Statistics and Hypothesis Testing

CISC 3225
Spring 2024
DSFS 5, 6, 7

Introduction

When we ask questions about data, we often form a ***hypothesis***: a supposition or proposed explanation made on the basis of limited evidence as a starting point for further investigation.

To formulate a hypothesis, it is useful to have domain knowledge or a familiarity with your data.

Wine example

Hypothesis: Sweeter wine has less alcohol than dry wine.

Basis:

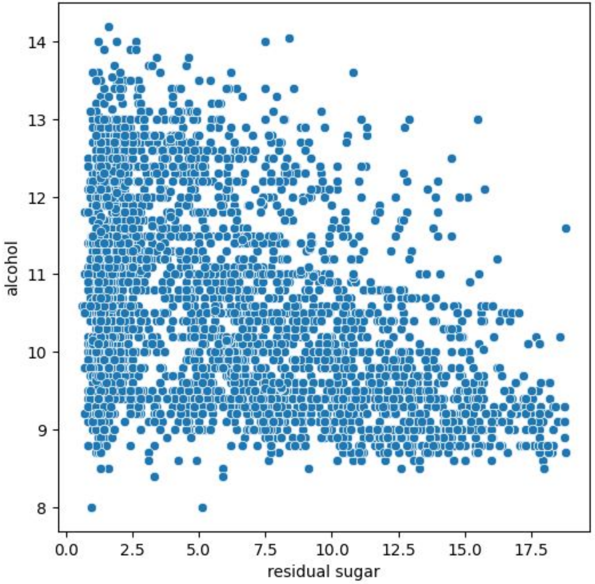
- Wine-related domain knowledge (homebrewer, wine enthusiast)
- Familiarity with biological processes (yeast convert sugar to alcohol)
- Independent research

Wine example

Hypothesis: Sweeter wine has less alcohol than dry wine.

Exploratory analysis: Visualize the relationship

- It *appears like* there is more alcohol when there is less sugar, and less alcohol when there is more sugar.
- Good for a quick exploration, but…
- …can we quantify this relationship?



Visualized from white wine dataset, outliers removed (residual sugar outside 99% quantile)

Statistical hypothesis testing

Statistical hypothesis testing allows us to determine *whether our data supports a hypothesis.*

Process for statistical hypothesis testing:

1. Formulate a hypothesis.
2. Gather relevant data (existing datasets or experiments)
3. Formulate a *null hypothesis* (H_0) that represents a default position
4. Formulate an *alternative hypothesis* (H_1) to compare with H_0
5. Use statistical tests to determine whether we can reject H_0

Statistical hypothesis testing

Null hypothesis (H_0): Default position
Alternative hypothesis (H_1) compared with H_0

As a result of hypothesis testing, we can do one of the following:

- Reject the null hypothesis
- Fail to reject the null hypothesis

Statistical hypothesis testing

Null hypothesis (H_0): Default position
Alternative hypothesis (H_1) compared with H_0

As a result of hypothesis testing, we can do one of the following:

- Reject the null hypothesis
- Fail to reject the null hypothesis

		Reality	
Test		H_0 true	H_0 false
	Fail to reject H_0	OK	Type 2 error (False negative)
	Reject H_0	Type 1 error (False positive)	OK

Flipping a Coin

Problem: How do we determine if a coin is fair?

1. Hypothesis: A coin has a 50% chance of landing on heads
   (The coin has a probability *p* of landing heads, and *p*=0.5)
2. Data: Flip a coin many times and count the number of heads (<u>X</u>).
3. H_0: p = 0.5
4. H_1: p ≠ 0.5
5. Statistical tests: ???

Random variable

Flipping a Coin: Results

X: Number of heads

$H_0: p = 0.5$

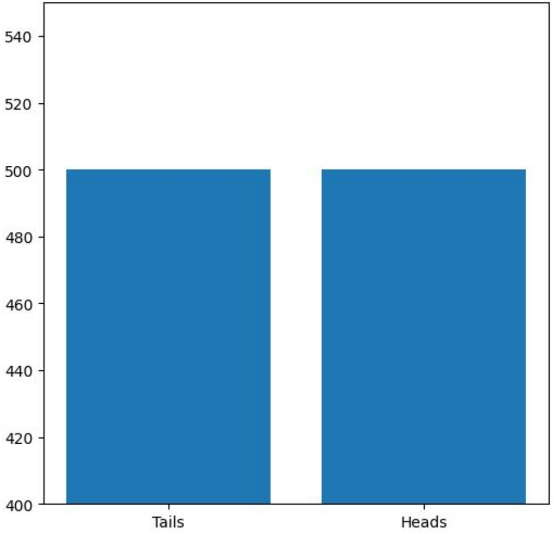
$H_1: p \neq 0.5$

Results from flipping the coin 1000 times:

Easy case:

X: 500

Is the coin fair?



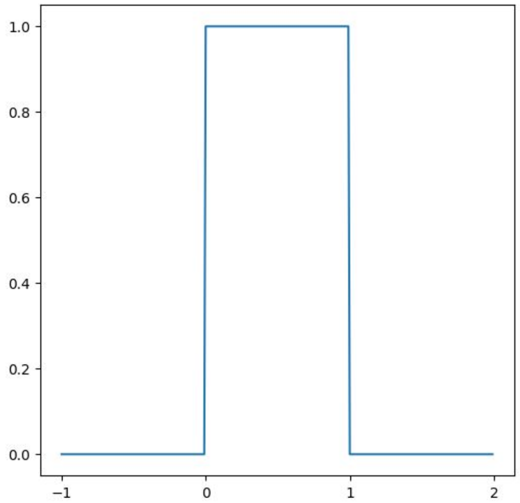
H_0: $p = 0.5$
H_1: $p \neq 0.5$
Heads = 1
Tails = 0

Flipping a Coin

If the coin were fair, what distribution of heads and tails would we expect to see?

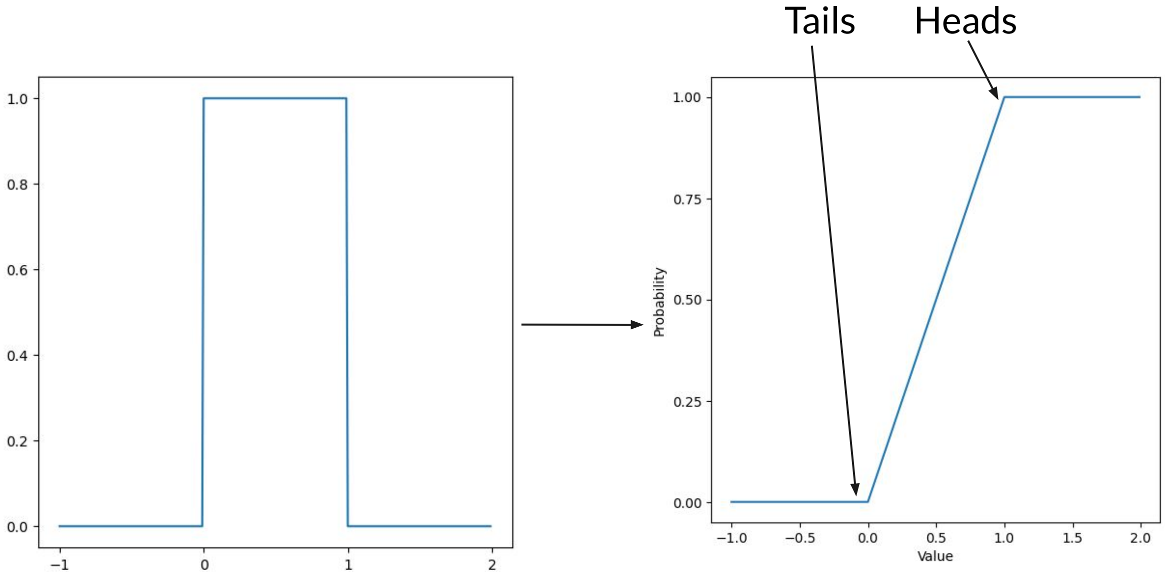
A **uniform distribution**: There is equal chance of heads (1) or tails (0) occurring.

The density function of the uniform distribution is shown to the right:



Flipping a Coin

$H_0: p = 0.5$
$H_1: p \neq 0.5$
Heads = 1
Tails = 0



Cumulative distribution function (CDF): The probability that a random variable is less than or equal to a certain value.

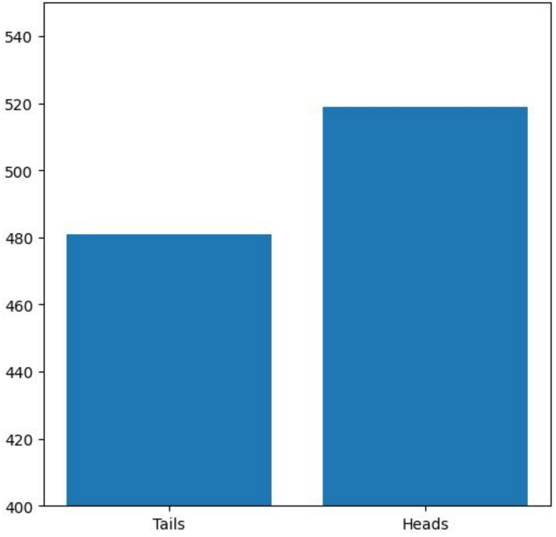
Flipping a Coin: Results

X: Number of heads

$H_0: p = 0.5$

$H_1: p \neq 0.5$

Results from flipping the coin 1000 times:



Harder case:

X: 519

Is the coin fair?

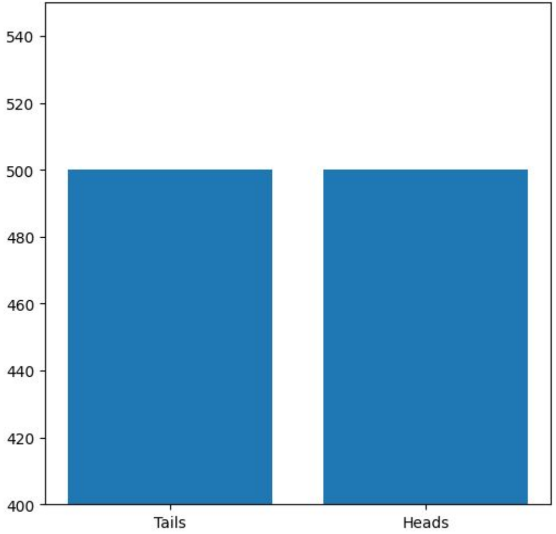
Flipping a Coin: Problem

$H_0: p = 0.5$

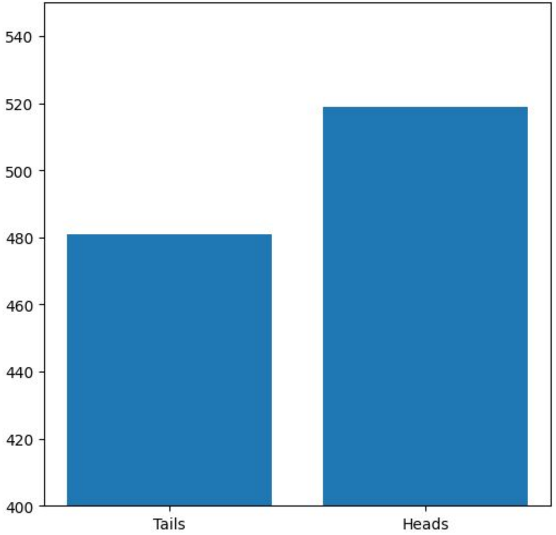
$H_1: p \neq 0.5$

Heads = 1

Tails = 0



This experiment produced perfectly uniform head counts.



This experiment *did not.*

Flipping a Coin: Binomial variables

X: Number of heads in a coin-flipping experiment

X is a ***random variable***: a variable whose possible values have an associated probability distribution.

Heads = 1
Tails = 0

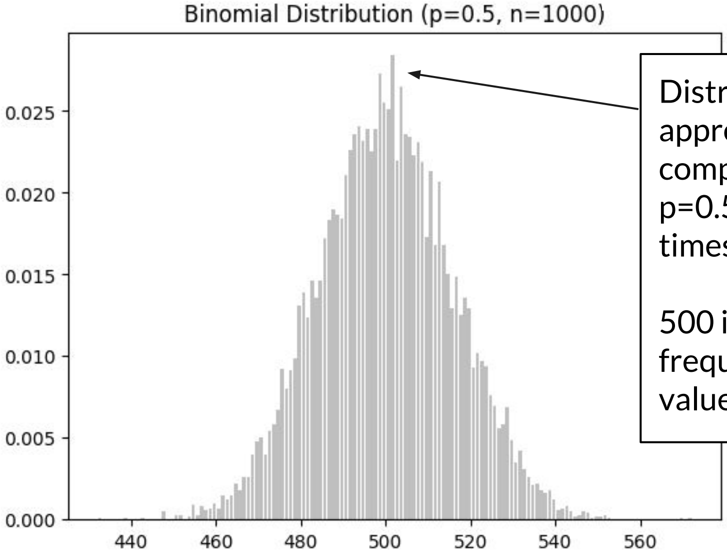
X is a *binomial random variable*, which has two parameters:

- n: Total number of trials, each with a binary outcome (a Bernoulli trial)
- p: The probability of one trial outputting 1

X: Number of heads

H_0: $p = 0.5$

H_1: $p \neq 0.5$



Distribution approximated by computing X with p=0.5, n=1000 many times.

500 is the most frequently occurring value of X

Flipping a Coin: Results

X: Number of heads

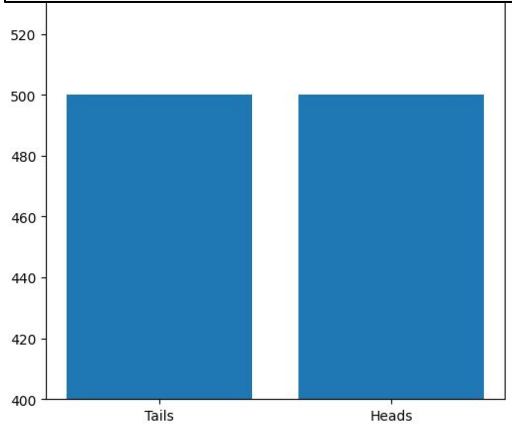
$H_0: p = 0.5$

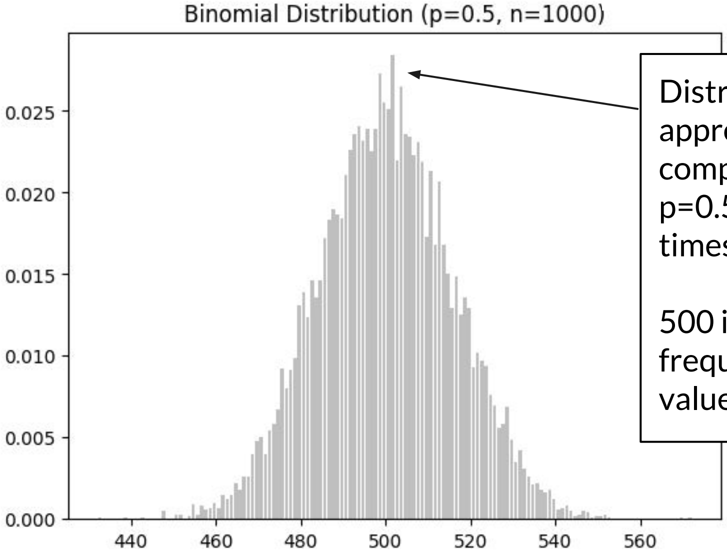
$H_1: p \neq 0.5$

Central question: If X is binomial random variable with p=0.5, n=1000, **how likely** is the outcome shown below?

- Unlikely: Reject the null hypothesis
- Likely: Fail to reject the null hypothesis

How likely is this outcome?





Distribution approximated by computing X with p=0.5, n=1000 many times.

500 is the most frequently occurring value of X

Flipping a Coin: Results

X: Number of heads

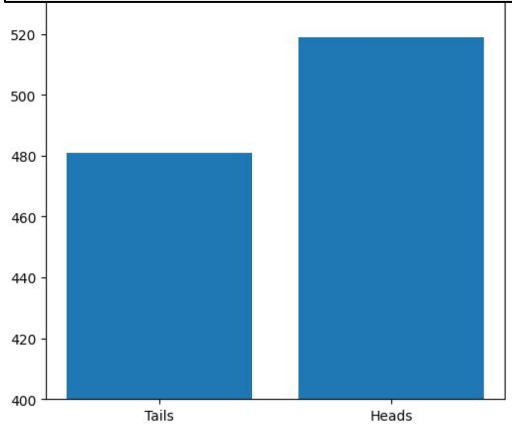
H_0: p = 0.5

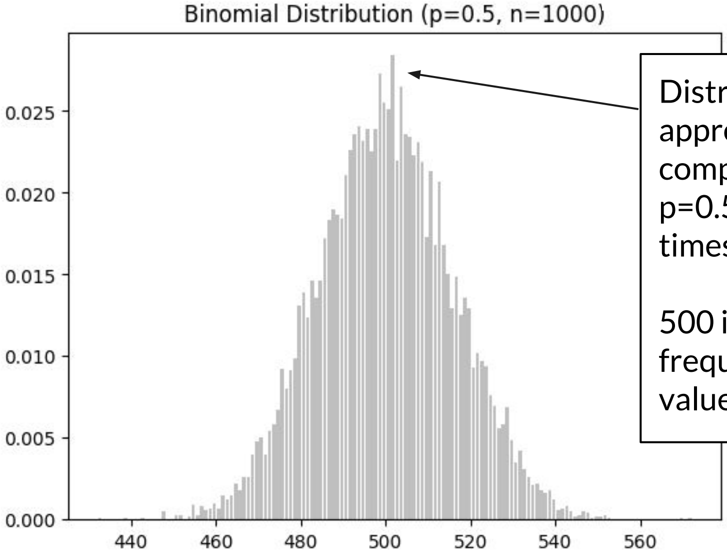
H_1: p ≠ 0.5

Central question: If X is binomial random variable with p=0.5, n=1000, **how likely** is the outcome shown below?

- Unlikely: Reject the null hypothesis
- Likely: Fail to reject the null hypothesis

How likely is this outcome?





Distribution approximated by computing X with p=0.5, n=1000 many times.

500 is the most frequently occurring value of X

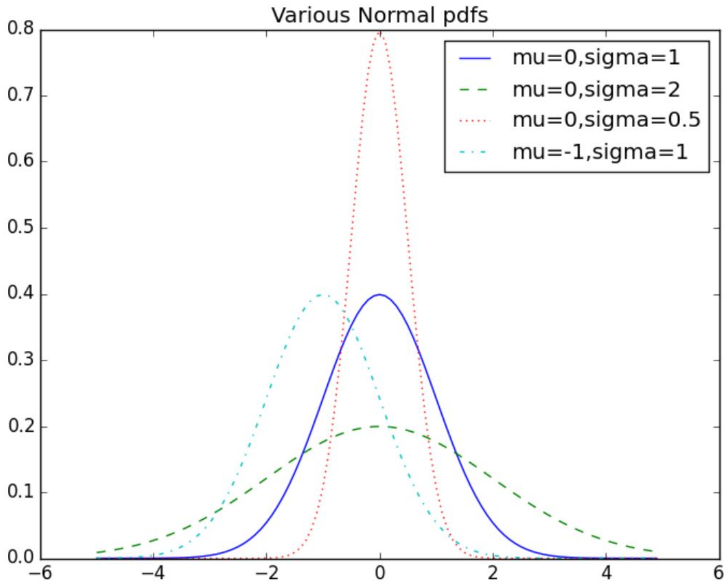
Side note: The normal distribution

The classic bell curve.

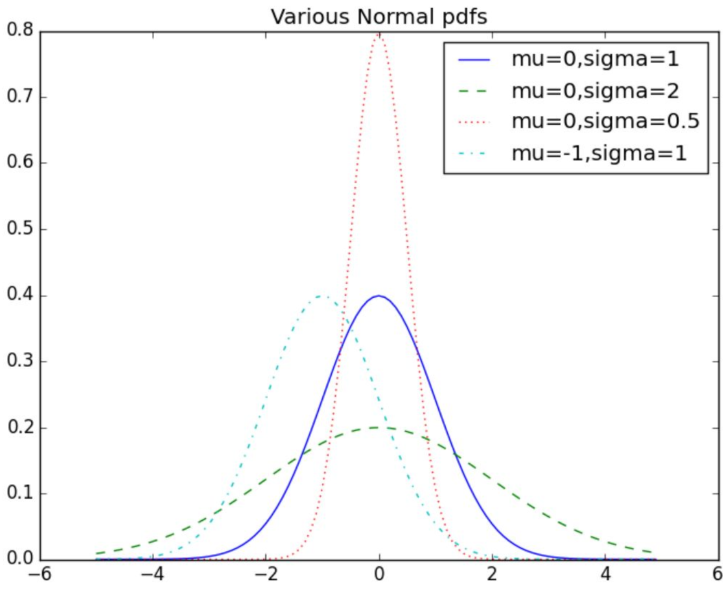
Probability density function: $f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

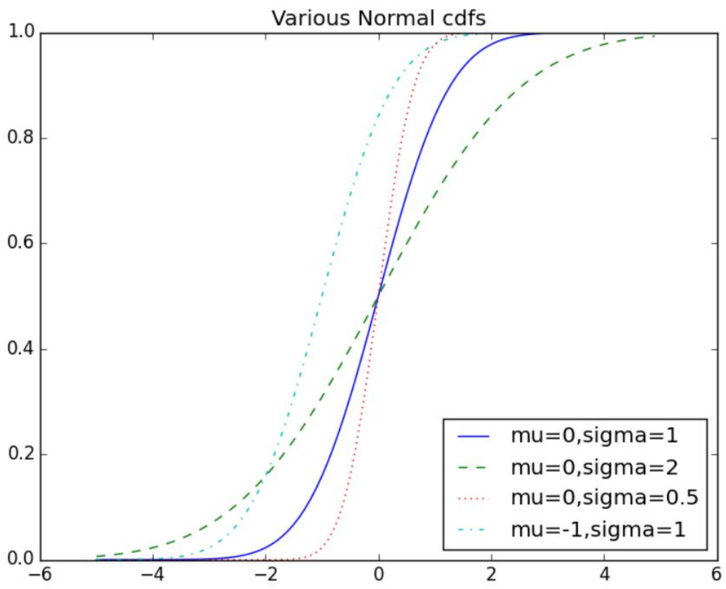
Two parameters:

- **μ (Mu): Mean**, or center of the bell
- **σ (Sigma): Standard deviation**, or how wide the bell is



Side note: The normal distribution CDF





H_0: p = 0.5
H_1: p ≠ 0.5
Heads = 1
Tails = 0

Approximating binomial variables

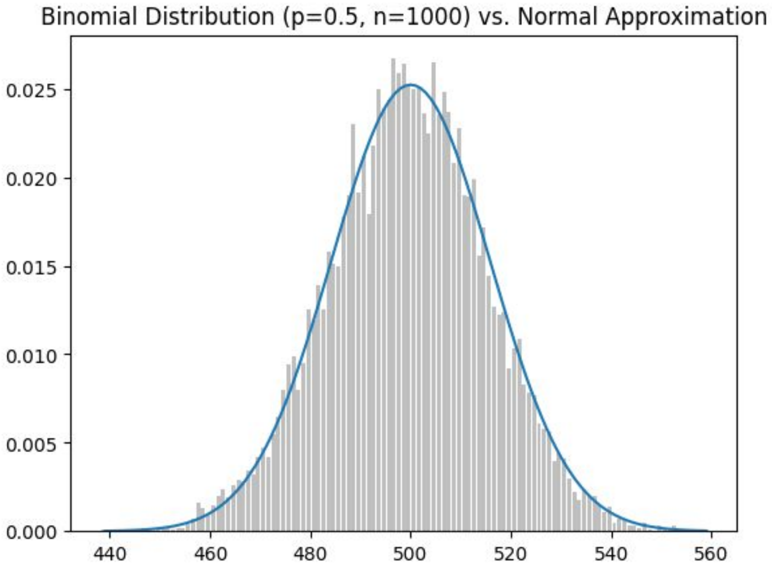
Variables:

- *p*: Probability of flipping heads (0.5)
- 1-*p*: Probability of flipping tails (0.5)
- *n*: Number of trials (1000)

$\mu = pn = (0.5)(1000) = 500$

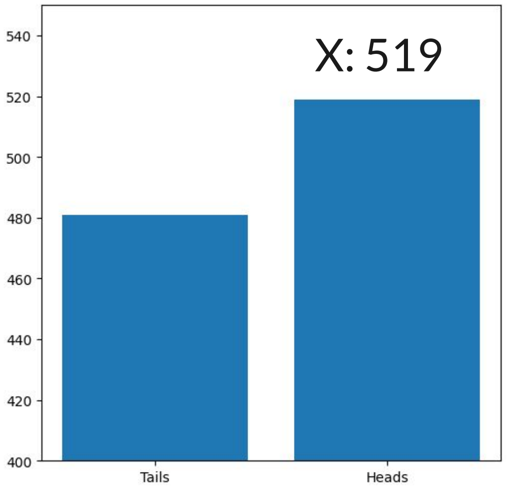
$\sigma = \sqrt{p(1 - p)n} = \sqrt{(0.5)(0.5)(1000)} = \sqrt{250} \approx 15.8$

If H_0 is true, X should be distributed *approximately normally* with mean 500 and standard deviation 15.8



Significance testing

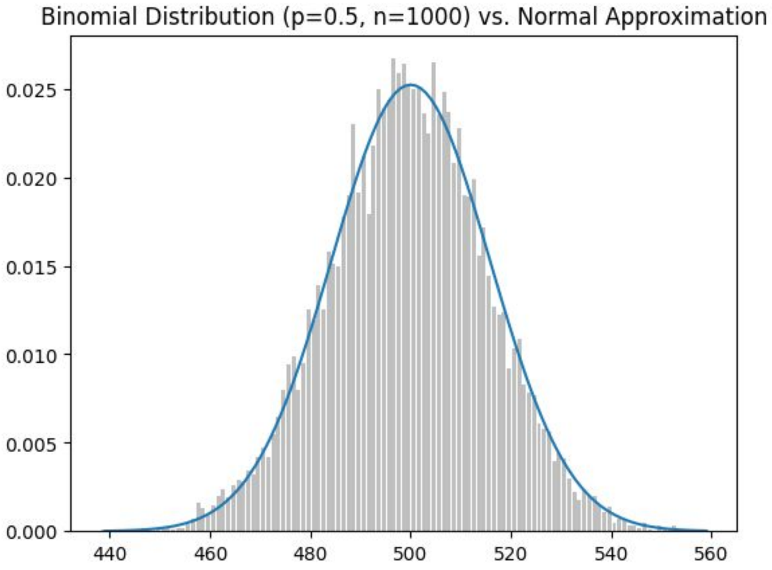
H_0: p = 0.5
H_1: p ≠ 0.5
Heads = 1
Tails = 0



Should we reject the null hypothesis?

Alternate question: **How willing are we to make a type 1 (false positive) error? (Significance)**

By convention, the significance is usually **5%** (low tolerance for type 1) or **1%** (very low tolerance for type 1)



Statistical hypothesis testing

Null hypothesis (H_0): Default position
Alternative hypothesis (H_1) compared with H_0

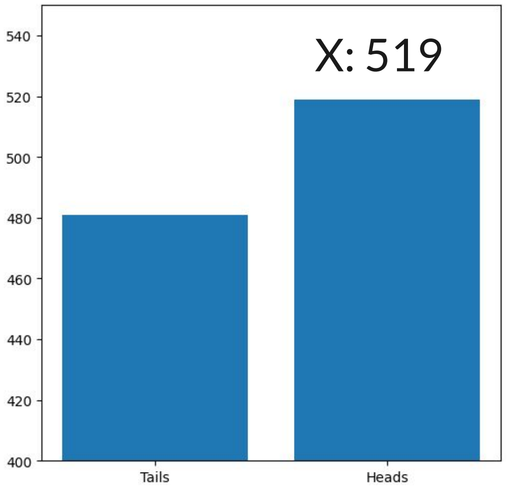
As a result of hypothesis testing, we can do one of the following:

- Reject the null hypothesis
- Fail to reject the null hypothesis

	Reality		
Test		H_0 true	H_0 false
	Fail to reject H0	OK	Type 2 error (False negative)
	Reject H0	Type 1 error (False positive)	OK

Significance testing

$H_0: p = 0.5$
$H_1: p \neq 0.5$
Heads = 1
Tails = 0

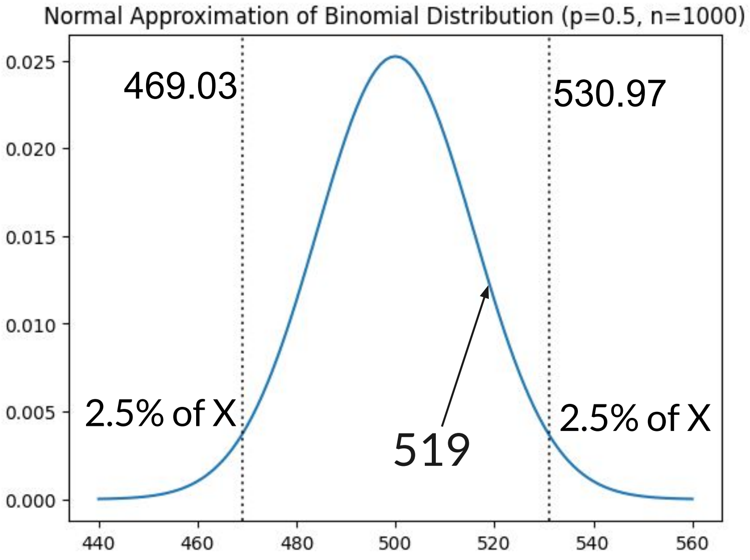


Significance of 5%:

Reject the null hypothesis if X falls outside of 95% of the distribution.

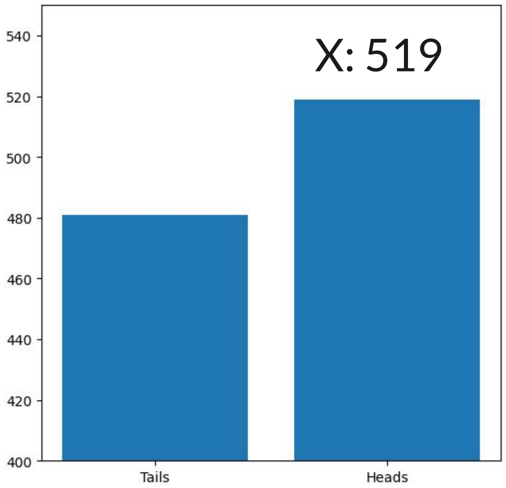
Fail to reject the null hypothesis if X falls within 95% of the distribution.

We *fail to reject the null hypothesis.*



H_0: p = 0.5
H_1: p ≠ 0.5
Heads = 1
Tails = 0

Significance testing

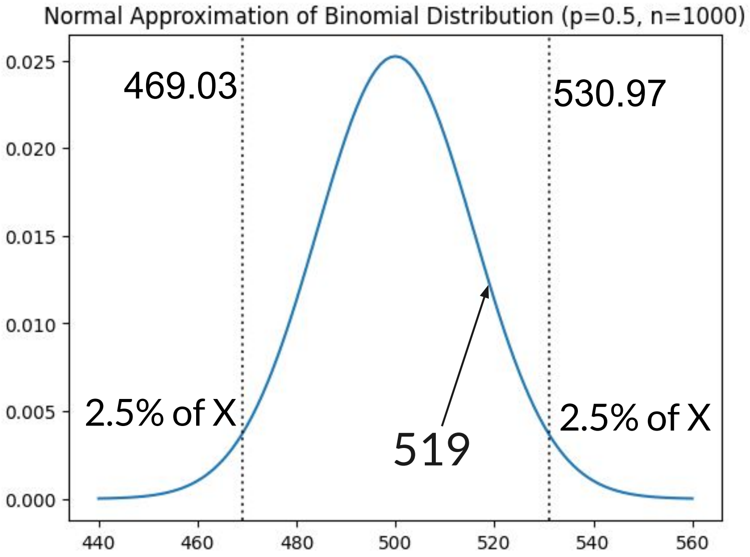


Significance of 5%:

Alternate interpretation: There is a 5% chance that a value of X falls below 469.03 or above 530.97.

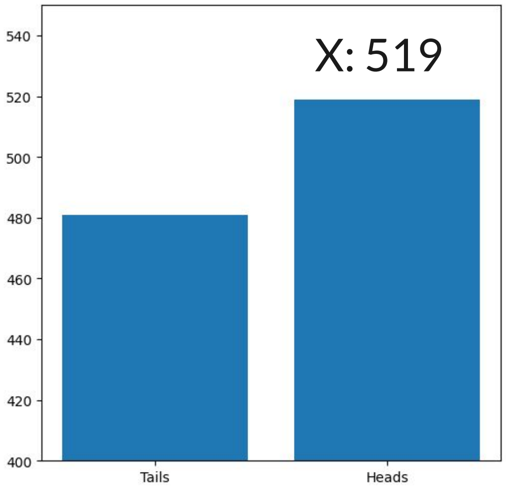
This test will show the correct result 95% of the time.

With 20 coin flip trials, assuming H_0 is true, p=0.5, n=1000, we will make a Type 1 error once.



Significance testing

H_0: p = 0.5
H_1: p ≠ 0.5
Heads = 1
Tails = 0

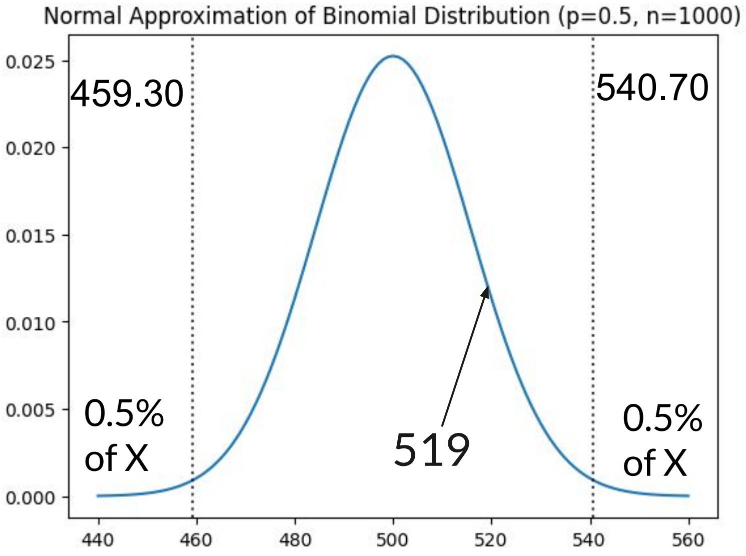


Significance of 1%:

Alternate interpretation: There is a 1% chance that a value of X falls below 459.30 or above 540.70.

This test will show the correct result 99% of the time.

With 100 coin flip trials, assuming H_0 is true, p=0.5, n=1000, we will make a Type 1 error once.



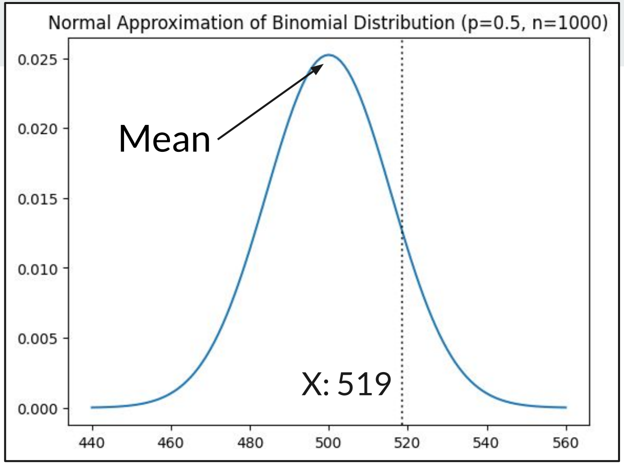
p-values

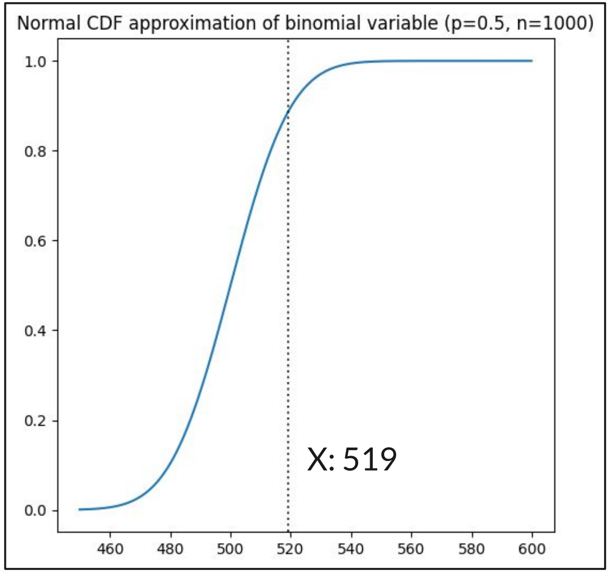
Definition of *p*-value:

- Assuming H_0 is true…
- …and we observe some value X…
- …*p* is the probability that we would see a value *at least as extreme as X*

Computing *p*-values

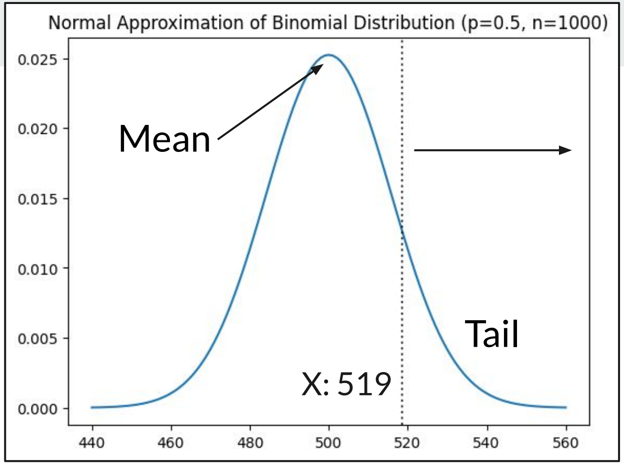
X is greater than the mean.

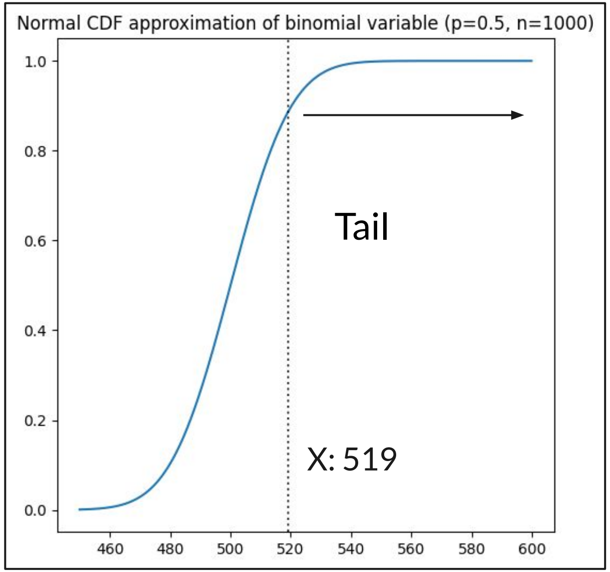




Computing *p*-values

X is greater than the mean, the *tail* is everything greater than X.



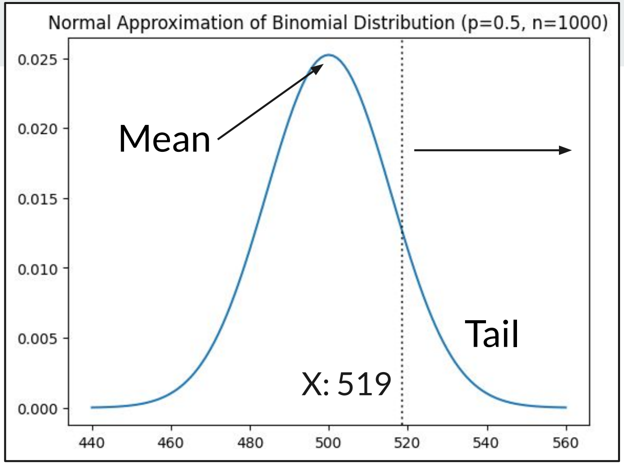


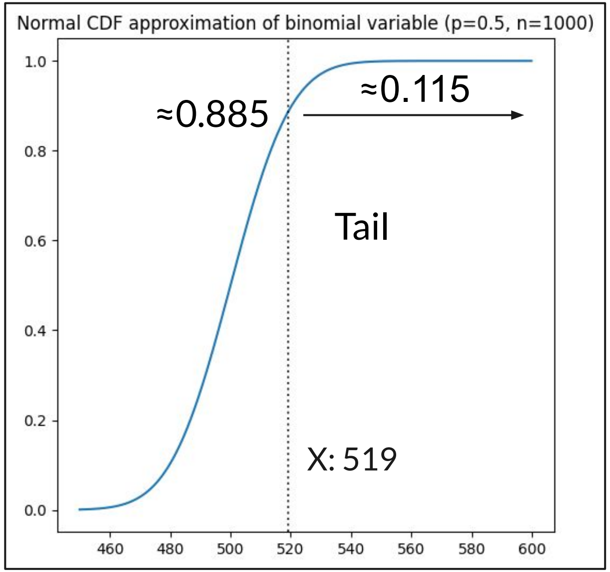
Computing *p*-values

X is greater than the mean, the *tail* is everything greater than X.

p equals 2 * the *probability of encountering a value in the tail*. (multiply 2 to consider both sides)

$p = 0.115 * 2 \approx 0.23$





$H_0: p = 0.5$
$H_1: p \neq 0.5$

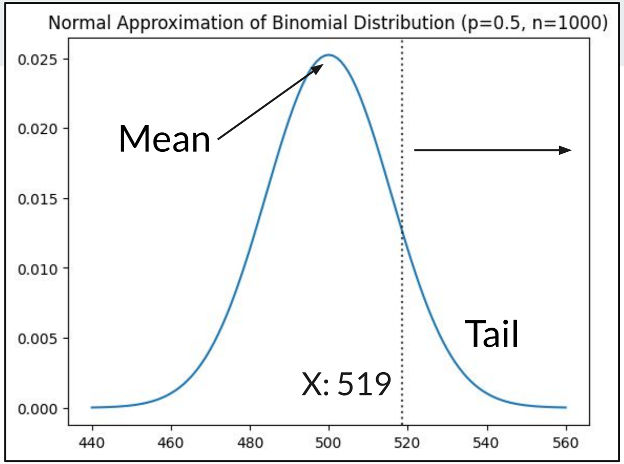
p-values and significance testing

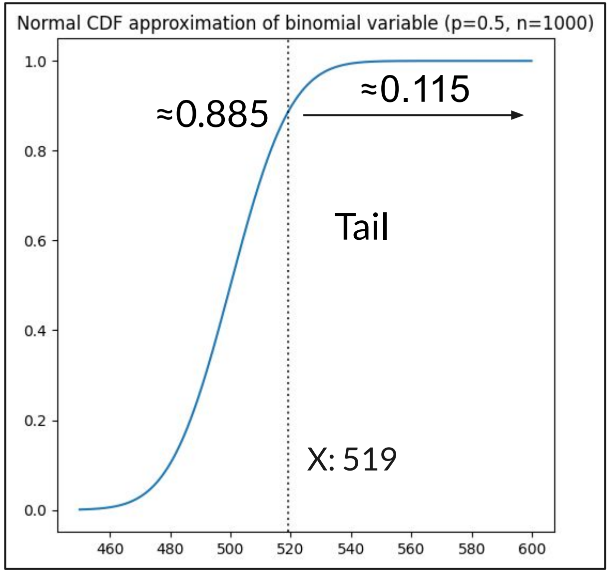
If *p* is less than or equal to our significance level (5% or 1%), reject the null hypothesis.

If *p* is greater than our significance level, fail to reject the null hypothesis.

$p = 0.115 * 2 \approx 0.23 > 0.05$
$\rightarrow$ Fail to reject the null hypothesis

The coin is likely a fair coin.





Closing notes

- *p* values are a simple and commonly used method of hypothesis testing.
- Many statistical tests output a *p* value:
  - *p*-values can be interpreted as a signal of the strength of what you are testing.
  - *p*-values are controversial, and some research relies on them too much.
  - *p*-values are susceptible to *p*-hacking, a technique to manipulate results.