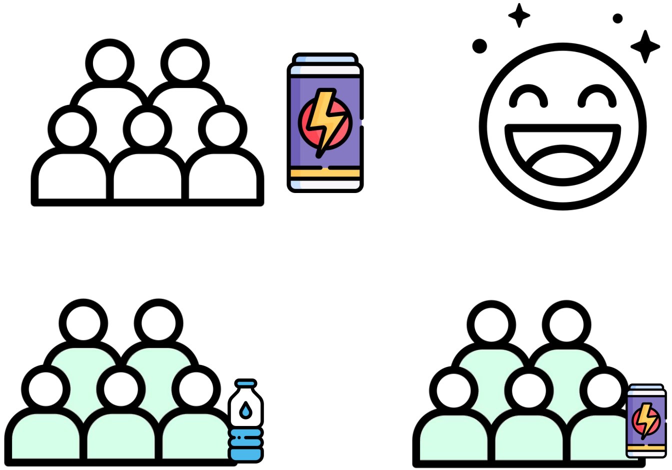


Paired t-test

Scenario: Working with data that represents a before and after condition

Test:

1. Test the same group of people twice
2. Calculate difference between each pair of measurements
3. Calculate the test statistic by dividing the mean difference by SE
4. **p-value (< 0.05)**: often considered significant



More information!

Chi-squared test

Scenario: Working with categorical data

Test:

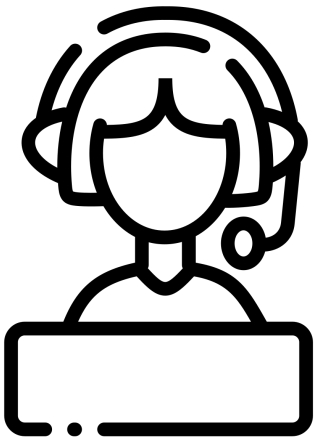
1. Create table of observed frequencies
2. Calculate expected frequencies assuming no relationship
3. Calculate chi-squared statistic: how much observed frequencies deviate

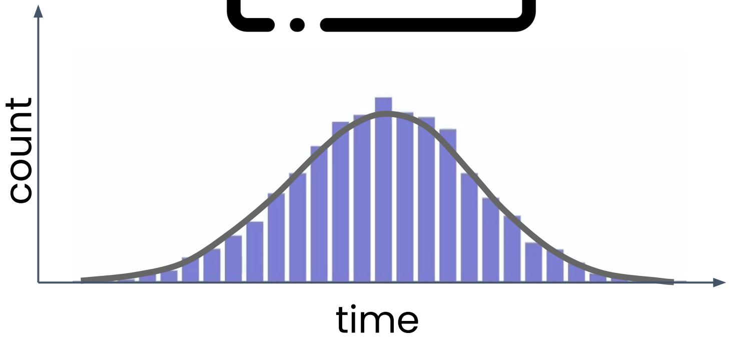
Goodness of fit test

Scenario: Test whether data follows a particular distribution

Test:

1. Calculate test statistic measuring deviation from distribution
2. Determine p-value





Hypothesis testing

Interpretation with LLMs

DeepLearning.AI

Hypothesis testing

Inference with LLMs

DeepLearning.AI

Applied Statistics for Data Analytics

Your next steps

DeepLearning.AI