Hypothesis testing
& statistical significance

Why?

To demonstrate that an observed pattern is unlikely to be occur by random chance

Samples and population

Samples and population

In the statistical framework, observations/measurements, or statistics of the data are *samples* of a complete *population*.

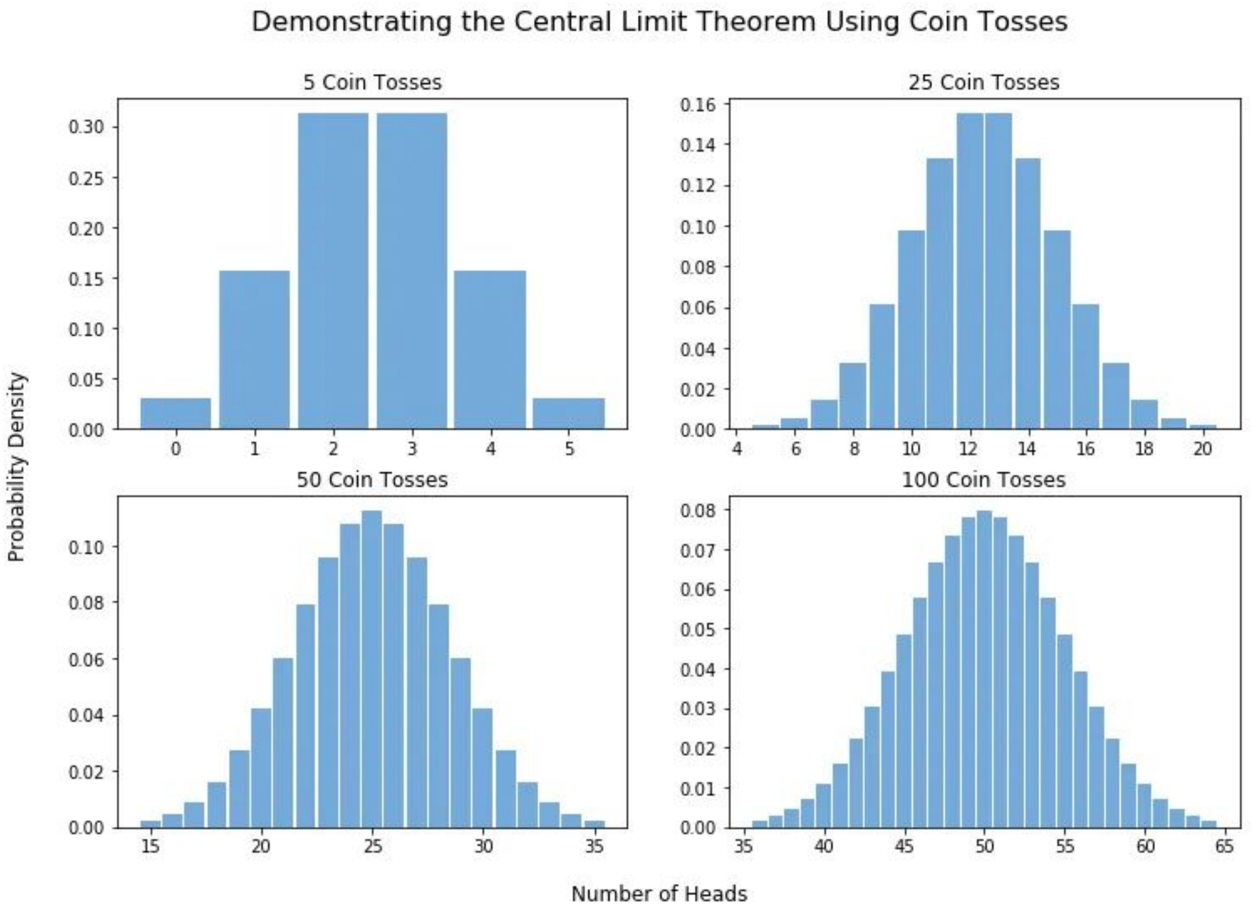
Samples and population

- μ (*mu*) represents the population mean, while $\bar{x}$ (*x-bar*) represents the sample mean
- σ (*sigma*) represents the population standard deviation, and s represents the sample standard deviation

Intuition:

more data =

more certainty



https://medium.com/@WDSS/gibrats-law-the-central-limit-theorem-s-forgotten-twin-b67b23c45ee1

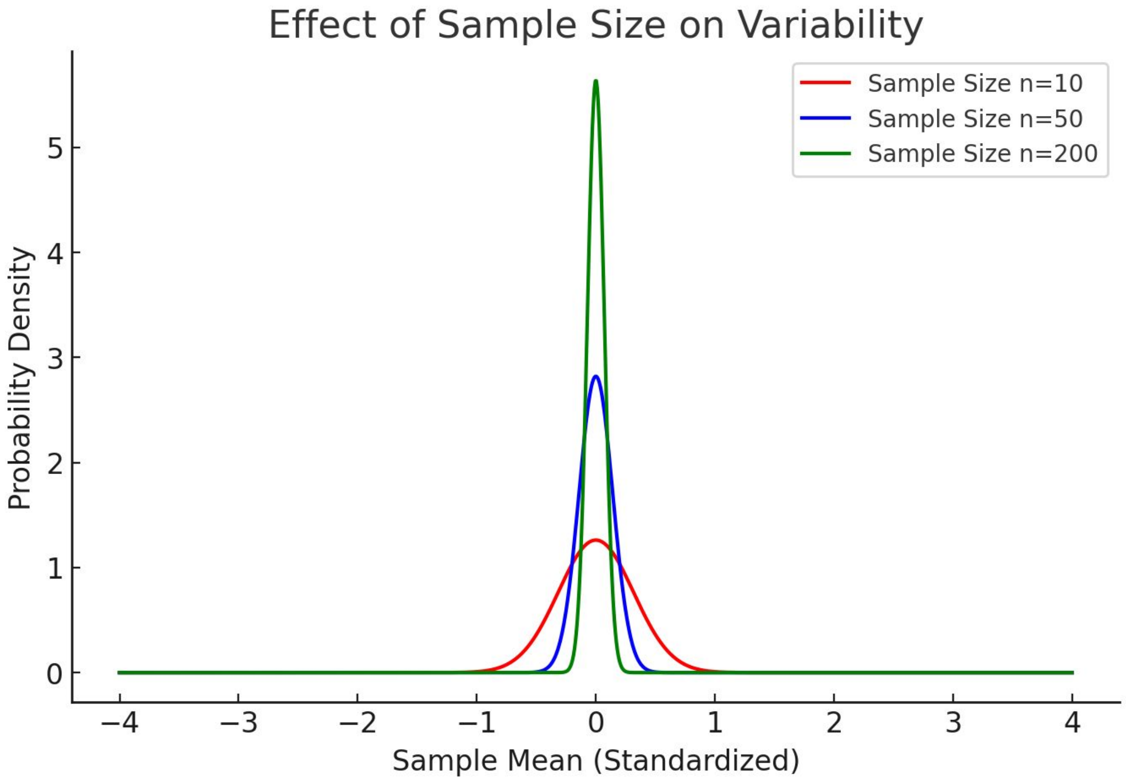
Central limit theorem

- If we take many samples of a population, statistics (like the mean or mean difference) will show variability that follows a normal distribution.
- The distribution of means of a summary statistic over repeated samples follows a normal distribution.

Sample size effects variability

- With a small sample, the sample mean can fluctuate greatly due to random variation.
- When increasing the sample size, the mean estimates become more stable, reducing variability.

Sample size effects variability



Sample size effects variability

With larger samples:

- Get more precise estimates of the true mean.
- The likelihood of observing extreme values decreases.
- Statistical tests become more powerful, detecting smaller differences.

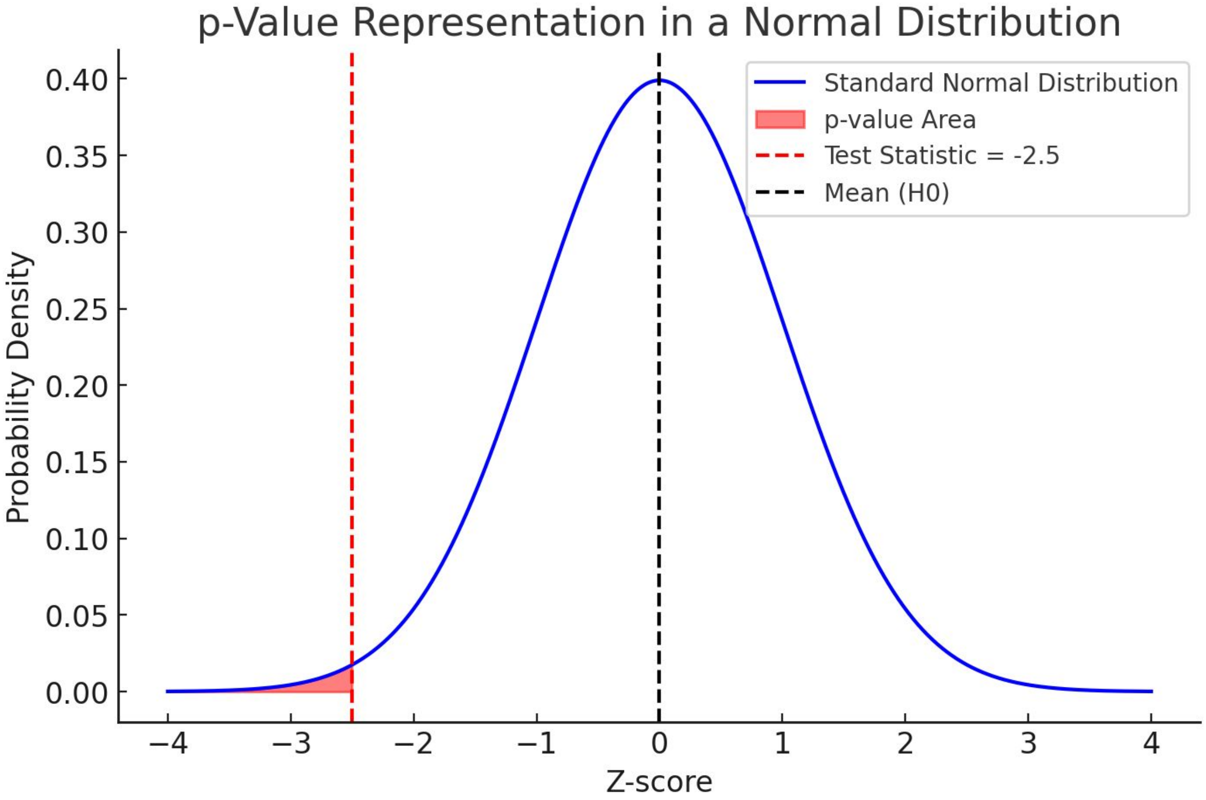
Measure unlikelihood

Because statistics follow a normal distribution, can measure *how unlikely* and observed statistic of a sample is

Statistical tests

Compares how *extreme* or unusual the statistic is, compared to the normal distribution of values

p-value



p-Value Representation in a Normal Distribution
Standard Normal Distribution
p-value Area
Test Statistic = -2.5
Mean (H0)
Probability Density
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−4
−3
−2
−1
0
1
2
3
4
Z-score

Hypothesis testing

Hypothesis testing

1. Formulating a *null hypothesis* (H0) and an *alternative hypothesis* (H1).
2. Choosing an appropriate statistical test.
3. Setting a significance level, *alpha*.
4. Calculating a test statistic and corresponding *p-value*.
5. *Reject* or *fail to reject* H0.

Formulate a hypothesis

H0: The coin is fair

H1: The coin is weighted

Formulate a hypothesis

H0: The two samples are drawn from the same underlying population

H1: The two samples are from fundamentally different populations

Formulate a hypothesis

H0: The two samples are generated by the same process

H1: The two samples are generated by fundamentally different processes

Formulate a hypothesis

H0: Speeds are no different after congestion pricing was implemented

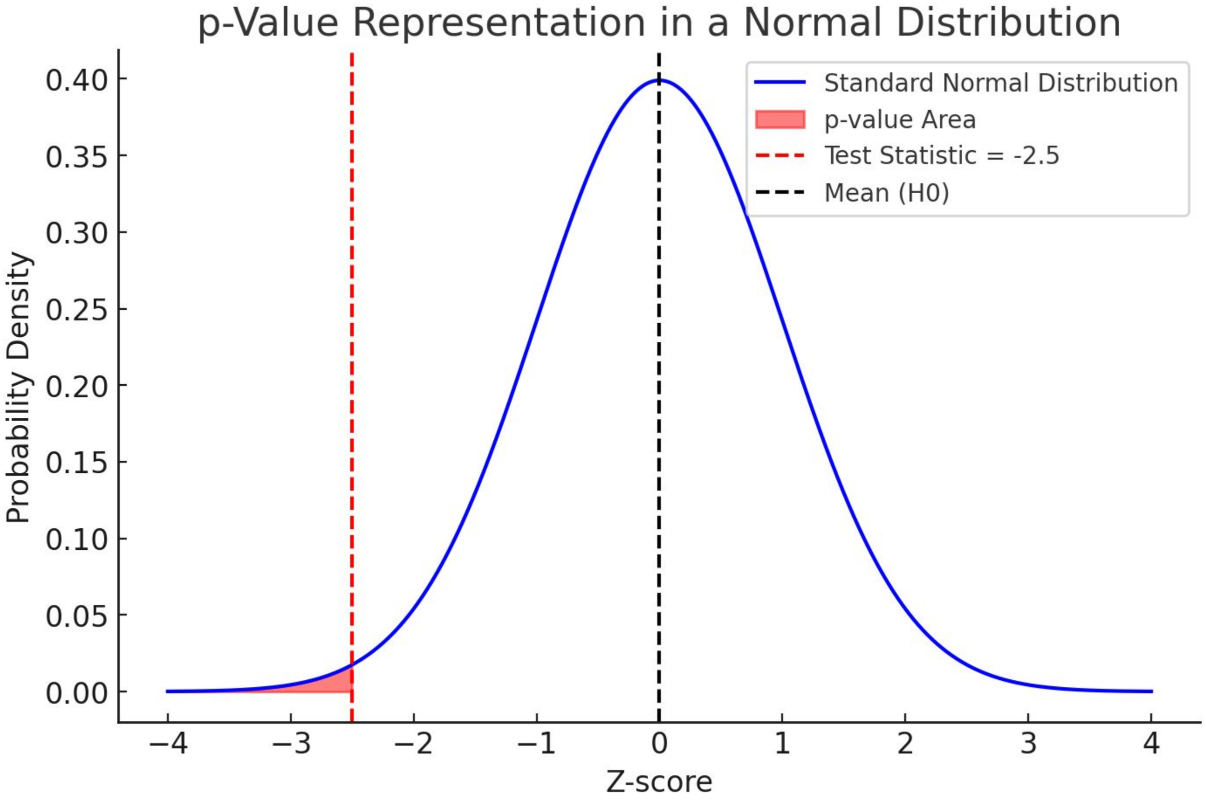
H1: Speeds are faster after congestion pricing was implemented

p-values

p-value : measures the probability of observing results *as extreme* as those in the sample, assuming the null hypothesis is true.

- A small *p*-value (< 0.05) suggests strong evidence *against* H0, leading to *reject* the null hypothesis
- A large *p*-value (> 0.05) suggests weak evidence against H0, meaning we *do not reject* it.

p-value



p-Value Representation in a Normal Distribution
Standard Normal Distribution
p-value Area
Test Statistic = -2.5
Mean (H0)
Probability Density
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−4
−3
−2
−1
0
1
2
3
4
Z-score

- The mean of the normal distribution represents the expected value under H0 (usually 0 difference).
- The spread (variance) of the distribution depends on sample size and population variability.
- If our test statistic falls in the extreme tails of the normal curve (low *p*-value), it suggests that the observed result is unlikely due to chance alone.

alpha

: The probability of incorrectly rejecting a true null hypothesis

- The significance level of a test
- Should be set in advance of conducting the test
- *alpha* levels are typically 0.05, 0.01, and 0.001 (corresponding to 95%, 99%, and 99.9% certainty.

“statistical significance”

: Reject the null hypothesis at greater than threshold level.

If *p*-value < *alpha*, then reject the null hypothesis.

If *p*-value >= *alpha*, then *cannot reject* the null hypothesis

Common misinterpretations of p-values

❌ A p-value of 0.05 means there's a 5% chance the null hypothesis is true.
✔ A p-value of 0.05 means that if H0 were true, we would observe a result as extreme as ours (or more extreme) only 5% of the time.

❌ A high p-value proves the null hypothesis is true.
✔ A high p-value means we do not have enough evidence to reject H0, but it does not confirm H0H0 is true.

T tests

T-tests

: compare the *means* of two groups to determine if they are significantly different

T-tests

- One-sample T-test: Compares a sample mean to a known population mean.
- Independent (Two-sample) T-test: Compares means from two different groups.
- Paired T-test: Compares means from the same group at two different times.

T-tests

- *two-sided T*-tests check whether one sample mean is *different* (higher *or* lower) than the other
- *one-sided T*-tests check whether one sample is either *greater than* or *less than* the other

T statistic

: difference of sample mean and hypothesized population mean, or difference between to sample means, relative to variation of samples

- sign (positive/negative) indicates which mean is larger
- scale indicates size of difference.
- but practically, not that useful. *p*-value is what is interesting.

T-test example

Does a new bus lane shorten average commute times?

1. Hypotheses:

- Null hypothesis (H0): average commutes are no shorter
- Alternative hypothesis (H1): average commutes are shorter

T-test example

Does a new bus lane shorten average commute times?

2. Compute test statistic: *T* statistic of difference between means
&

3. Compute *p*-value of test statistic

T-test example

Does a new bus lane shorten average commute times?

4. Interpret *p*-value