UCLA QCBio Collaboratory

Introduction to Modern Statistics

Wenbin Guo
Bioinformatics, UCLA
wbguo@ucla.edu

2025 Winter

UCLA Bioinformatics

Notation of the slides

- `Code` or `Pseudo-Code` chunk starts with " ➢ ", e.g.
  ➢ `print("Hello world!")`

- Link is underlined

- Important terminology is in **bold** font

- Practice comes with

Agenda

- Day 1: Probability and Statistics basics
  - Uncertainty; Probability; Distribution
  - Descriptive statistics

- Day 2: Inference
  - Hypothesis testing and *p*-values
  - Permutation test and bootstrap
  - False discovery rate control

- Day 3: Modeling
  - Regression techniques
  - Model selection

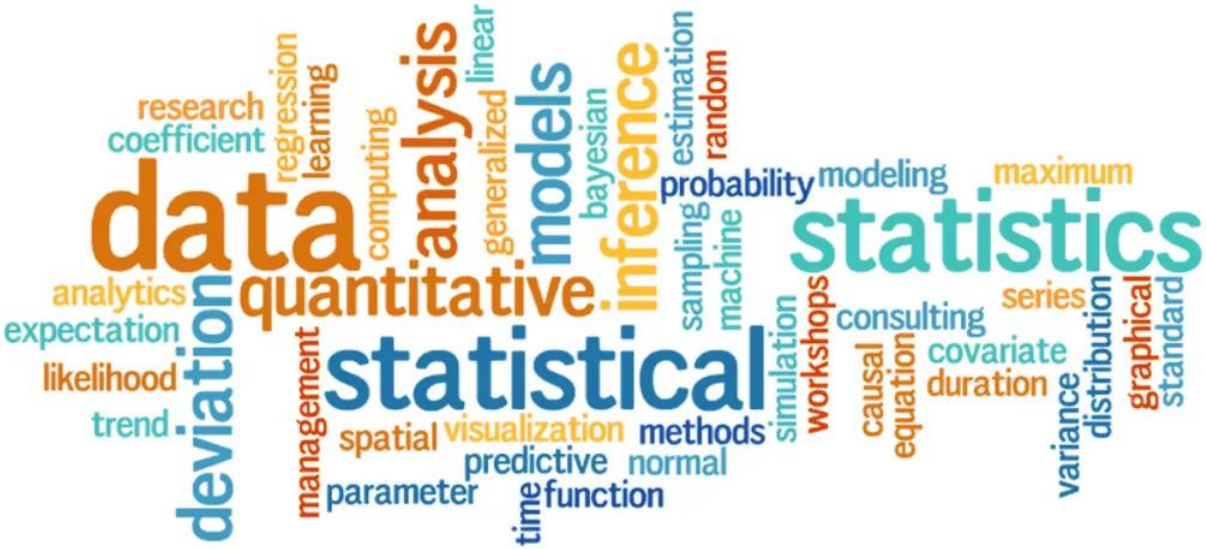


Image from Internet

UCLA QCBio Collaboratory

Day 2: Statistical inference

Wenbin Guo
Bioinformatics IDP, UCLA
wbguo@ucla.edu
2025 Winter

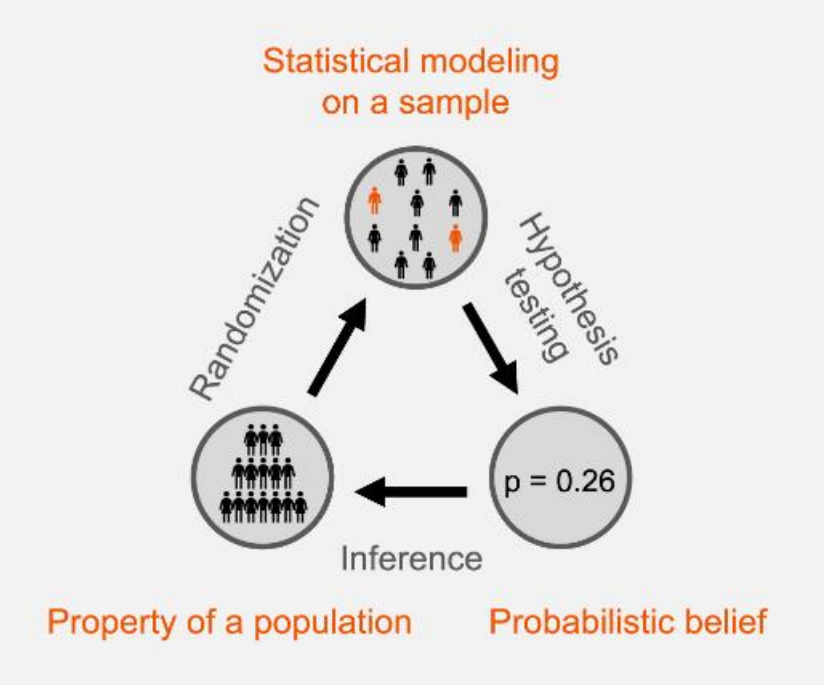
Overview

Time

- 2-hour workshop (45min + 45min + practice/Q&A)

Topics

- ❑ Inferential statistics basics
- ❑ Hypothesis testing
- ❑ Permutation test
- ❑ Bootstrap
- ❑ Multiple test correction



Summary – Day1

Introduction to statistics

- ❑Concept
- ❑History
- ❑Importance

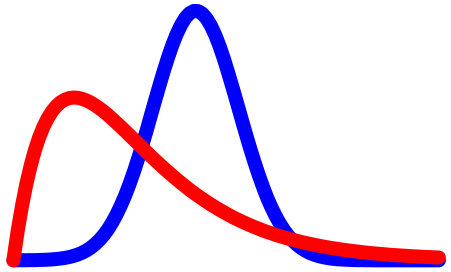
Uncertainty

- ❑Cause
- ❑Examples in real life
- ❑How do we react with it

Summary – Day1

Probability

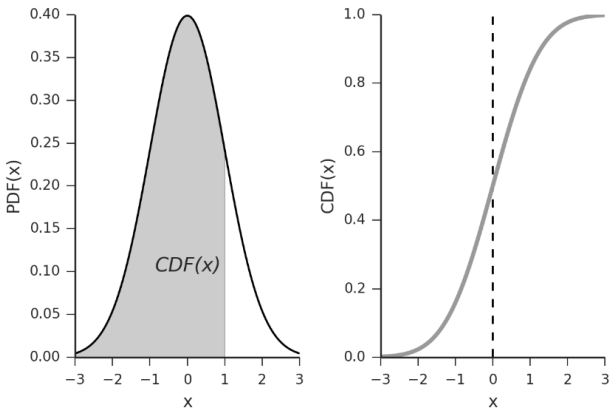
- ❑Events and sample space
- ❑Probability and its properties $P(\mathbf{A}) = \frac{|\mathbf{A}|}{|\mathbf{\Omega}|}$
- ❑Conditional probability $P(\mathbf{A} \mid \mathbf{B}) = \frac{P(\mathbf{AB})}{P(\mathbf{B})}$
- ❑Bayes Theorem

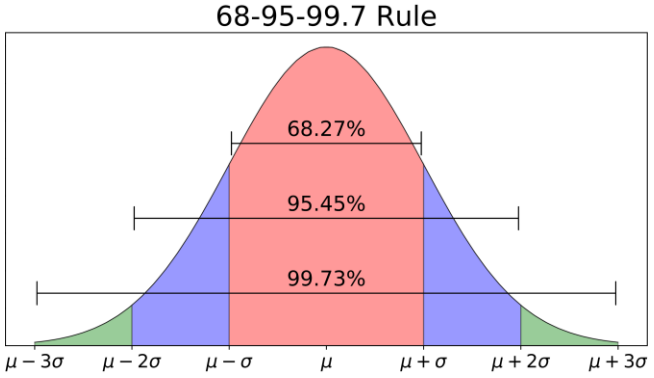


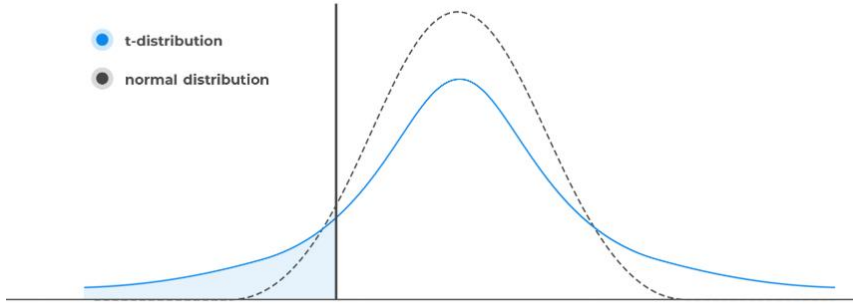
Summary – Day1

Distributions

Category	Name	Description
Discrete	Uniform	equal probability to every integer between a and b
	Bernoulli	single trial with two possible outcomes with success rate p
	Binomial	number of successes in n independent Bernoulli trials
	Negative Binomial	number of failures before achieving r successes
	Geometric	number of trials needed to get the first success in a series of Bernoulli trials
	Hypergeometric	the number of successes in a sample of size n drawn without replacement from a population of size N containing K successes
	Poisson	number of events occurring in a fixed interval of time or space
Continuous	Uniform	equally probability to every value in an interval $[a,b]$
	Normal	bell curve center around mean μ with standard deviation σ
	t	Similar to Normal but with heavier tails
	Chi-square	models the sum of the squares of k independent standard normal variables
	F	compare variances between two groups
	Exponential	models the time between events in a Poisson process
	Beta	model proportions and probabilities



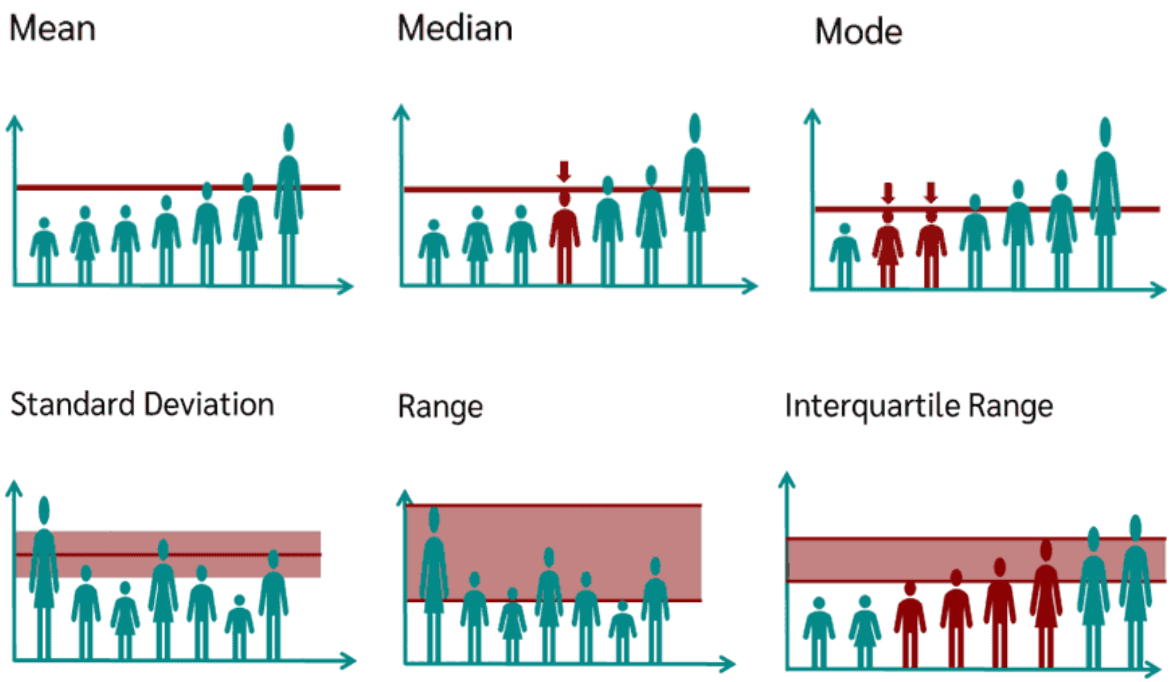


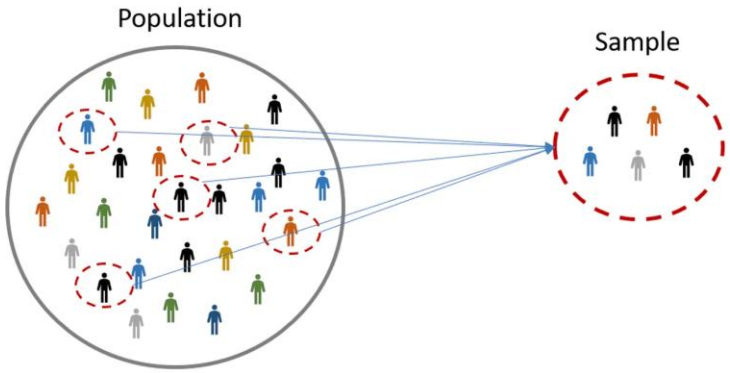


Summary – Day1

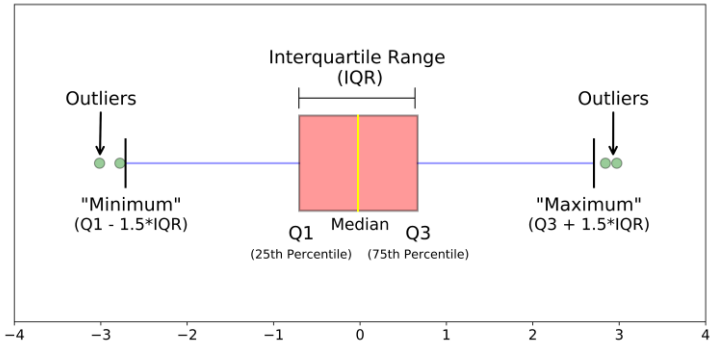
Descriptive statistics

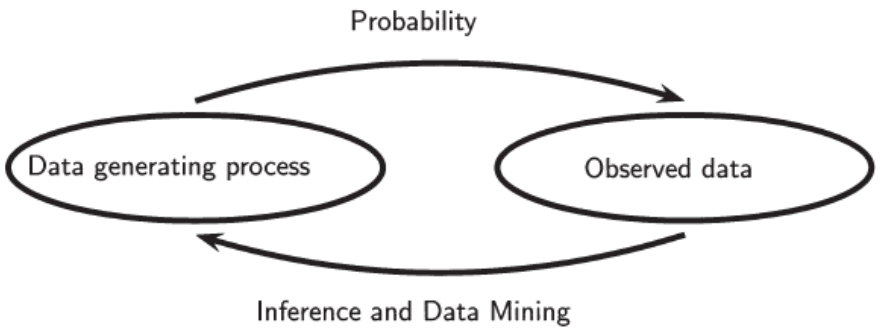
- ❑Tendency
- ❑Dispersion
- ❑Skewness
- ❑Distribution





Normal distribution
(unimodal, symmetric,
the “bell curve”)



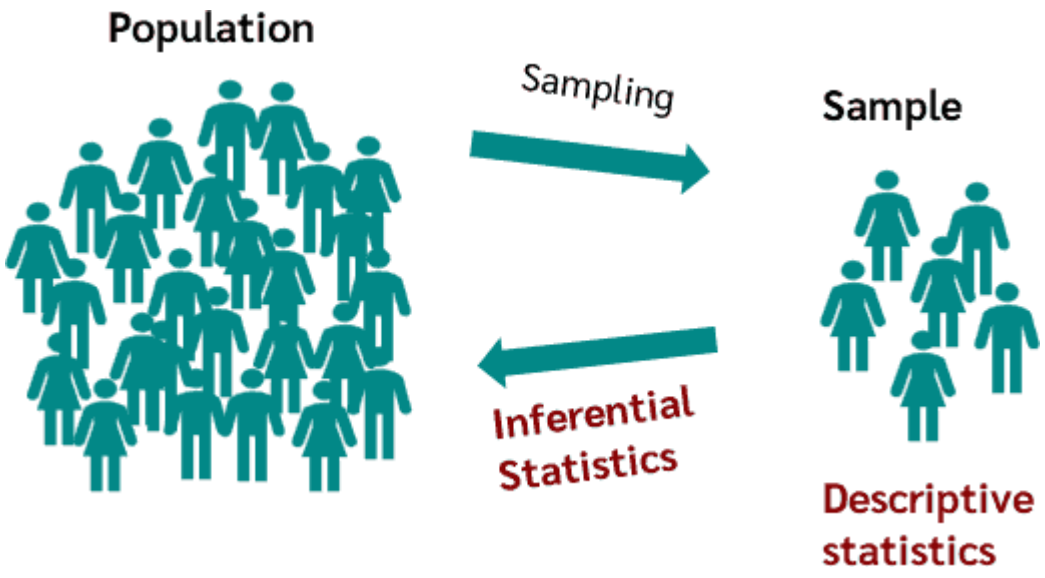


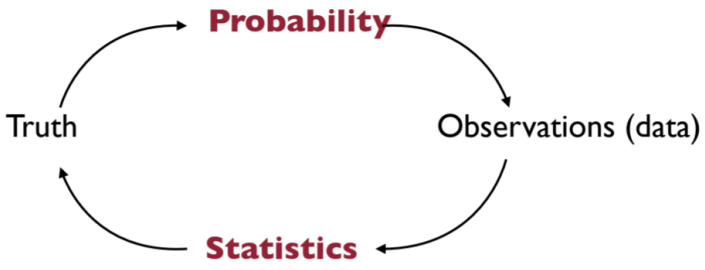
Inferential statistics

Use the part to learn about the whole

Inferential statistics

draw conclusions about the population based on sample data





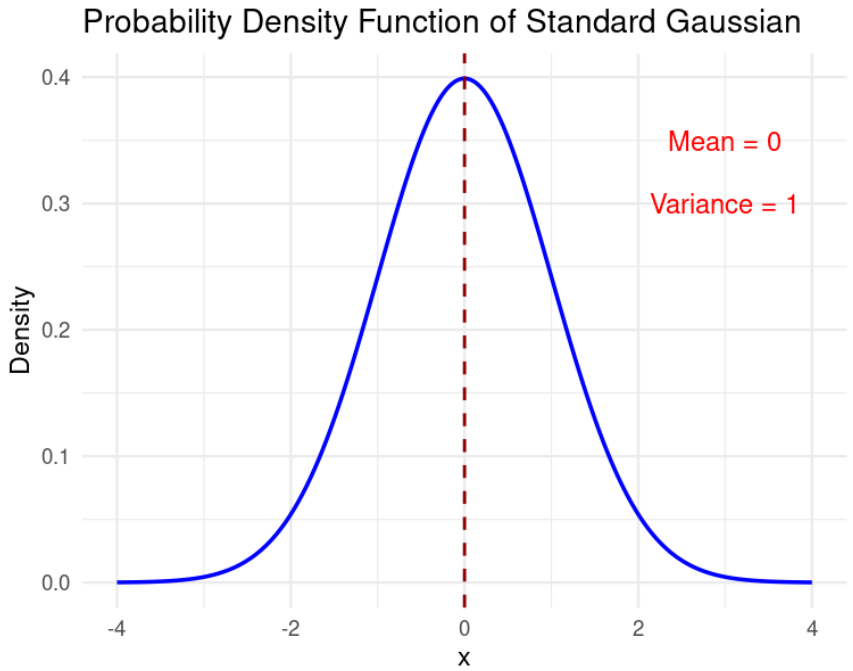
Typical tasks

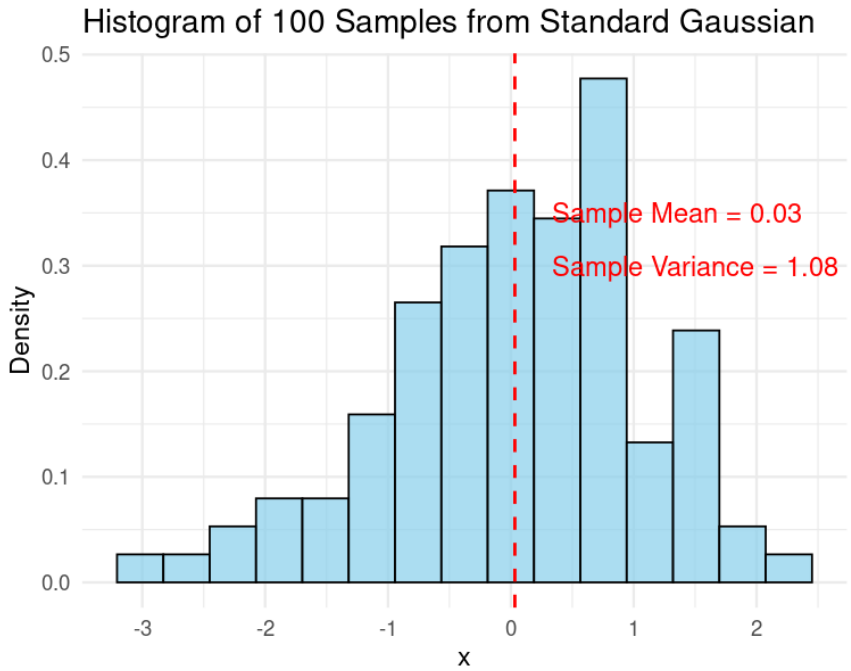
- ❑Parameter estimation
- ❑Hypothesis testing

Given a sample $x_1,\ldots,x_n \sim F$, how do we infer F?

Population vs samples

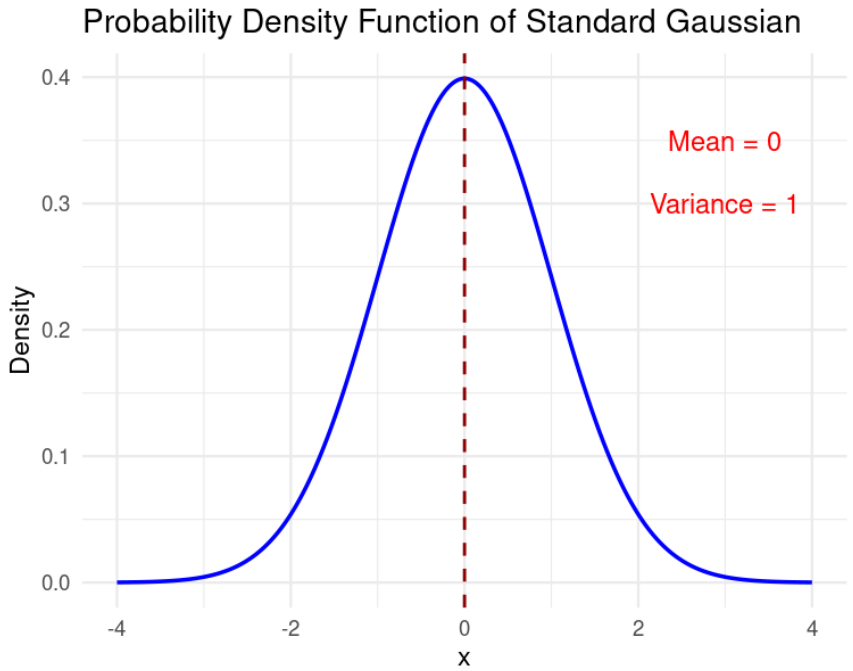
Use standard gaussian as an example

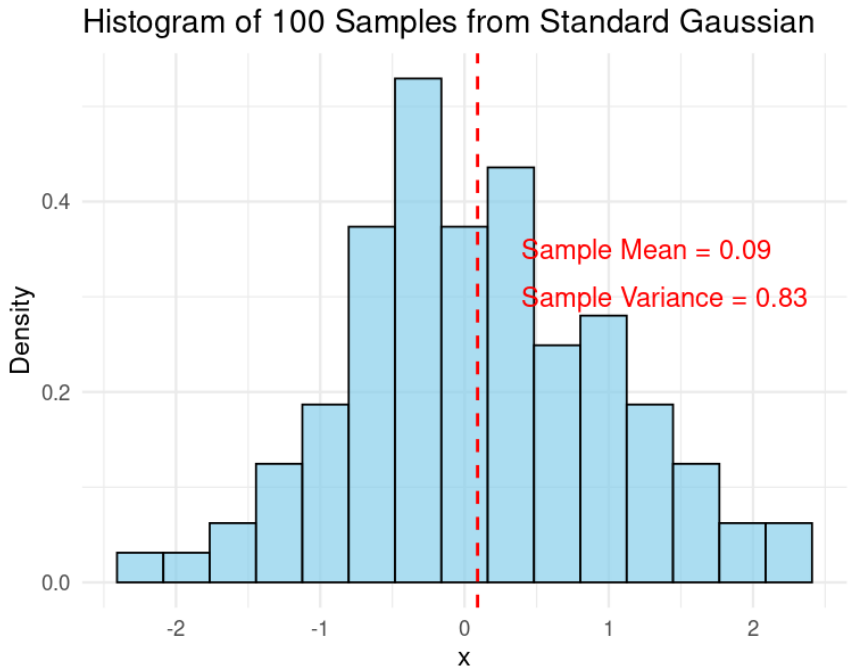




Population vs samples

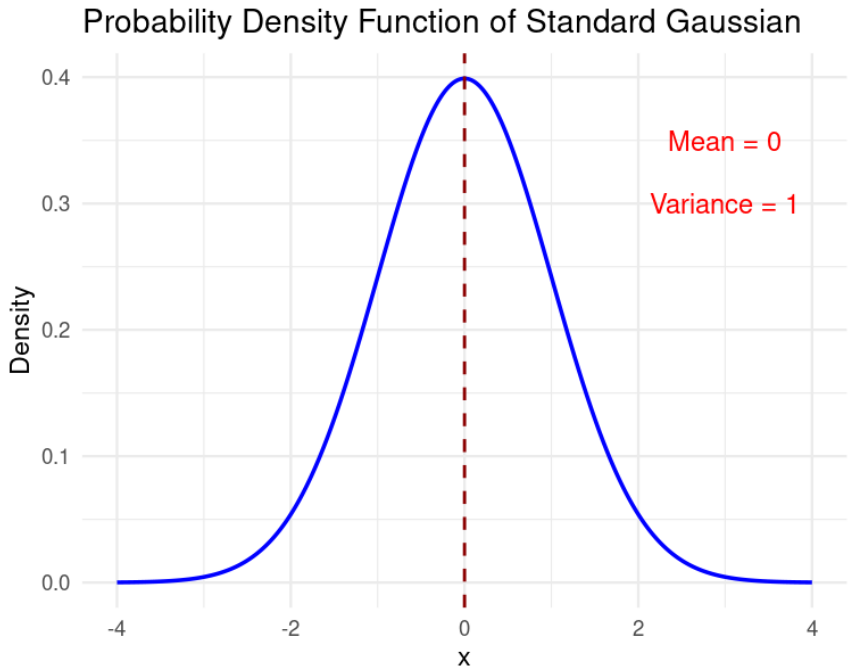
Use standard gaussian as an example

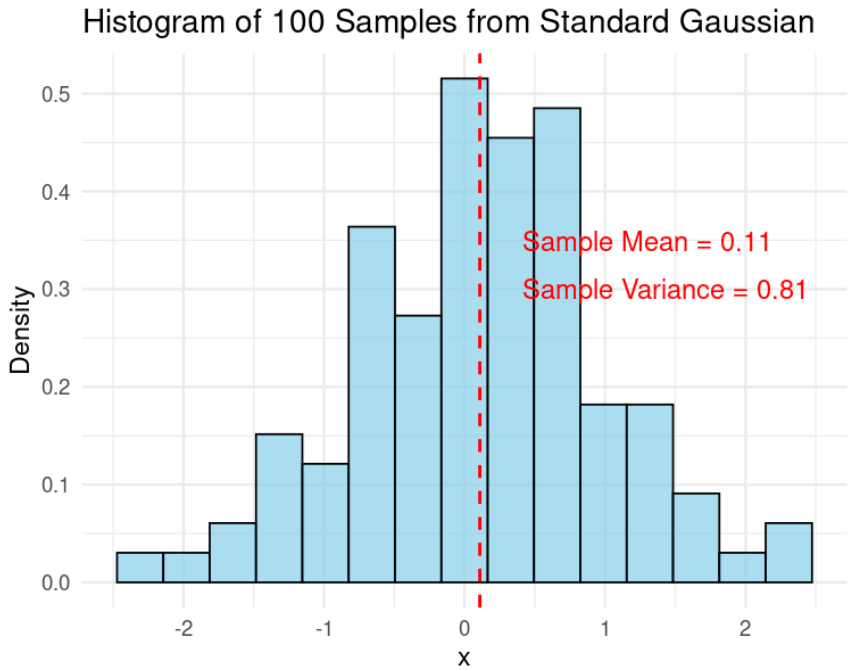




Population vs samples

Use standard gaussian as an example





Sampling is **random**, so are the summary statistics (sample mean/sample variance)

Expectation and Variance

Population

Mean:

$$E(X) = \sum x_i \; P(x_i) = \int_{-\infty}^{+\infty} x \, f(x) dx$$

Variance:

$$Var(X) = E[\left(X - E(X)\right)^2]$$

Samples

Mean:

$$\frac{\sum_{i=1}^{n} x_i}{n}$$

Variance:

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

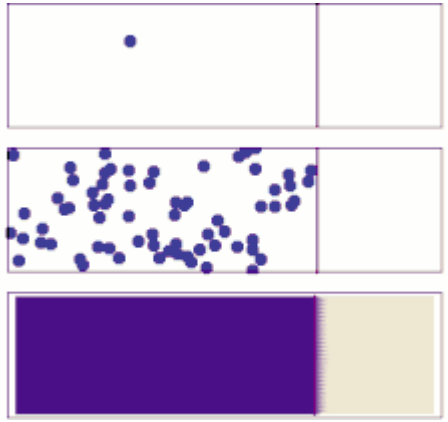
Parameter: a property of the distribution

Statistic: a summary calculated from samples

Law of large numbers (LLN)

5.6 Theorem (The Weak Law of Large Numbers (WLLN)).
If $X_1, \ldots, X_n$ *are* IID, *then* $\overline{X}_n \xrightarrow{\mathrm{P}} \mu$.

$$\mathbb{P}\left(|\overline{X}_n - \mu| > \epsilon\right) \leq \frac{\mathbb{V}(\overline{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$



Diffusion is an example of the law of large numbers.

In summary:

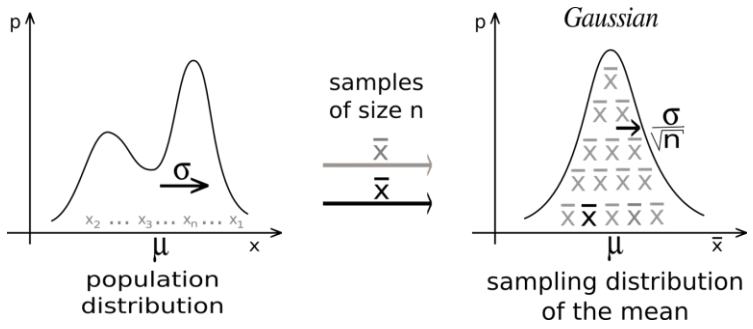
- Sample mean converges to population mean as sample size increases
- LLN provides Implications for estimation and reliability

Central limit theorem (CLT)

5.8 Theorem (The Central Limit Theorem (CLT)). *Let $X_1, \ldots, X_n$ be* IID *with mean μ and variance σ^2. Let $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$. Then*

$$Z_n \equiv \frac{\overline{X}_n - \mu}{\sqrt{\mathbb{V}(\overline{X}_n)}} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \rightsquigarrow Z$$

where $Z \sim N(0,1)$. In other words,



In summary:

- Distribution of sample means approaches normal distribution as sample size grows.
- CLT allows for hypothesis testing and constructing confidence interval using normal approximation.

Let’s do some practice!

- git clone https://github.com/wbvguo/qcbio-Intro2ModernStats.git

Hypothesis testing

“To reject or not to reject, that is the question.”

Wenbin Shakespeare

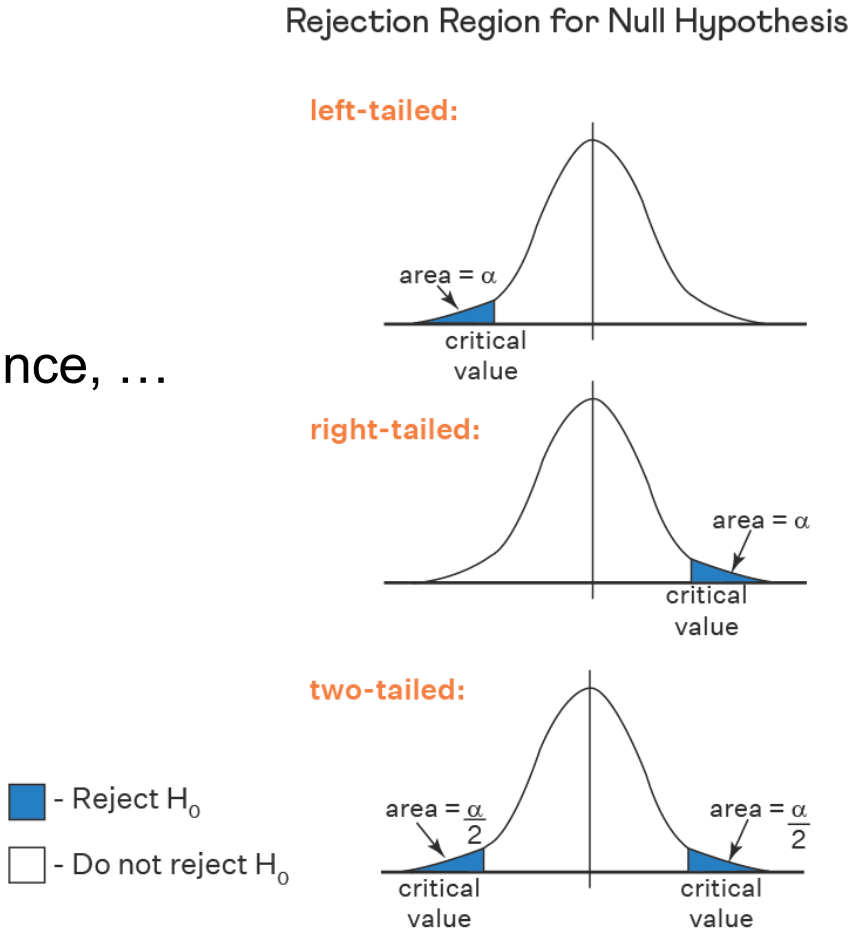
A famous nobody

Hypothesis testing

A process to determine if there is enough evidence to reject a null hypothesis

Procedure:

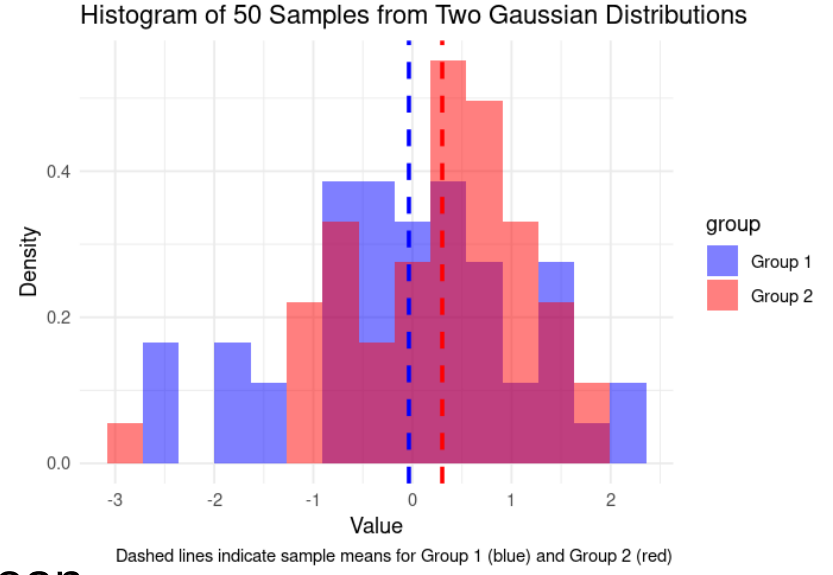
- ❑ Frame the hypothesis
  - Null hypothesis (H_0): no difference/effect, …
  - Alternative hypothesis (H_a or H_1): there is an effect/difference, …
- ❑ Choose an appropriate Test Statistics
- ❑ Calculate the statistical significance (*p*-value)
- ❑ Compare to a significance level (α)
- ❑ Make decision: reject/not reject

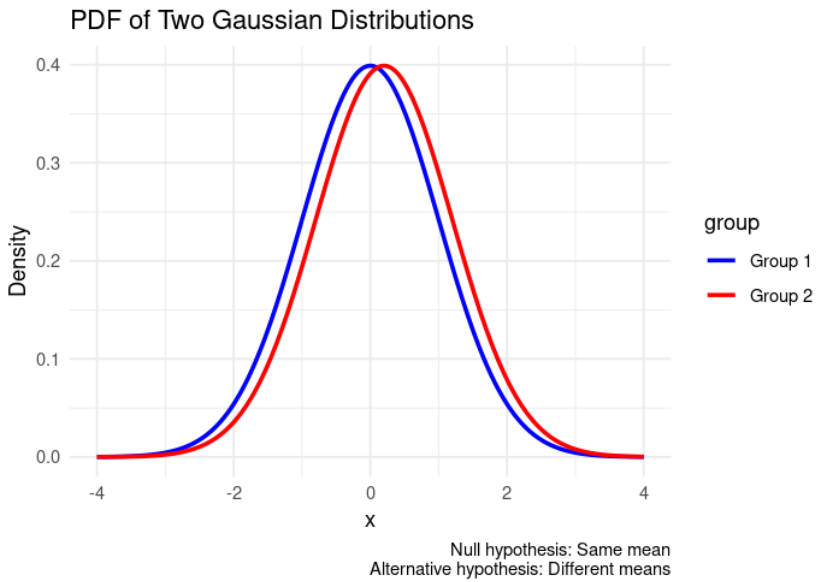


Let’s see an example

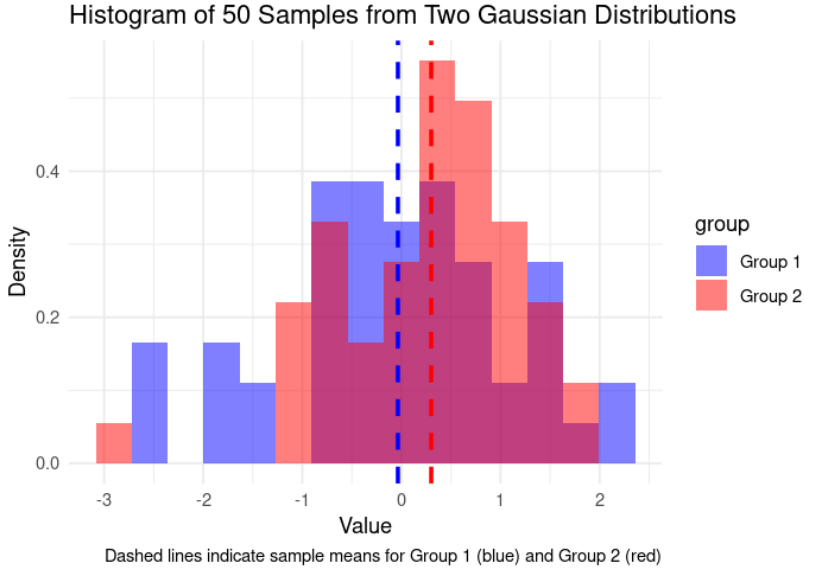
Use gaussian distributions as an example

- ❑Frame the hypothesis
  - Null hypothesis (H_0): two group have the same mean
  - Alternative hypothesis (H_a): two group have the different mean





Let’s see an example



Use gaussian distributions as an example

- ❑Frame the hypothesis
  - Null hypothesis (H_0): two group have the same mean
  - Alternative hypothesis (H_a): two group have the different mean
- ❑Choose an appropriate Test Statistics: t-test
- ❑Calculate the statistical significance (*p*-value)
- ❑Compare to a significance level (α)
- ❑Make decision: reject/not reject

```{r}
t.test(samples_group1, samples_group2)
```

```
        Welch Two Sample t-test

data:  samples_group1 and samples_group2
t = -1.6104, df = 93.647, p-value = 0.1107
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.7511248  0.0783784
sample estimates:
  mean of x   mean of y
-0.03567178  0.30070141
```

Rethink about hypothesis testing

Reject: Why our goal is to reject? not to prove?

- Proving something to be true is difficult, proving something to be false is usually easier, as long as you can find a counter-example.

Enough evidence:

- **Presumption of innocence**: any defendant in a criminal trial is assumed to be innocent until they have been proven guilty

In other words, hypothesis testing is conservative, we turn to not reject unless we have enough evidence

Better Call Saul

Fail to reject == the hypothesis is true?

No, this is not a binary world

Example: Clinical Trail for new drug

A clinical trial tests whether a new drug reduces blood pressure. The *p*-value for the difference between the drug and placebo group is **0.16**—greater than the conventional **0.05** threshold.

- The **absence of evidence** (failing to reject the null hypothesis) **does not prove** that the drug has no effect.
- The trial might have been underpowered (small sample size), lacking the data needed to detect a true effect.

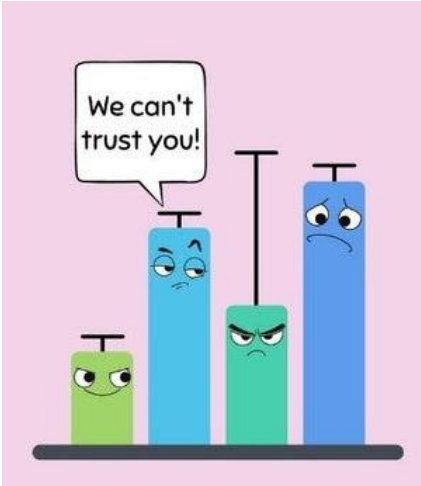
In statistics, **failing to find evidence** does not imply **evidence of no effect**—it may simply reflect study limitations.

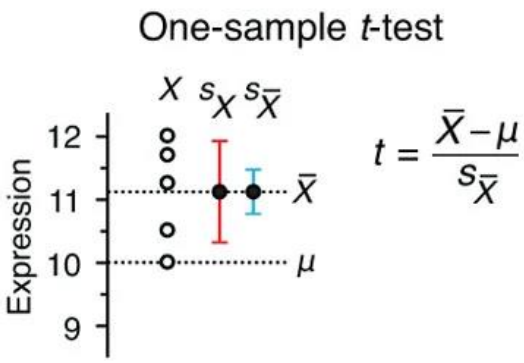
"*Absence of evidence is not evidence of absence*"

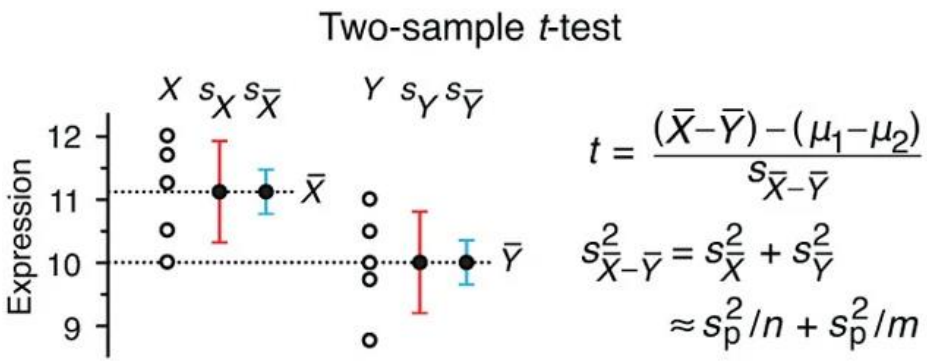
Carl Edward Sagan
(1934-1996)

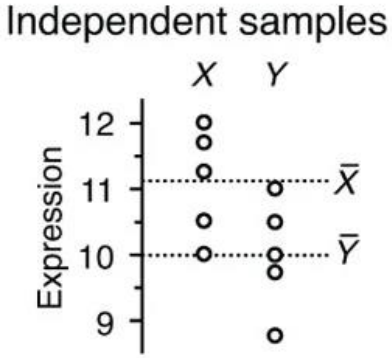
Statistical tests

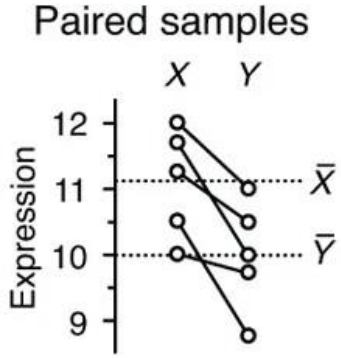
- Test mean:
  - ❑One group
  - ❑Two group
  - ❑Paired
  - ❑Multiple group





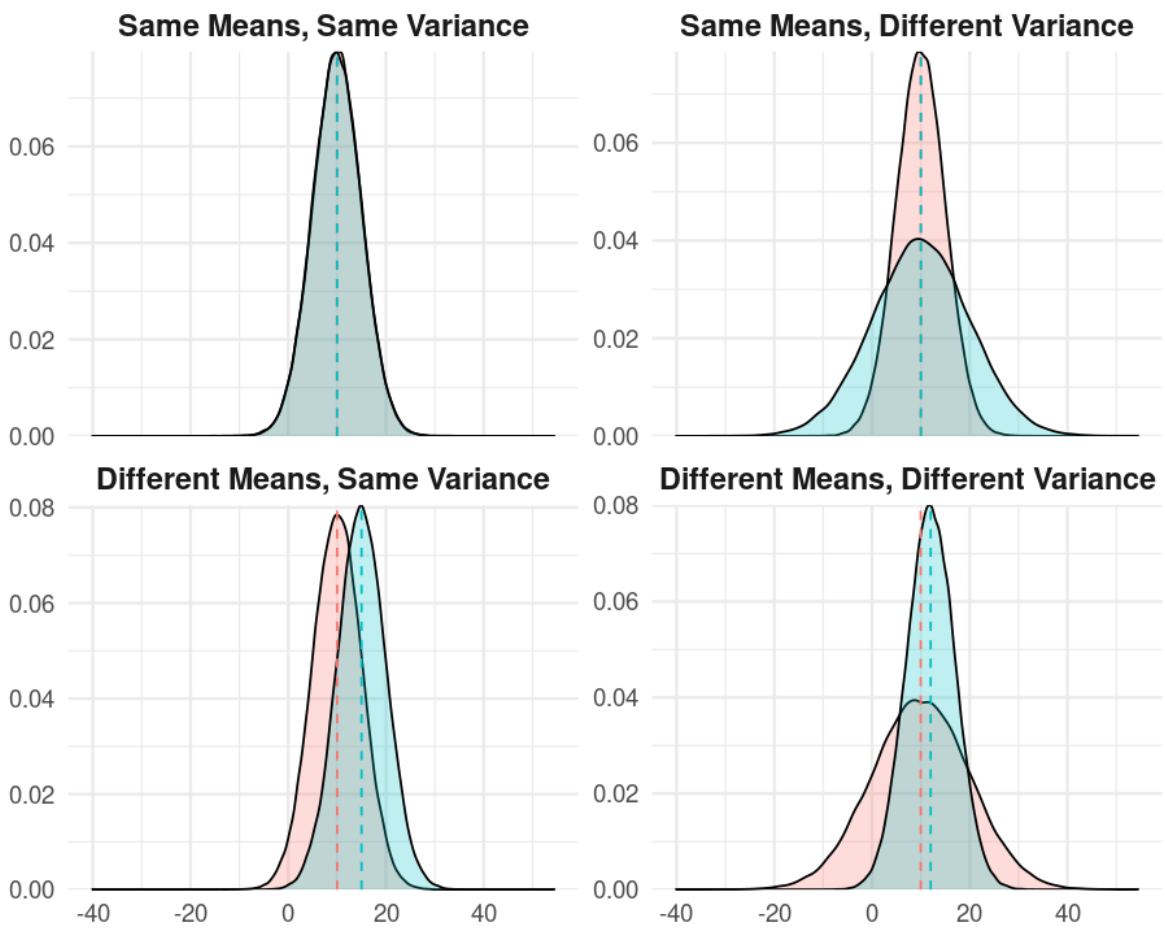






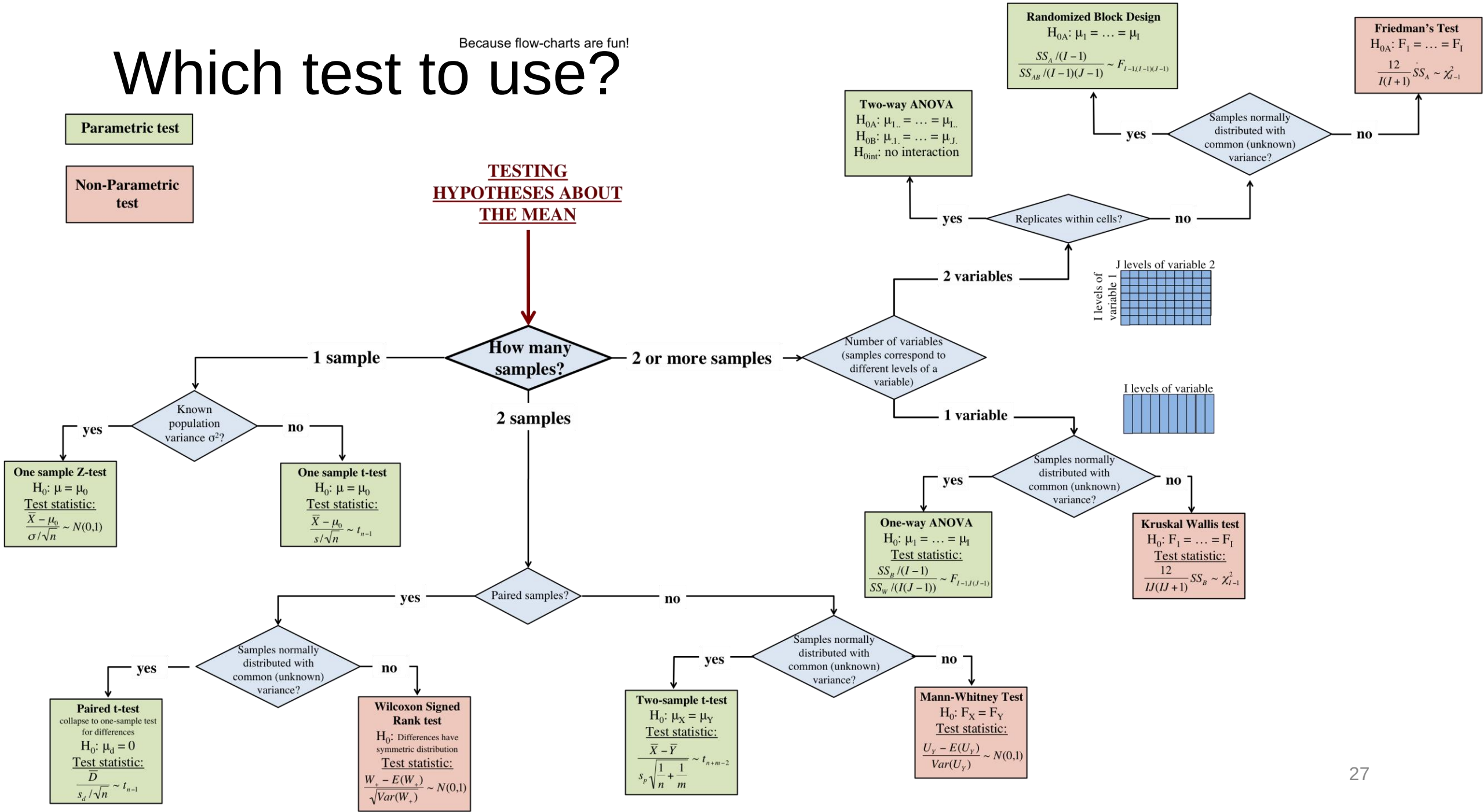
Statistical tests

- Test mean:
  - ❑One group
  - ❑Two group
  - ❑Paired
  - ❑Multiple group
- Test variance
- Test distribution



Which test to use?

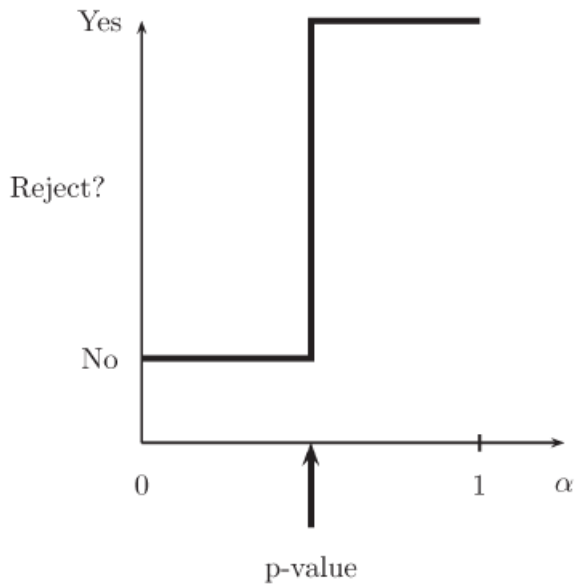
Because flow-charts are fun!



p-value

The p-value is the probability (under H_0) of observing a value of the test statistic the same as or more extreme than what was actually observed.

- a measure of the evidence against H_0
- the smaller the p-value, the stronger the evidence against H_0.



Typically,

p-value	evidence
$< .01$	very strong evidence against H_0
$.01 - .05$	strong evidence against H_0
$.05 - .10$	weak evidence against H_0
$> .1$	little or no evidence against H_0

True or false?

- large p-value is strong evidence in favor of H_0.
  - False
- The p-value is the probability that the null hypothesis is true
  - False
- "p = 0.05 means there's a 95% chance H_0 is true"
  - False

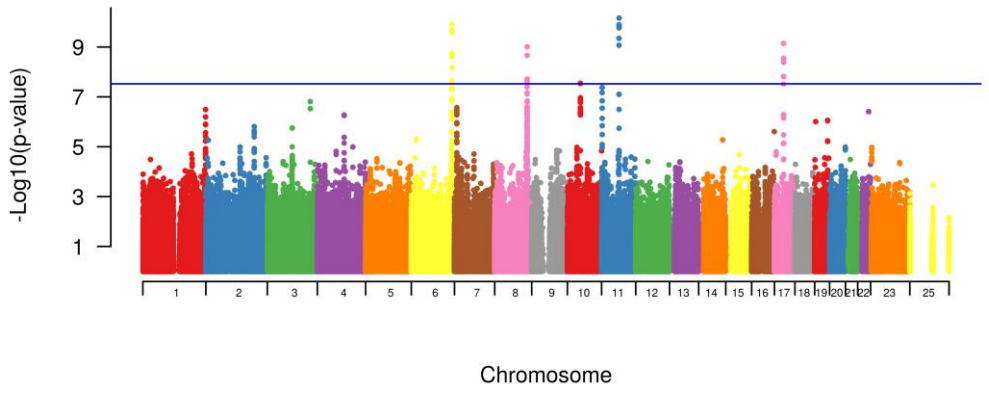
p-value distribution when H_0 is true

$$
\begin{aligned}
Pr(P < p) &= Pr\left(F^{-1}(P) < F^{-1}(p)\right) \\
&= Pr(T < t) \\
&\equiv p;
\end{aligned}
$$

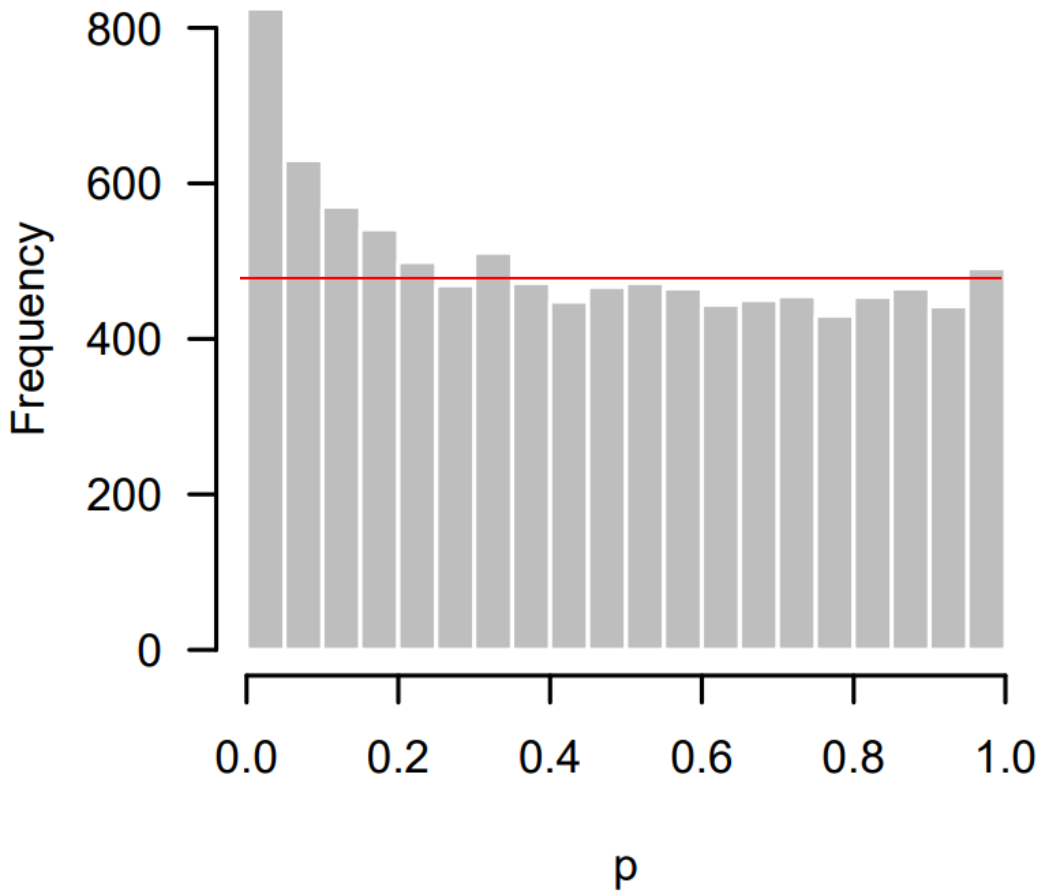
p-value follows **uniform** distribution under the Null

Can use this rationale to diagnostic

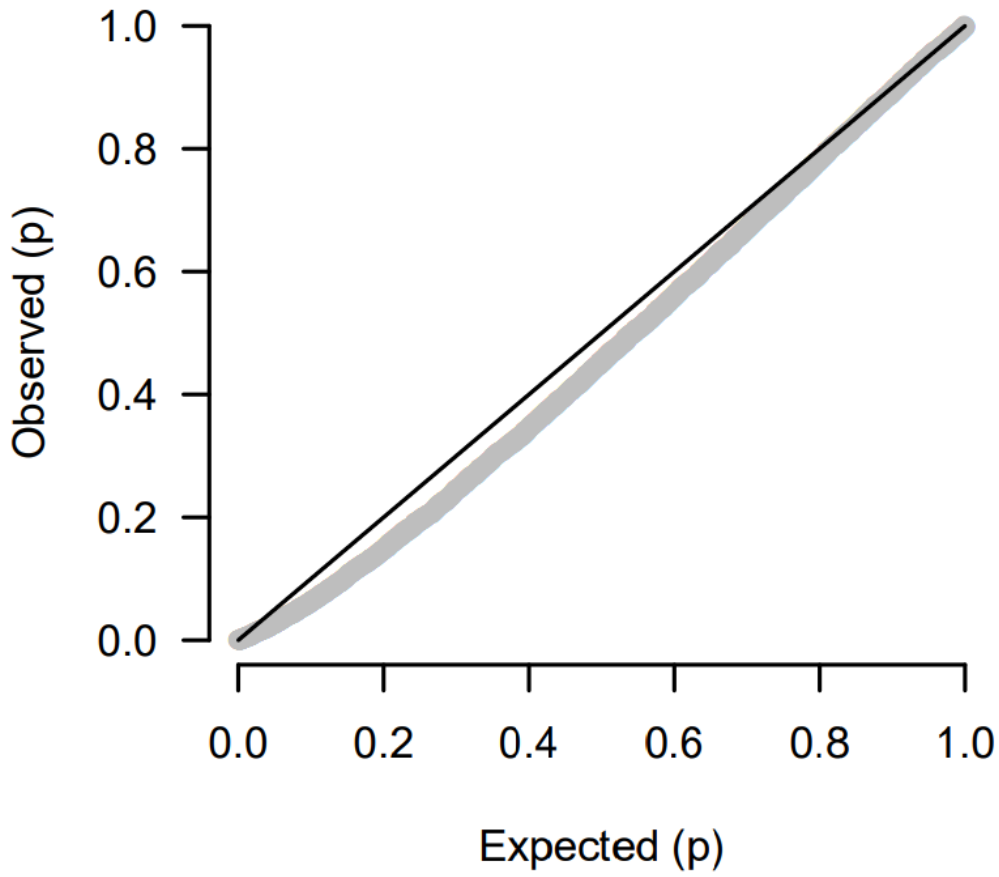
- Histogram of p-values
- Quantile-Quantile plot (QQ plot)



Examine *p*-value distribution



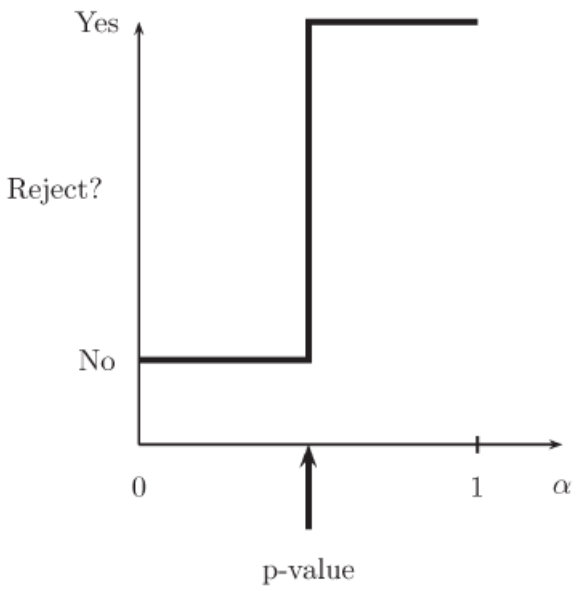
Histogram

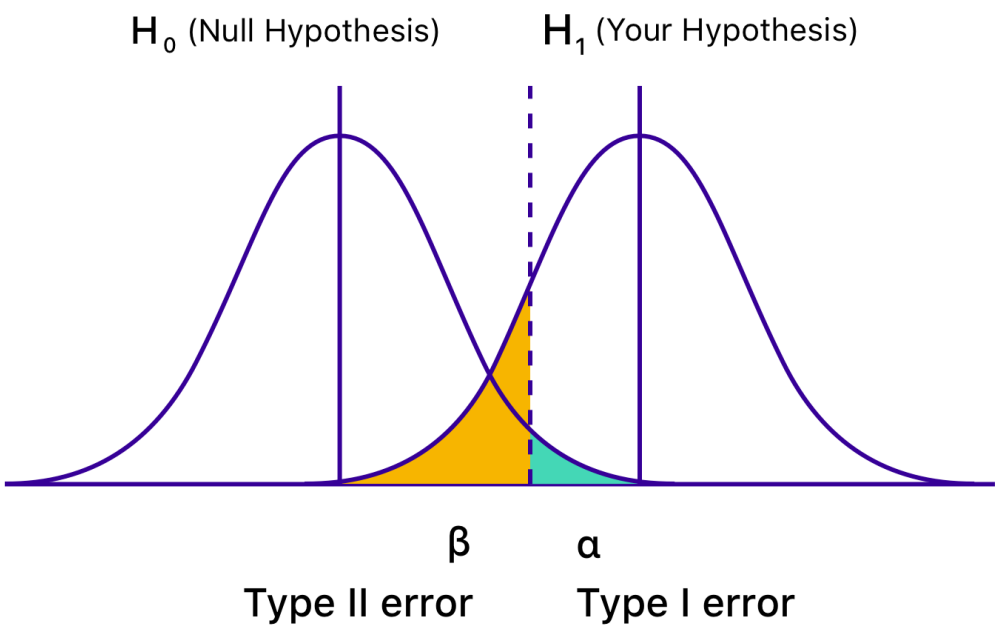


Quantile-Quantile plot

Decision errors

We use statistics to make decisions





Decision errors

But the decisions can be wrong…

- ❑Reject the null when it's true
- ❑Fail to reject the null when it's not true

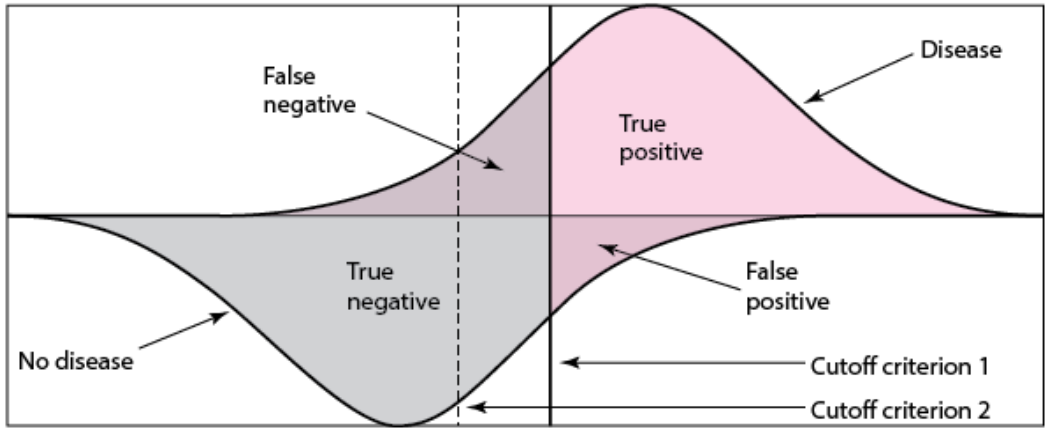
H0: Null Hypothesis
(Neutral / Status-quo)

Conclusions from observed data ↕	Reality → (H0 is True)	Reality → (H0 is False)
Reject H0	**Type-I error (Alpha, α)** - Reject a True Null Hypothesis. → **False-positive** conclusion	Correct Conclusion ☺
Do not Reject H0	Correct Conclusion ☺	**Type-II error (Beta, β)** - Do not reject a False Null Hypothesis → **False-Negative** conclusion

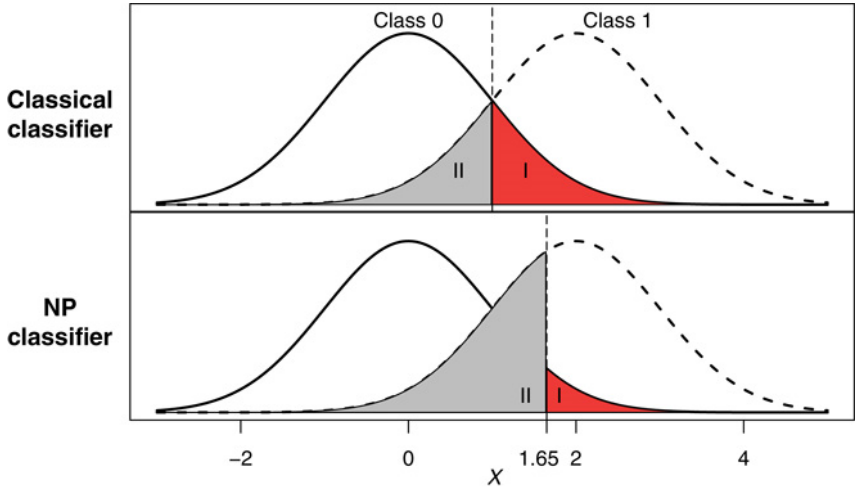
Type I Error

Type II Error

Decision errors and asymmetry classification



Put one type of error under control while minimizing the other one

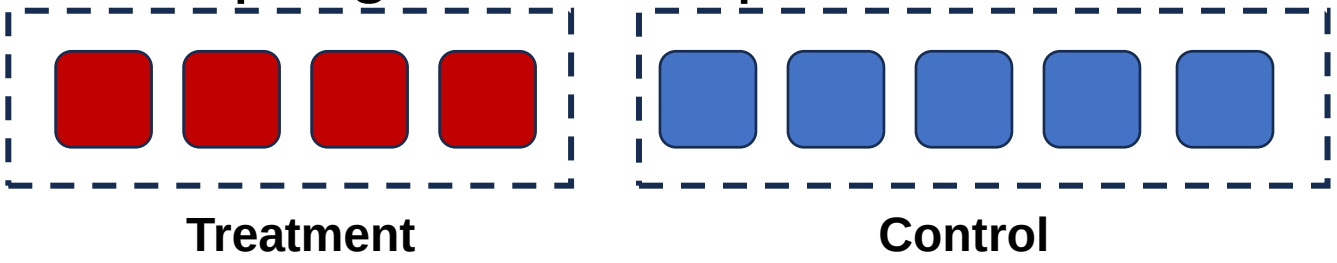


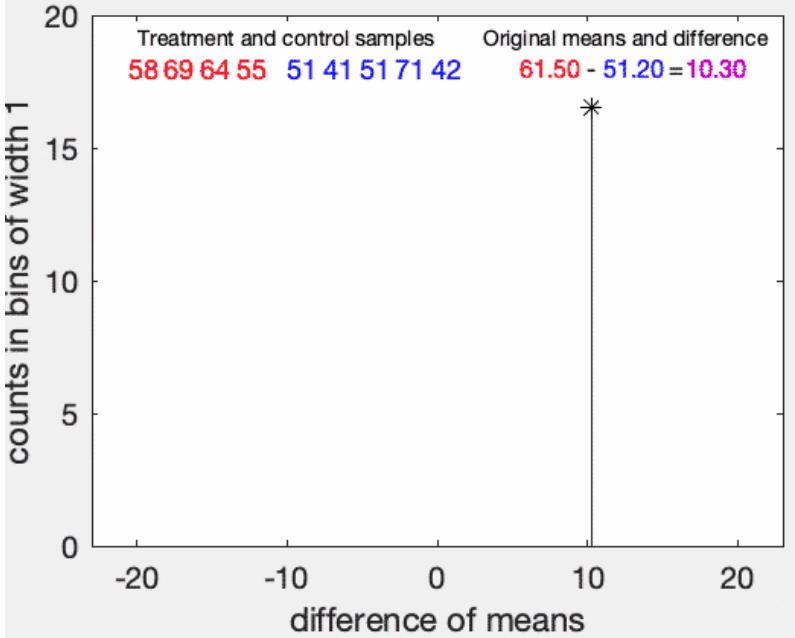
Permutation and Bootstrap

Computer Age Statistical Inference

Permutation test

Resampling without replacement





Procedure

- ❑ Define the hypothesis
- ❑ Define the test statistics
- ❑ Randomly permute the data
- ❑ Calculate the test statistics for each permutation
- ❑ Compare the observed test statistics with the permuted test statistic distribution

$$p = \frac{\sum_{i=1}^{B} I(\Delta\theta^{(i)} > t)}{B}$$

Permutation test

Take `sex_discrimination` data as an example

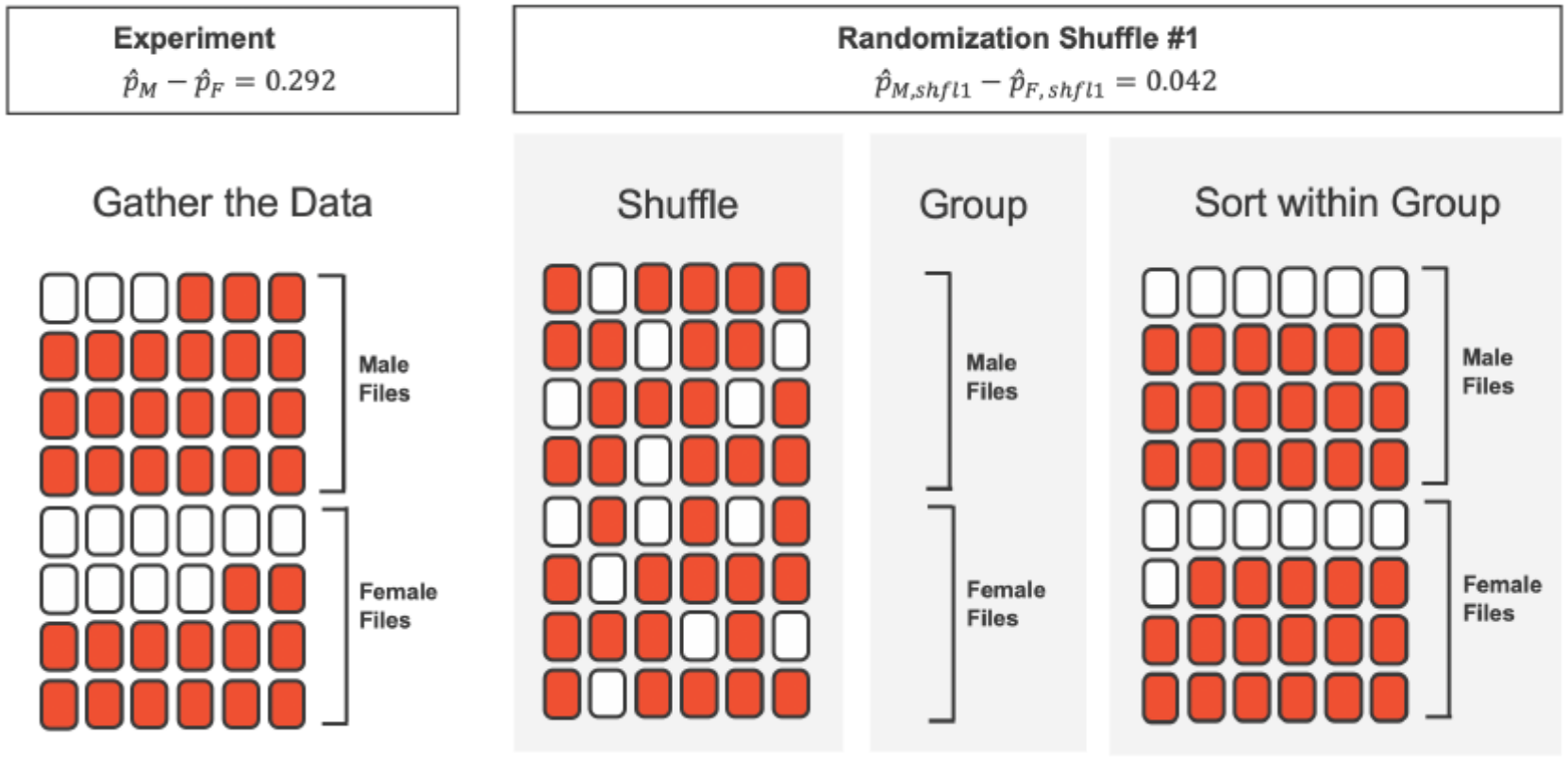
Table 11.1: Summary results for the sex discrimination study.

	decision		
sex	promoted	not promoted	Total
male	21	3	24
female	14	10	24
Total	35	13	48

We can clearly see that %promotion is higher in male population, but how to get the significance?

Permutation test

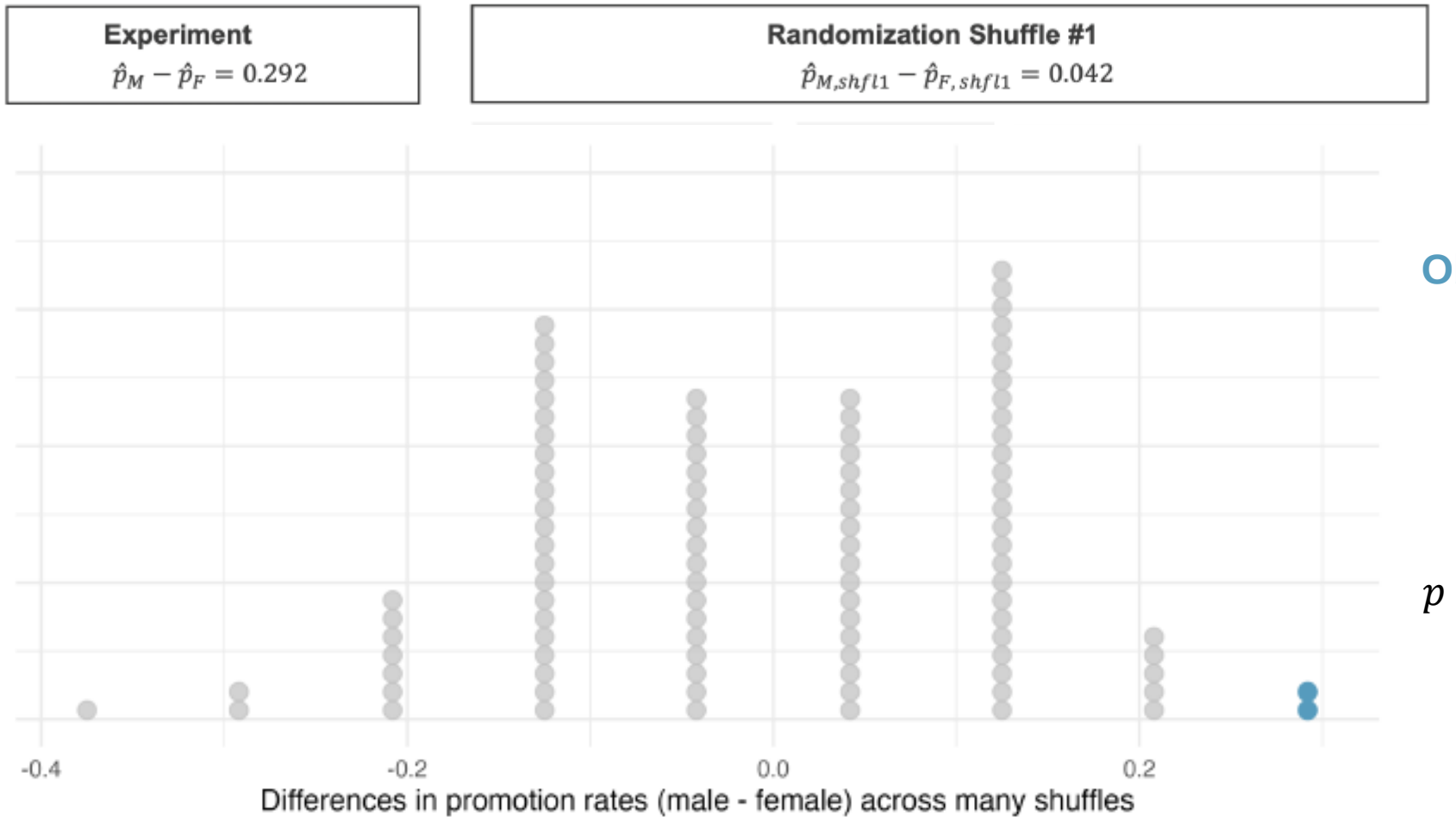
Take `sex_discrimination` data as an example



Repeat the shuffling many times

Permutation test

Take `sex_discrimination` data as an example



Observed statistic
vs.
null statistics

$$p = \frac{\sum_{i=1}^{B} I(\Delta\theta^{(i)} > t)}{B}$$

Permutation test

Pros:

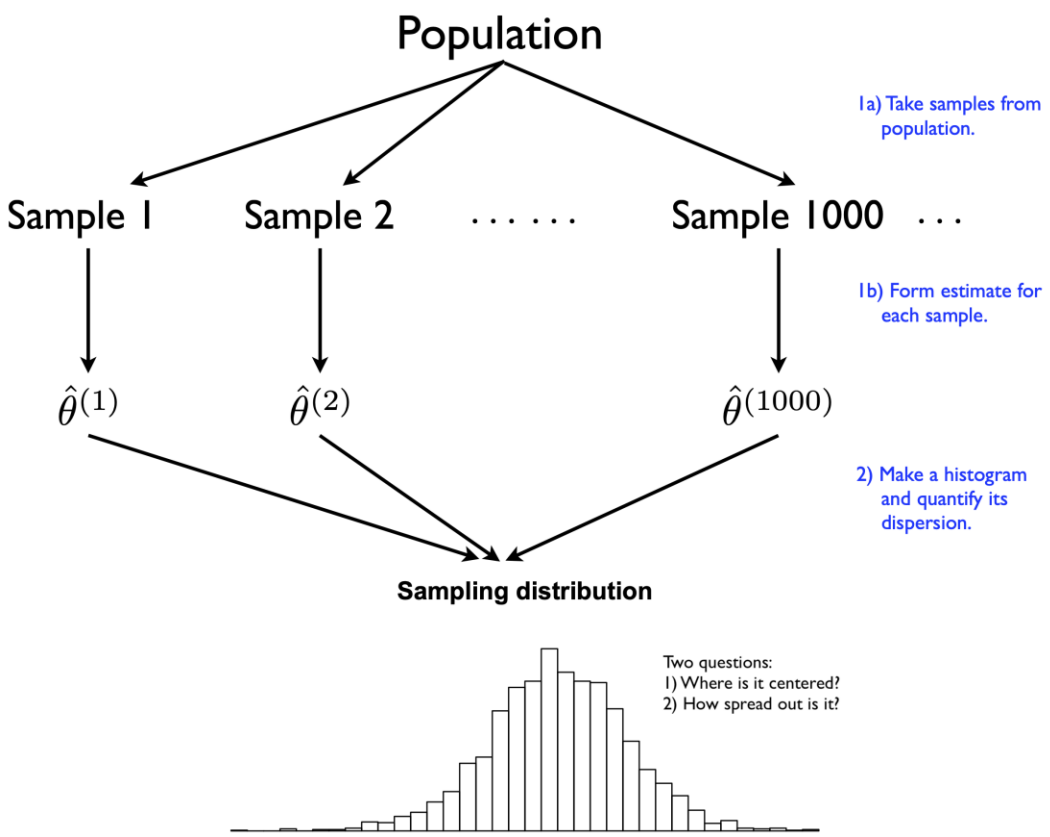
- **Distribution-free**: No assumptions about data distribution (non-parametric); Works well with skewed or non-normal data.
- **Flexible**: Applicable to a wide range of test statistics (mean, median, correlation, etc.).
- **Easy to implement**: Simple concept based on resampling.

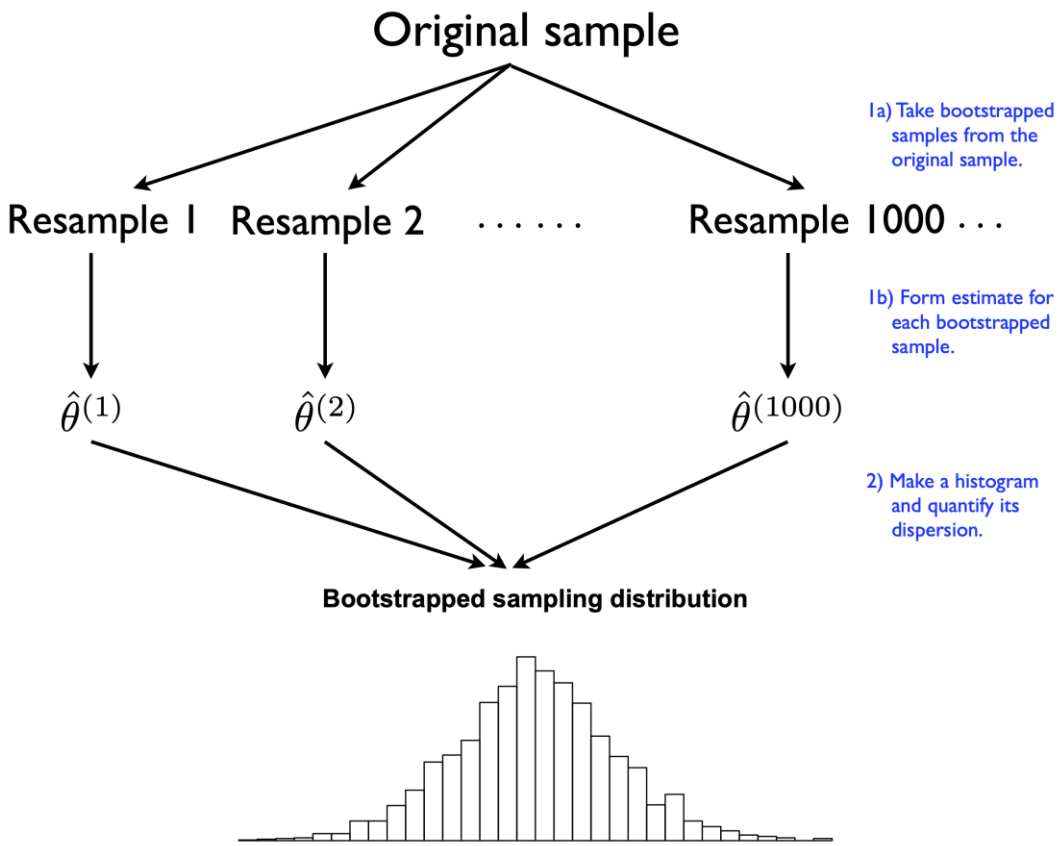
Cons:

- **Computationally intensive**: Requires a large number of resamples for accurate p-values, especially with large datasets.
- **Randomization required**: Assumes data can be randomly shuffled
- **Limited interpretability**: p-values are purely empirical, with no direct parameter estimation

Bootstrap

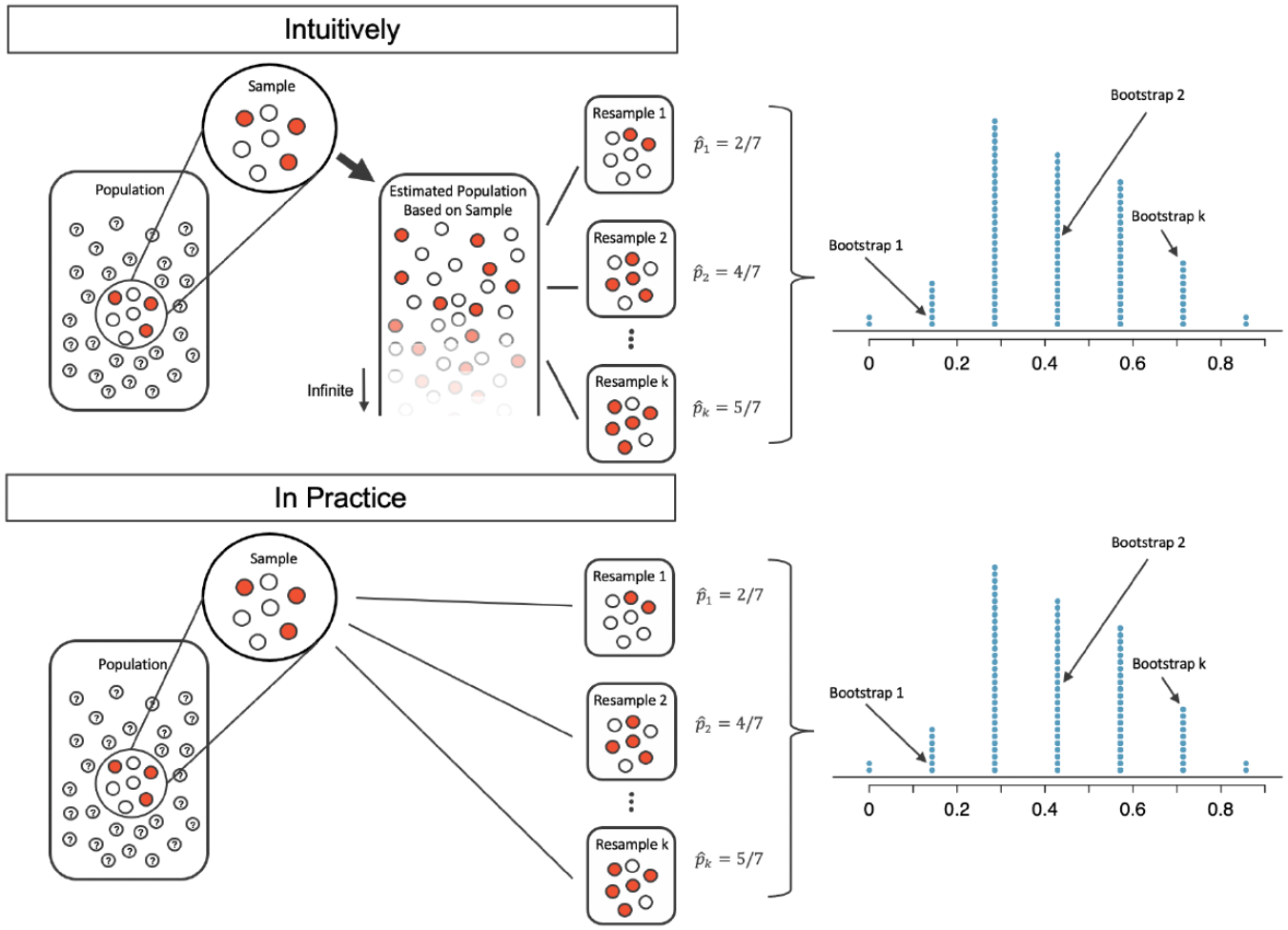
Resampling with replacement





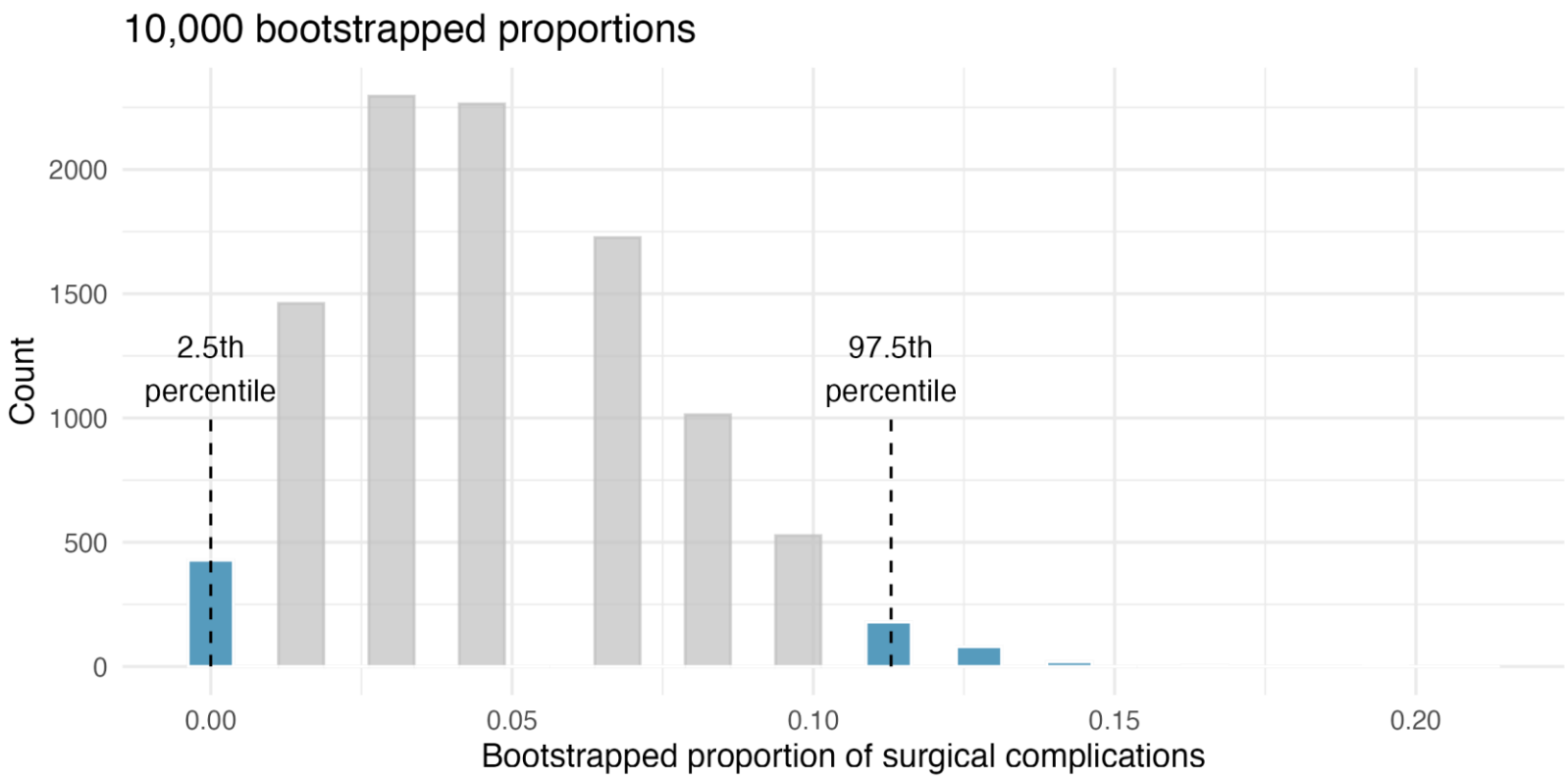
Bootstrap

Confidence interval with bootstrapping



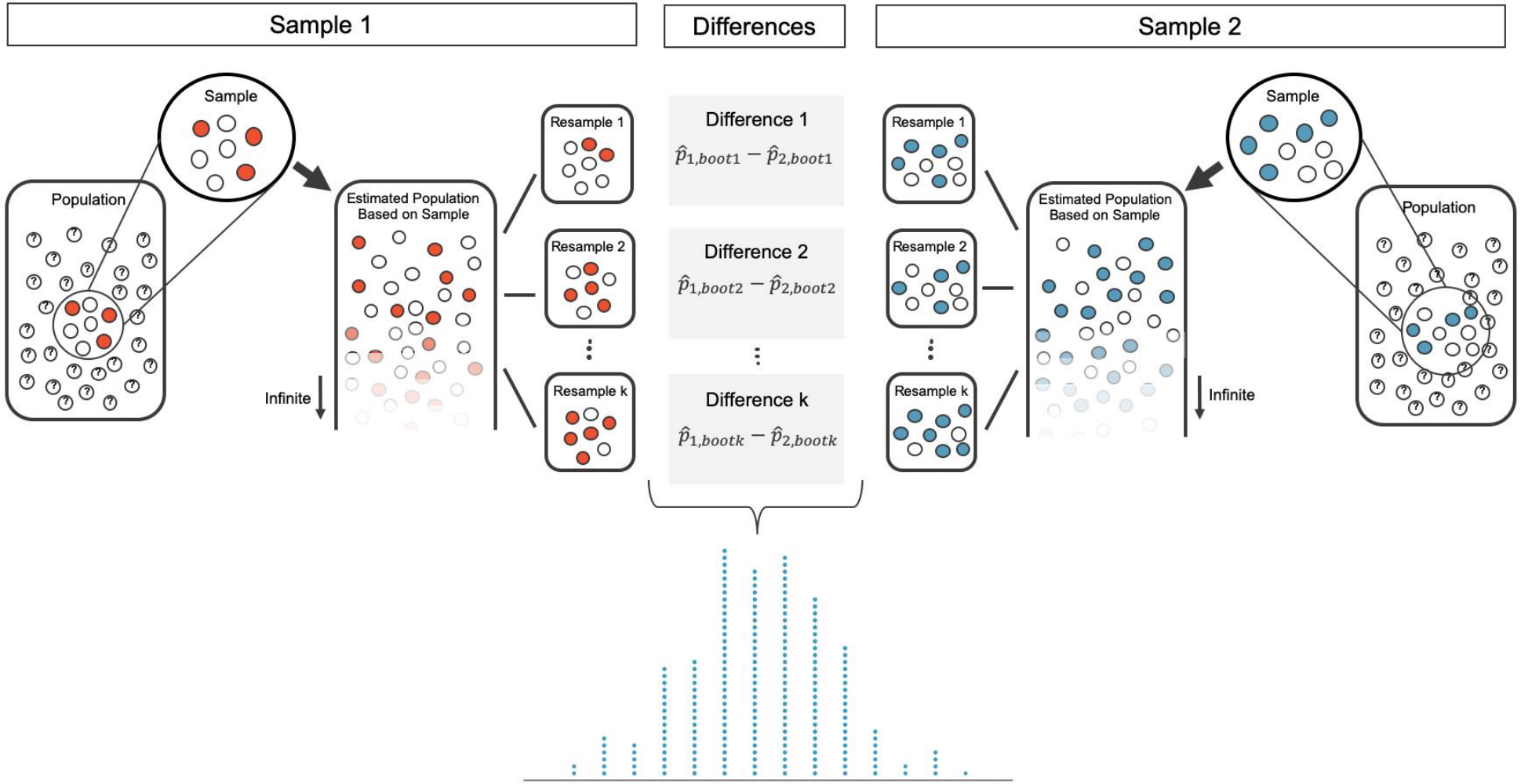
Bootstrap

Confidence interval with bootstrapping



Bootstrap

Two-sample bootstrap

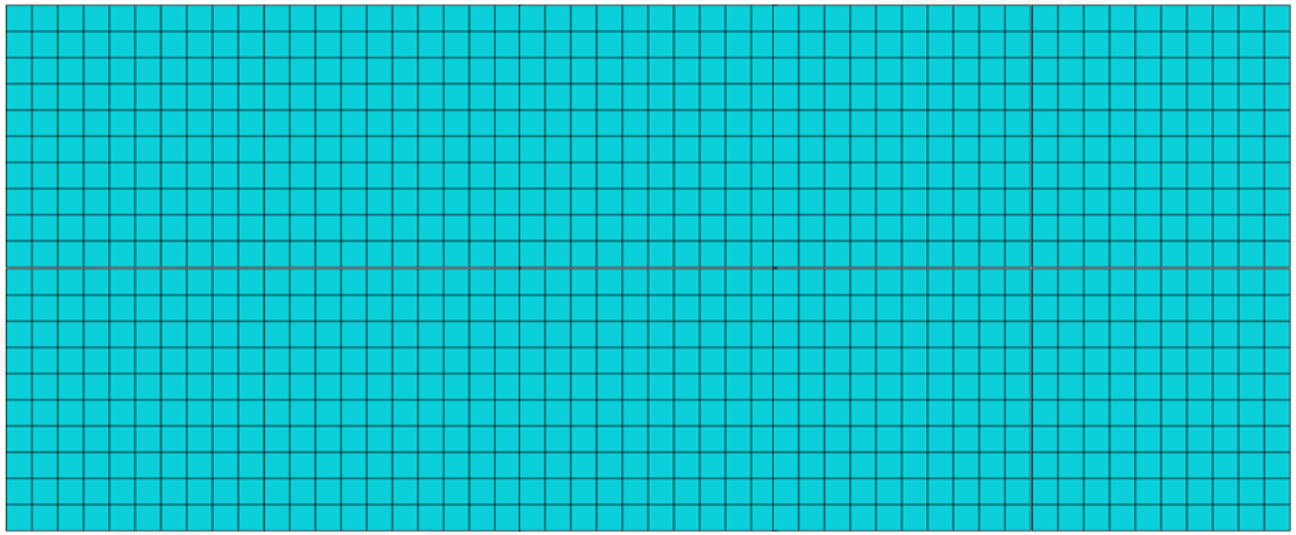


Multiple testing correction

False positives are a nightmare; false negatives are a tragedy.

Threshold on p-values for significant discoveries

p-value < 0.05 $= 5\%$ chance of false positive



Remember that *p*-value is uniform distributed under the null, small p-values can arise due to chance

False discovery rate

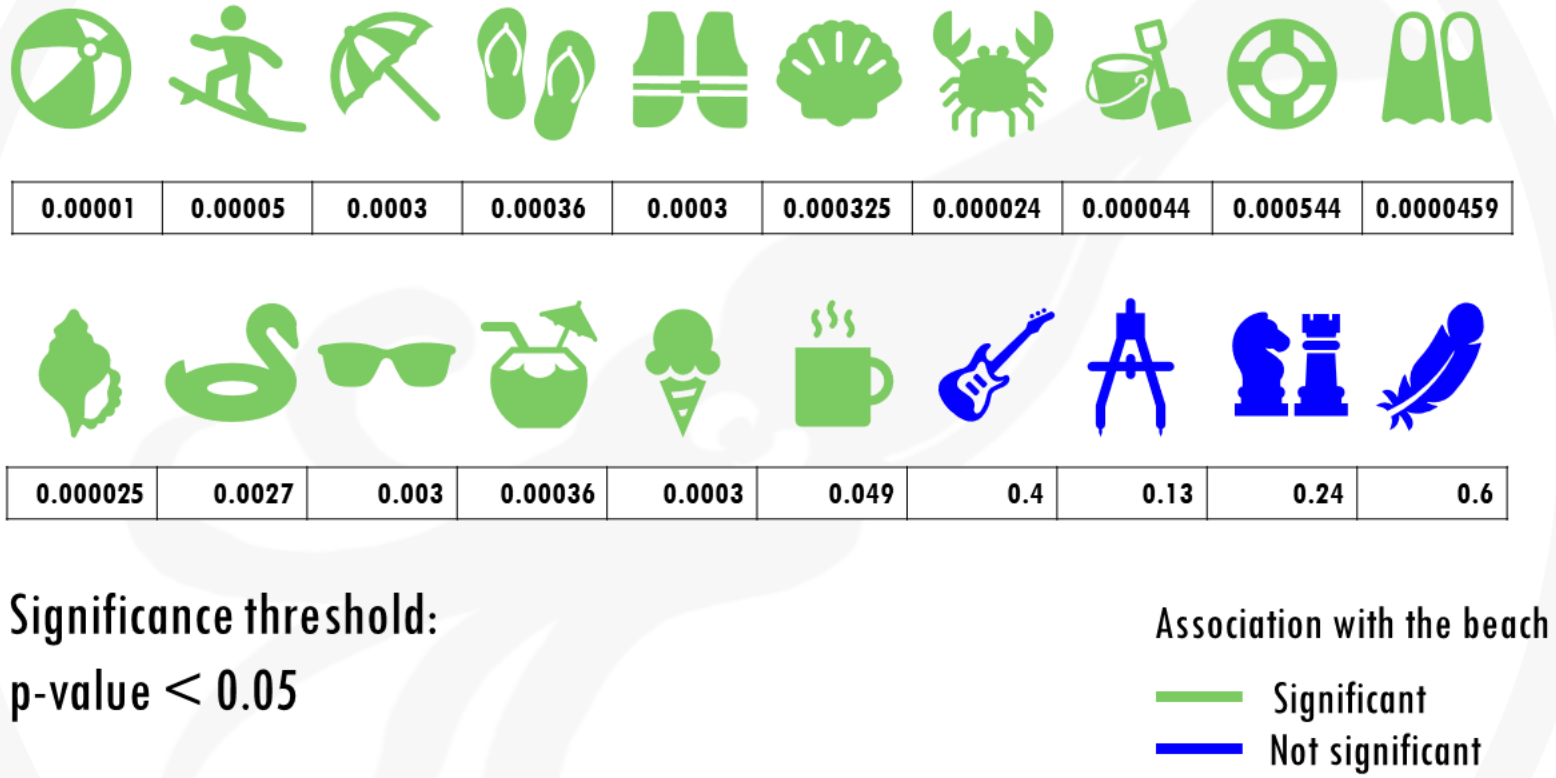
Genes with p-value < 0.05 which are actually not significant, it was just by chance that they got p-value < 0.05

$$\text{FDR} = \frac{\text{False positives}}{\text{All significant results}}$$

All significant genes (p-value < 0.05)

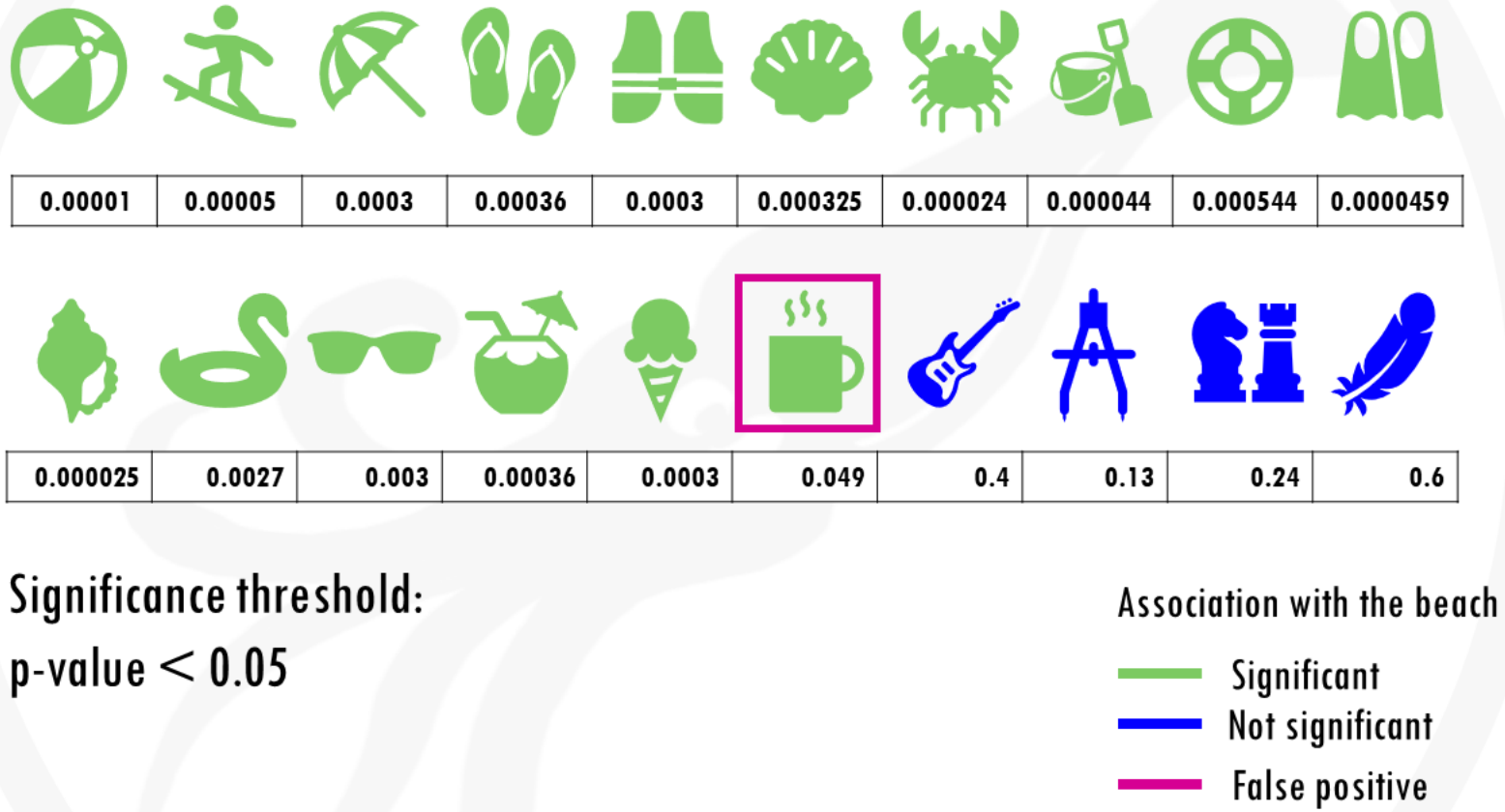
THE FALSE DISCOVERY RATE (FDR) IS THE PROPORTION OF FALSE POSITIVES AMONG ALL SIGNIFICANT RESULTS

Toy example



Test 20 objects to see if they are associated with beach

Toy example


0.00001
0.00005
0.0003
0.00036
0.0003
0.000325
0.000024
0.000044
0.000544
0.0000459
0.000025
0.0027
0.003
0.00036
0.0003
0.049
0.4
0.13
0.24
0.6
Significance threshold:
p-value < 0.05
Association with the beach
Significant
Not significant
False positive

Toy example

$$\text{FDR} = \frac{\textbf{False positives}}{\textbf{All significant findings}} = \frac{1}{16} = 6.25\%$$

p-value adjustment method

- Bonferroni correction: $\text{reject } H_0 \text{ if } p_i \leq \frac{\alpha}{m}$

- Benjamini-Hochberg procedure (one of the most cited stats paper, ~114k)

Step 1: Considering we have m p-values, each obtained from a single test: $p_1, p_2, ..., p_m$, we order the p-values in increasing order:

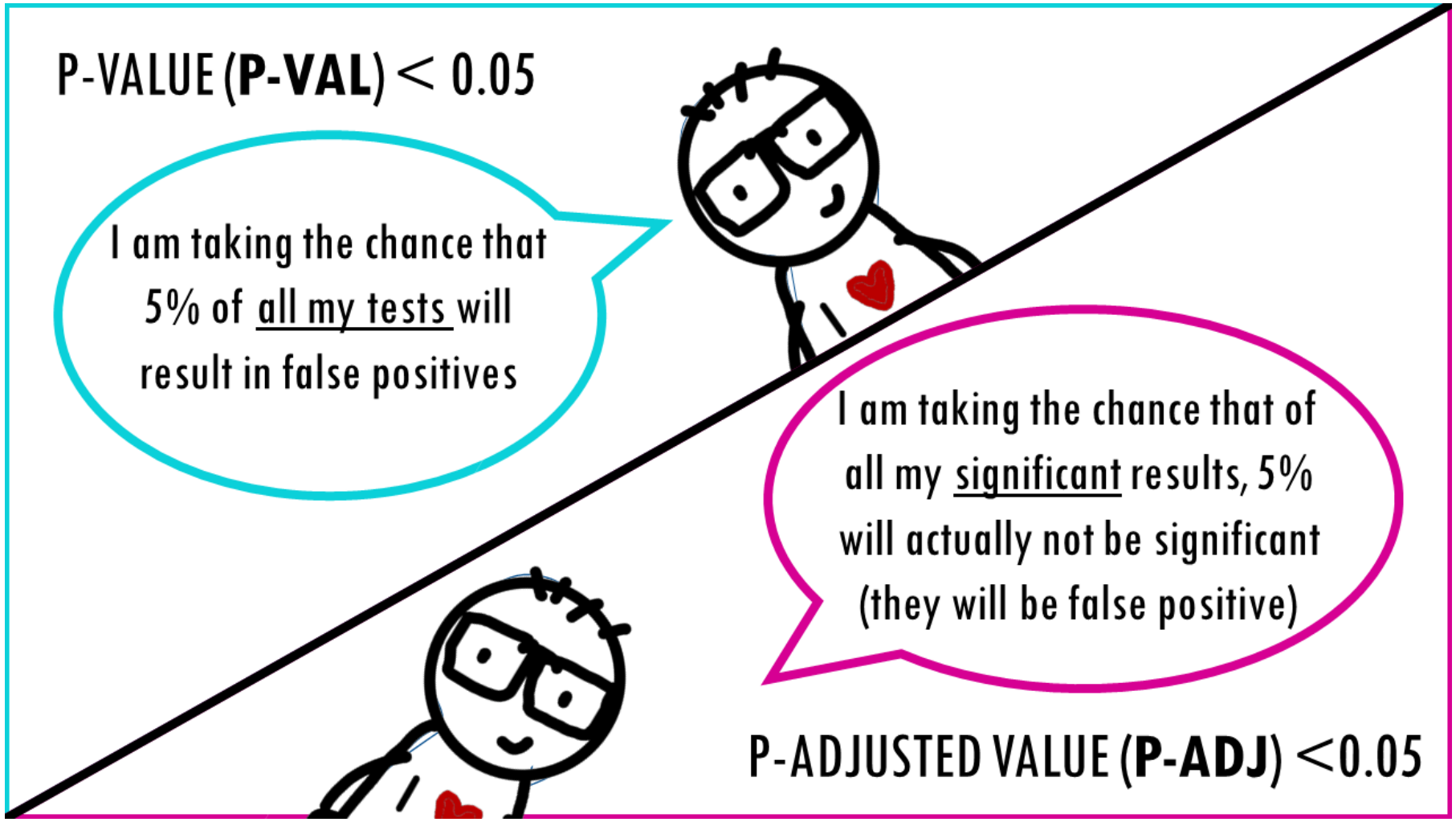
$$p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m-1)} \leq p_{(m)}$$

and denote their corresponding null hypotheses as:

$$H_{(1)}, H_{(2)}, ..., H_{(m-1)}, H_{(m)}$$

Step 2: Find integer k as the largest i such that: $p_{(i)} \leq \frac{i}{m}\alpha$. Then we reject all $H_{(i)}$ for which $i \leq k$.

Comparison


P-VALUE (P-VAL) < 0.05
I am taking the chance that 5% of all my tests will result in false positives
I am taking the chance that of all my significant results, 5% will actually not be significant (they will be false positive)
P-ADJUSTED VALUE (P-ADJ) <0.05

Toy example

0.00001	0.00005	0.0003	0.00036	0.0003	0.000325	0.000024	0.000044	0.000544	0.0000459

0.000025	0.0027	0.003	0.00036	0.0003	0.049	0.4	0.13	0.24	0.6

Significance threshold:

p-value < 0.05

→ p-value < 0.0025 (Bonferroni correction)

Association with the beach

- Significant
- Not significant
- False positive

Toy example

p-val	0.00001	0.00005	0.0003	0.00036	0.0003	0.000325	0.000024	0.000044	0.000544	0.0000459
p-adj	**0.0005**	**0.000075**	**0.00083**	**0.0008**	**0.0009**	**0.00084**	**0.0005**	**0.00094**	**0.008**	**0.00046**

p-val	0.000025	0.0027	0.003	0.00036	0.0003	0.049	0.4	0.13	0.24	0.6
p-adj	**0.0004**	**0.0055**	**0.02**	**0.0046**	**0.004**	**0.053**	**0.6**	**0.28**	**0.45**	**0.7**

Significance threshold:

p-adj < 0.05

0.05 * 15 = 0.75 objects falsely significant (falsely associated with the beach)

Association with the beach

- Significant
- Not significant
- False positive

Let’s do some practice!

Thanks

Q & A