

Introduction to Modern Statistics

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

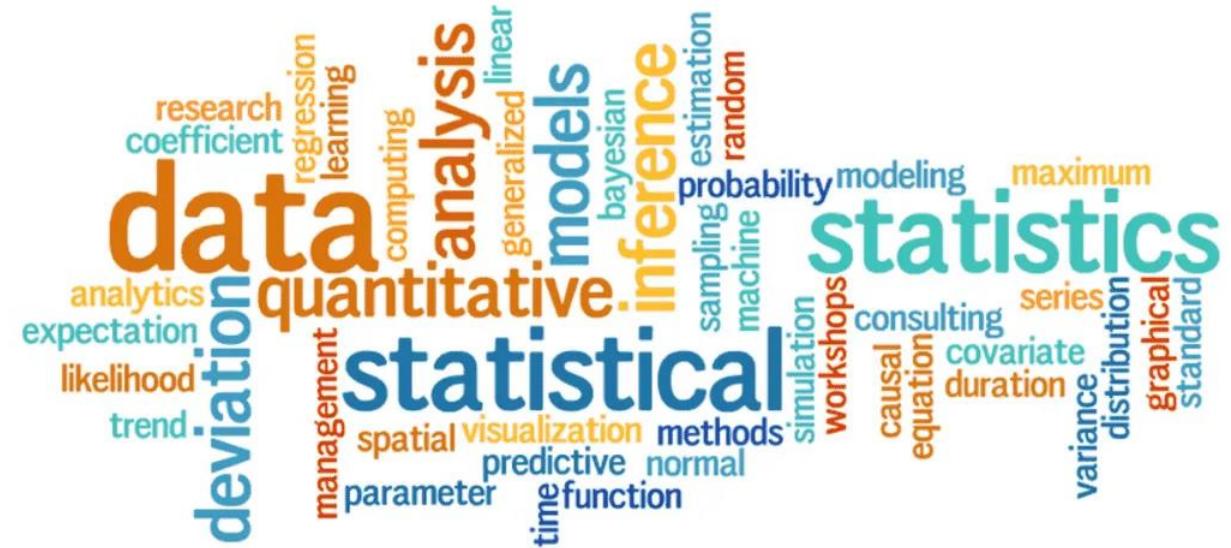
Notation of the slides

- 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



Day 2: Statistical inference

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

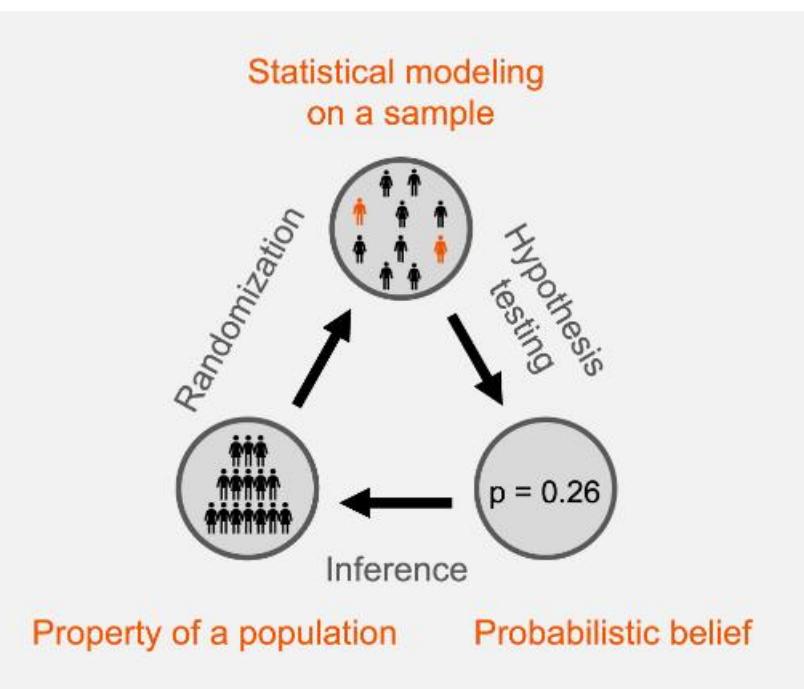
Overview

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

- ❑ Causes
- ❑ Examples in real life
- ❑ How do we react to it



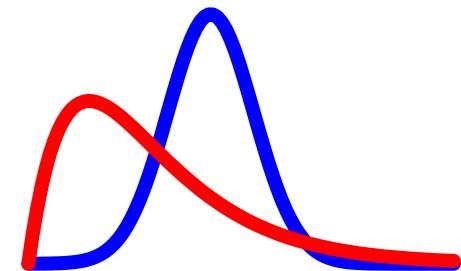
Summary – Day1

Probability

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



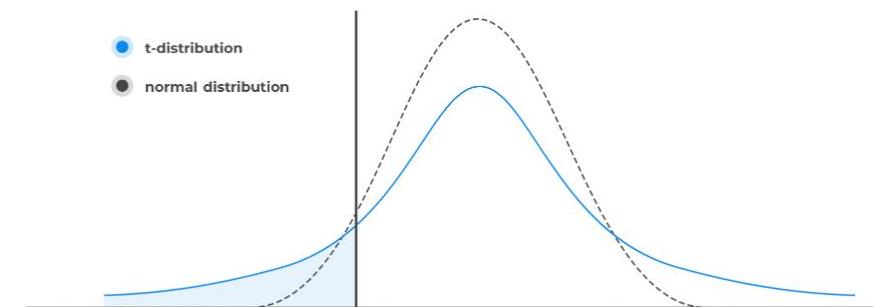
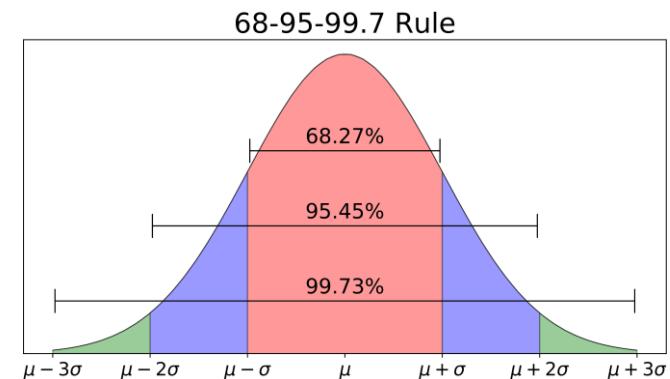
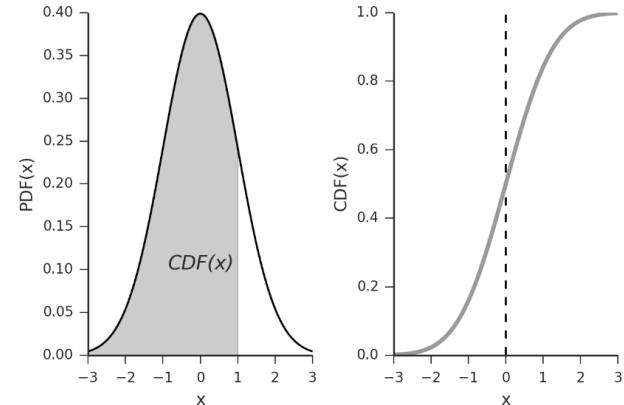
$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$



Summary – Day1

Distributions

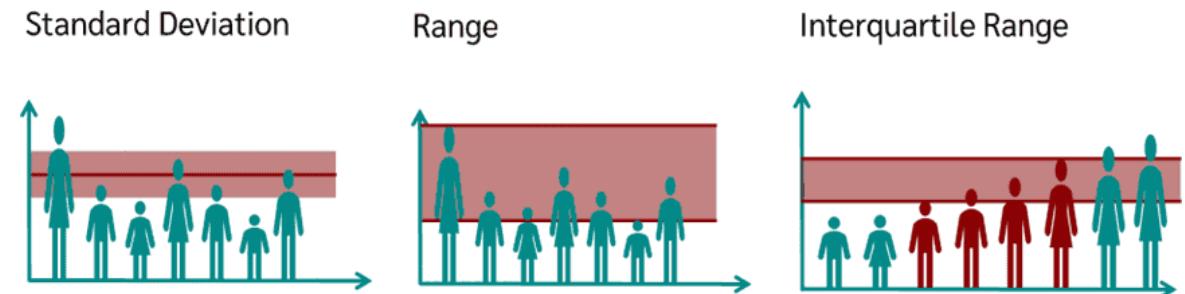
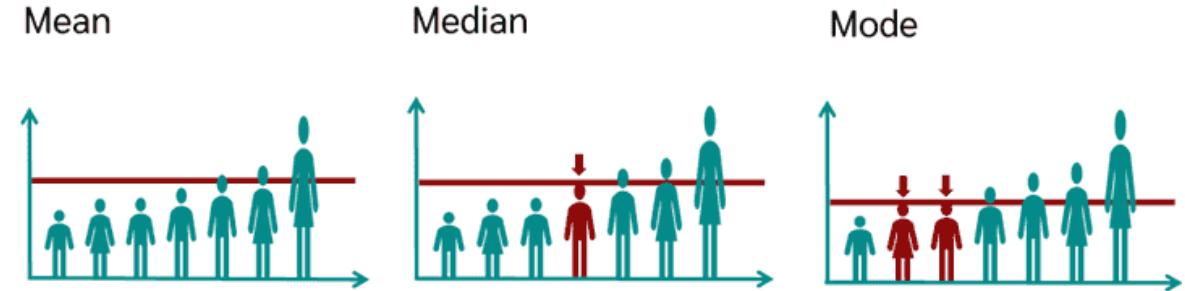
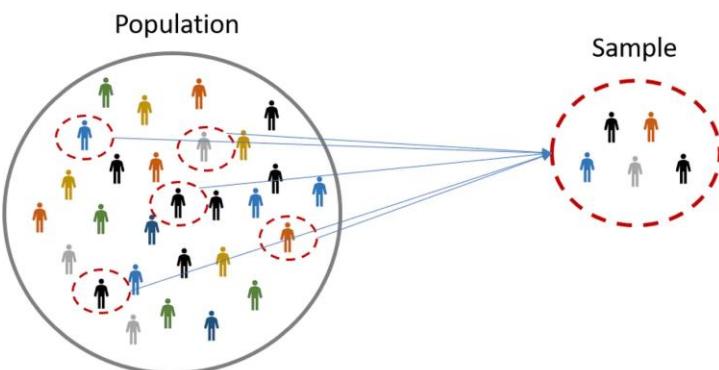
Category	Name	Description
Discrete	Uniform	equal probability to every integer between a and b
	Bernoulli	single trial with two possible outcomes with success rate p
	Binomial	number of successes in n independent Bernoulli trials
	Negative Binomial	number of failures before achieving r successes
	Geometric	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



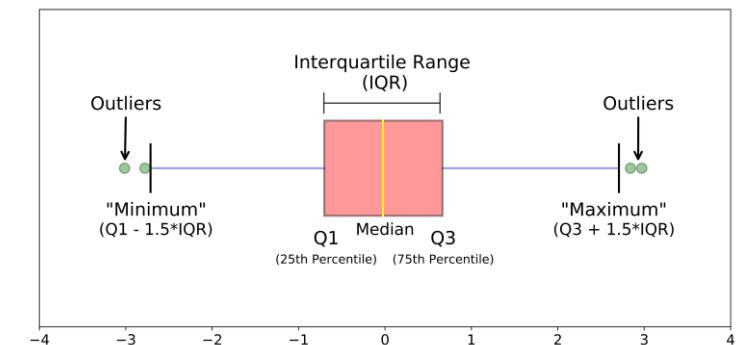
Summary – Day1

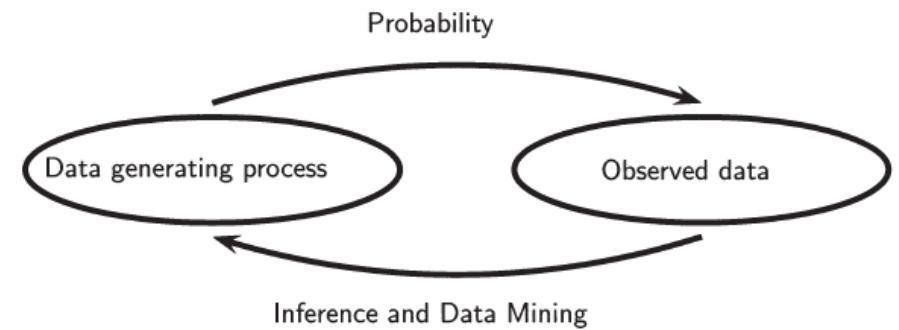
Descriptive statistics

- Tendency
- Dispersion
- Skewness
- Distribution



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



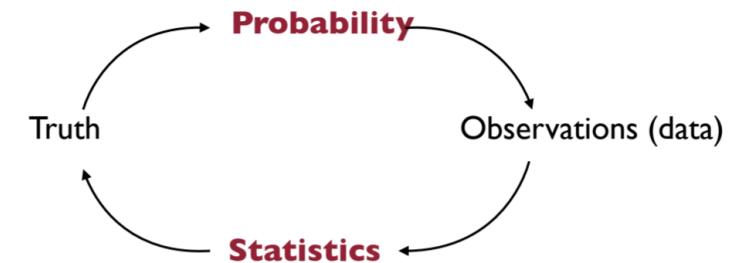
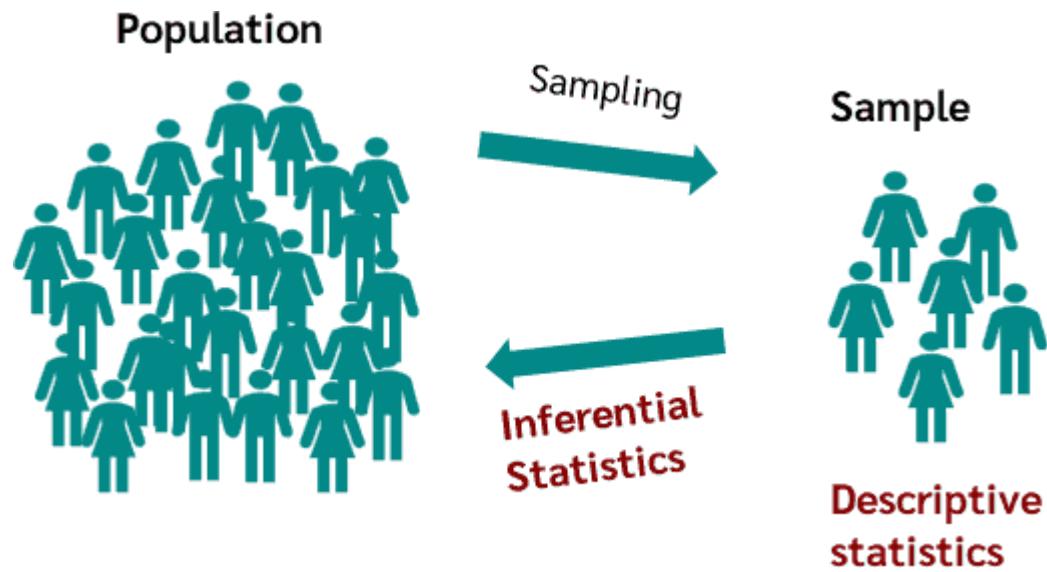


Inferential statistics

Use the part to learn about the whole

Inferential statistics

Draw conclusions about the population based on sample data



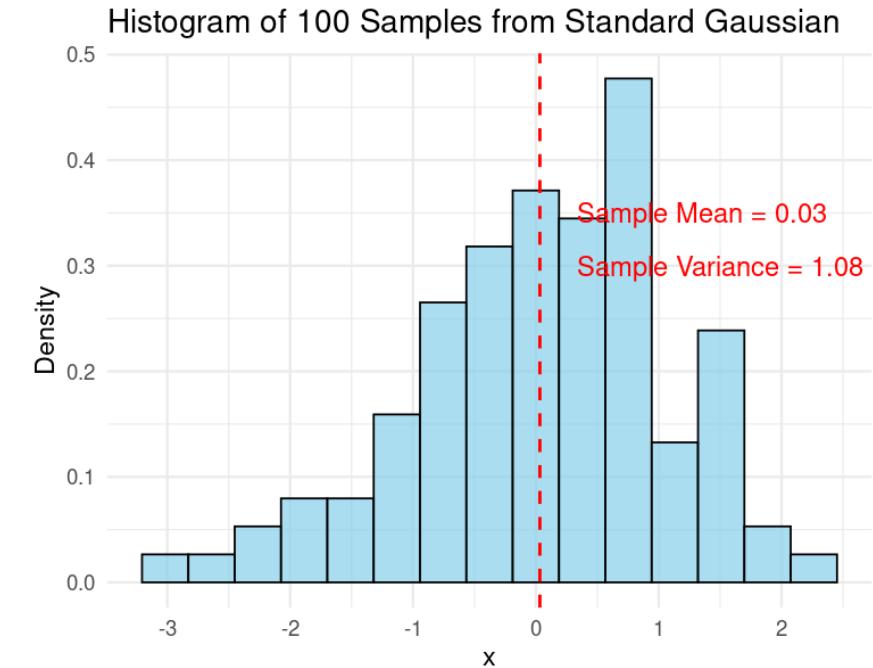
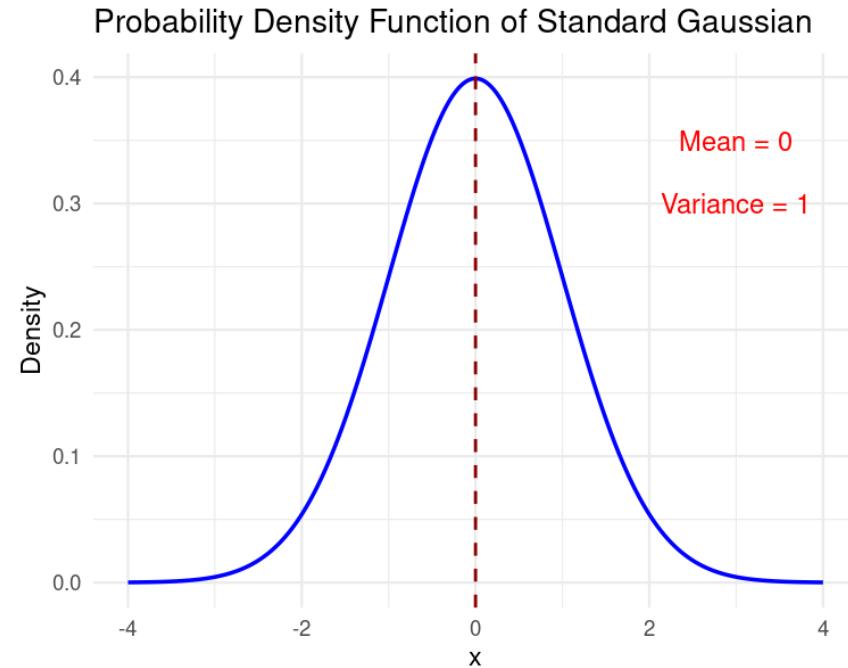
Typical tasks

- Parameter estimation
- Hypothesis testing

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

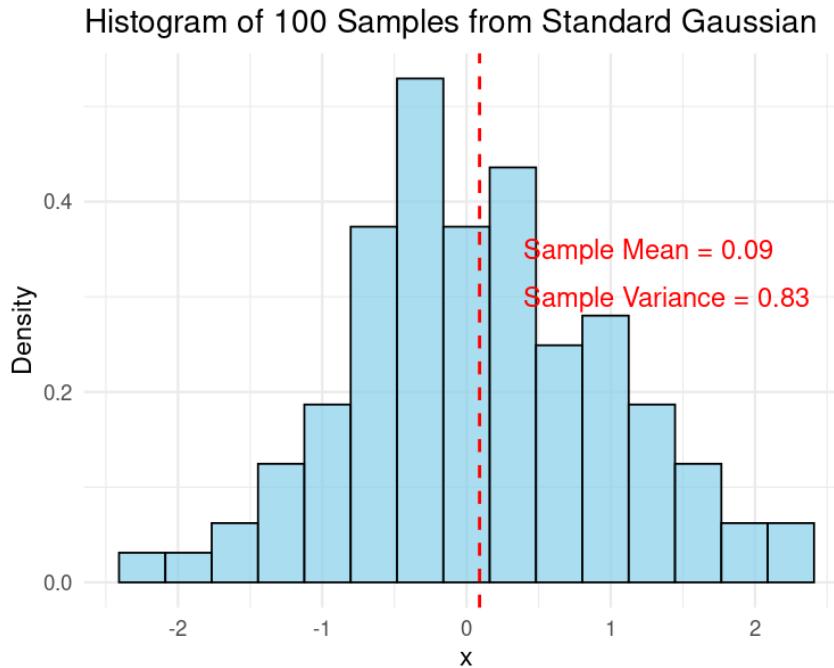
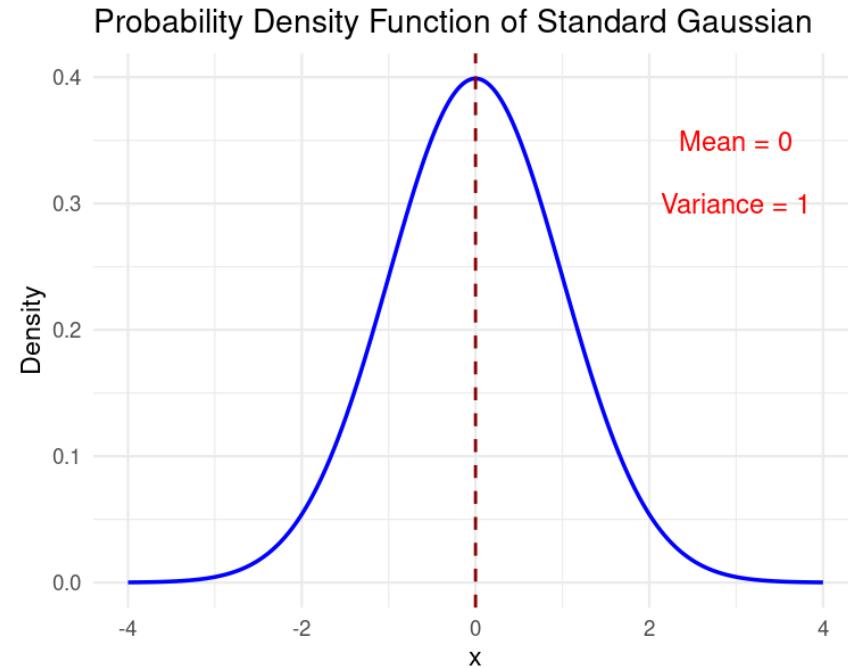
Population vs samples

Use standard gaussian as an example



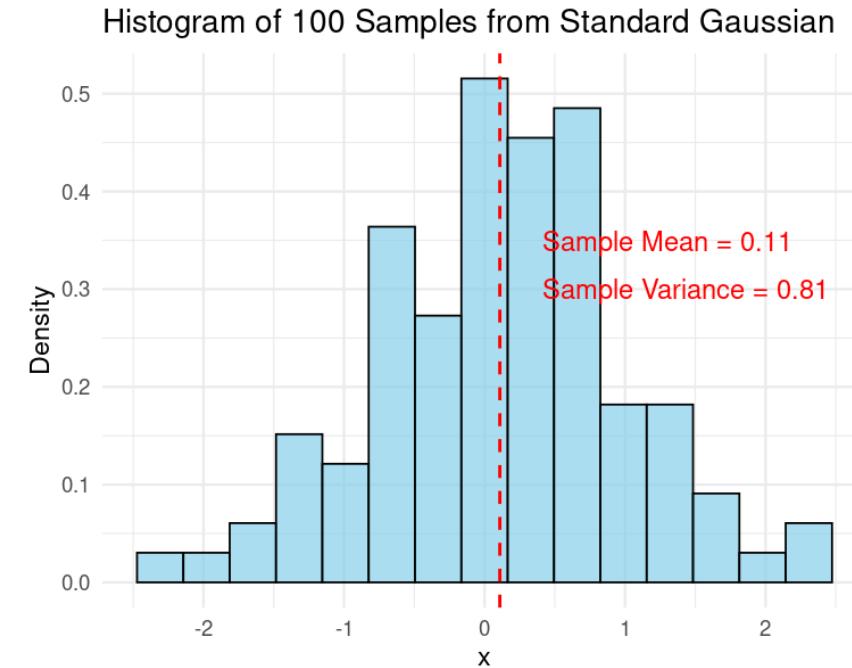
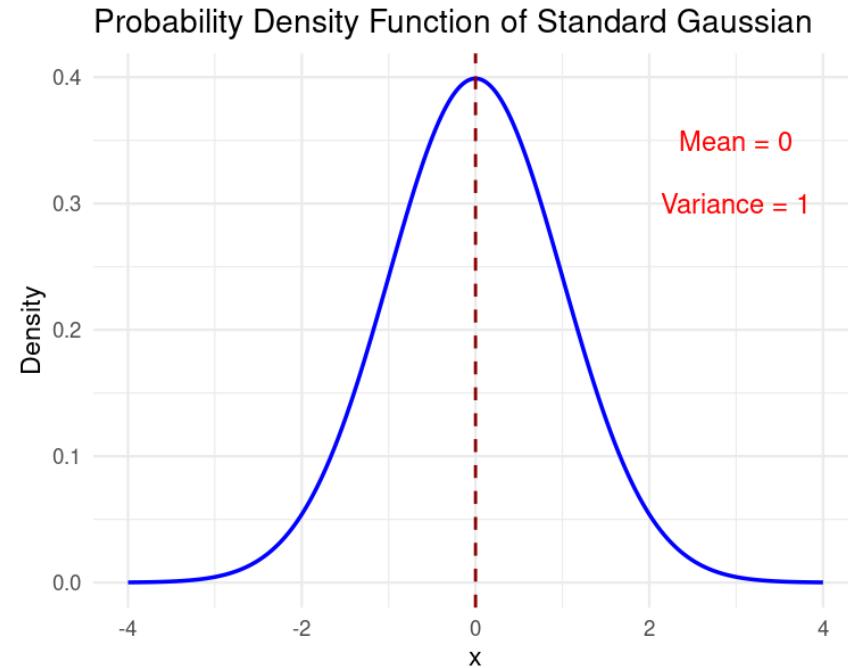
Population vs samples

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:

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

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

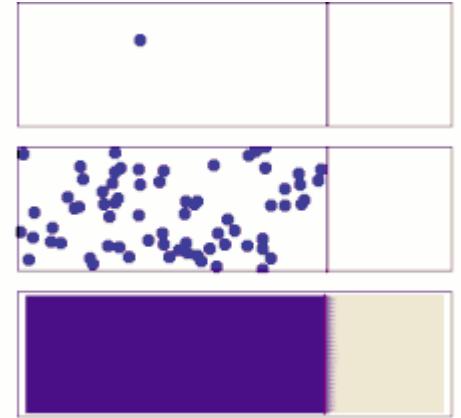


Statistic: a summary calculated from samples

Law of large numbers (LLN)

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

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



Diffusion is an example of the law of large numbers.

In summary:

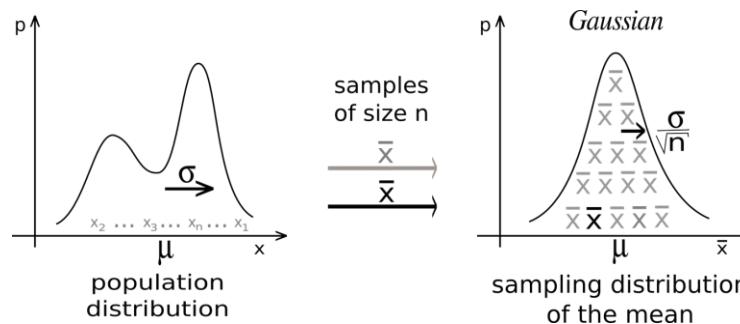
- Sample mean converges to population mean as sample size increases
- LLN provides Implications for estimation and reliability

Central limit theorem (CLT)

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

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

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

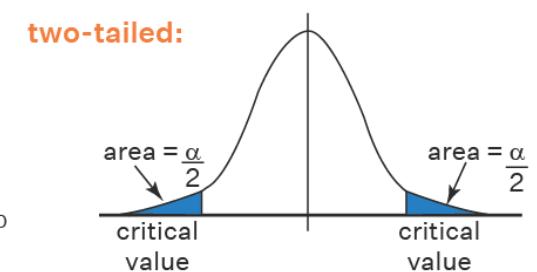
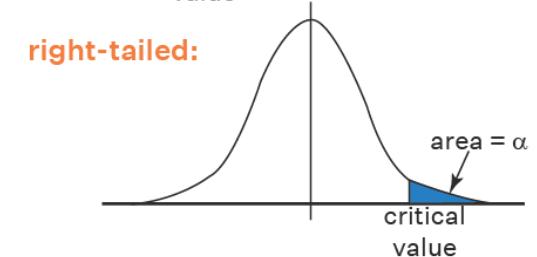
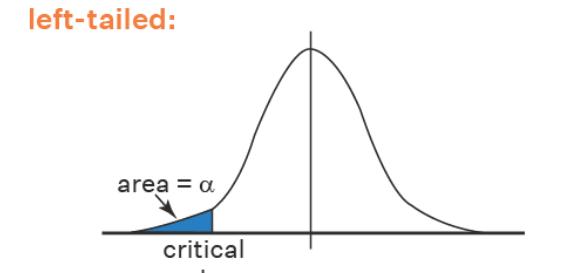
A process to determine if there is enough evidence to **reject** a null hypothesis

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

- Reject H_0
- Do not reject H_0

Rejection Region for Null Hypothesis

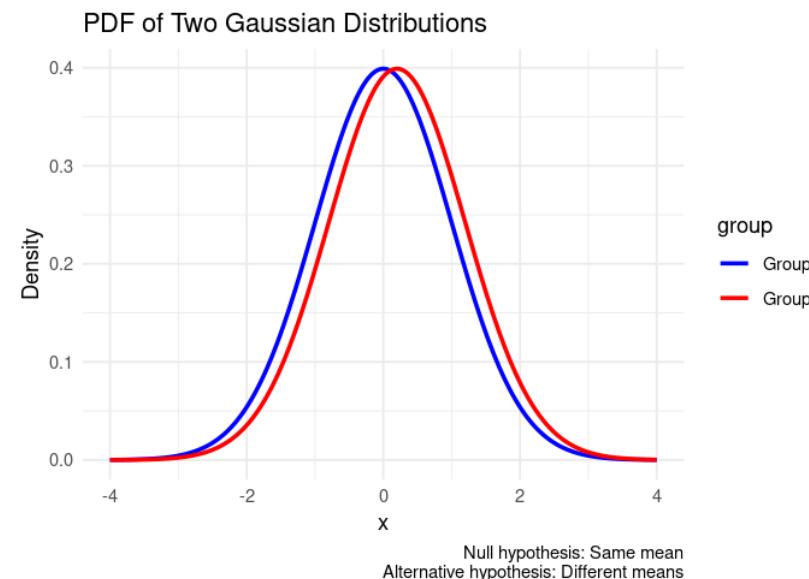
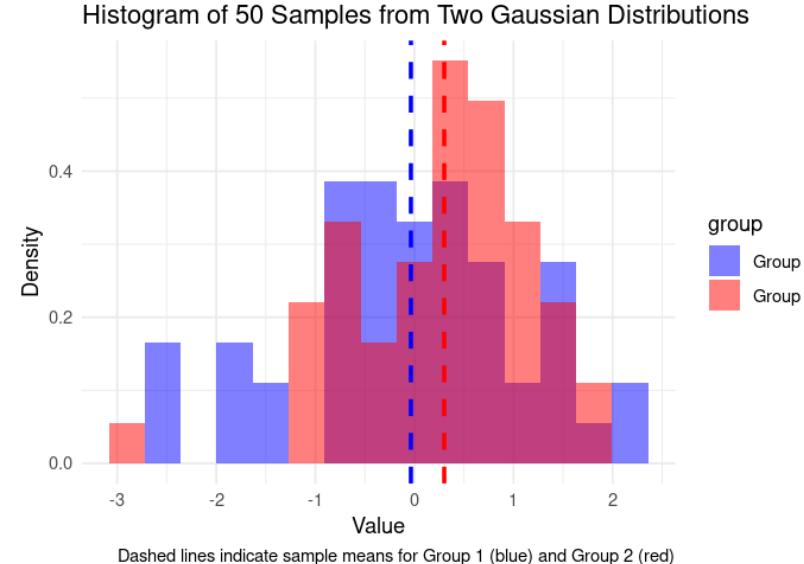


Let's see an example

Use gaussian distributions as an example

❑ Frame the hypothesis

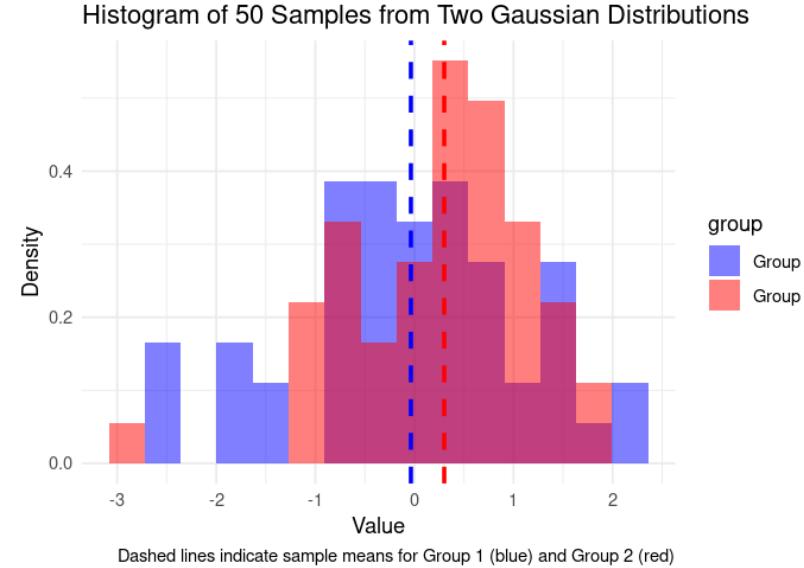
- Null hypothesis (H_0): two group have the same mean
- Alternative hypothesis (H_a): two group have the different mean



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?

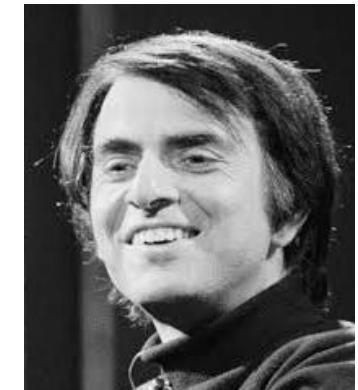
No, this is not a binary world

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.



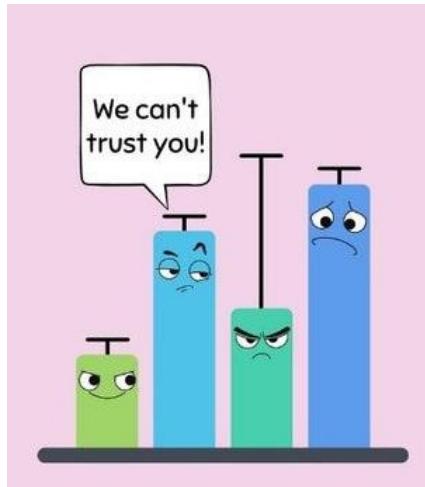
Carl Edward Sagan  
(1934-1996)

*“Absence of evidence is not evidence of absence”*

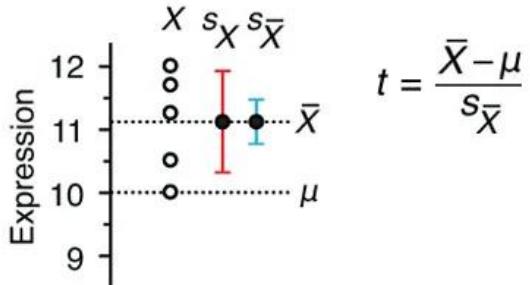
# Statistical tests

- Test mean:

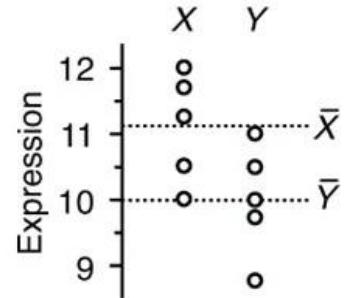
- One group
- Two group
- Paired
- Multiple group



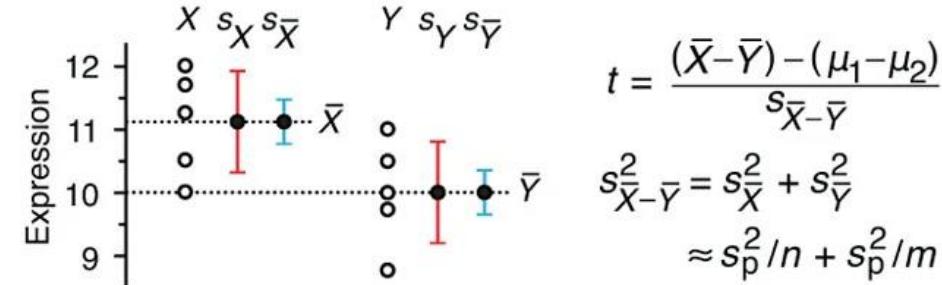
One-sample  $t$ -test



Independent samples

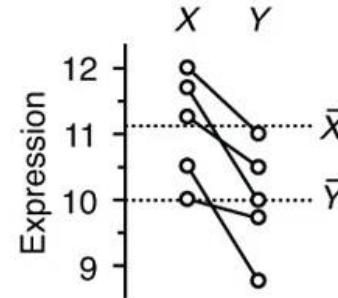


Two-sample  $t$ -test



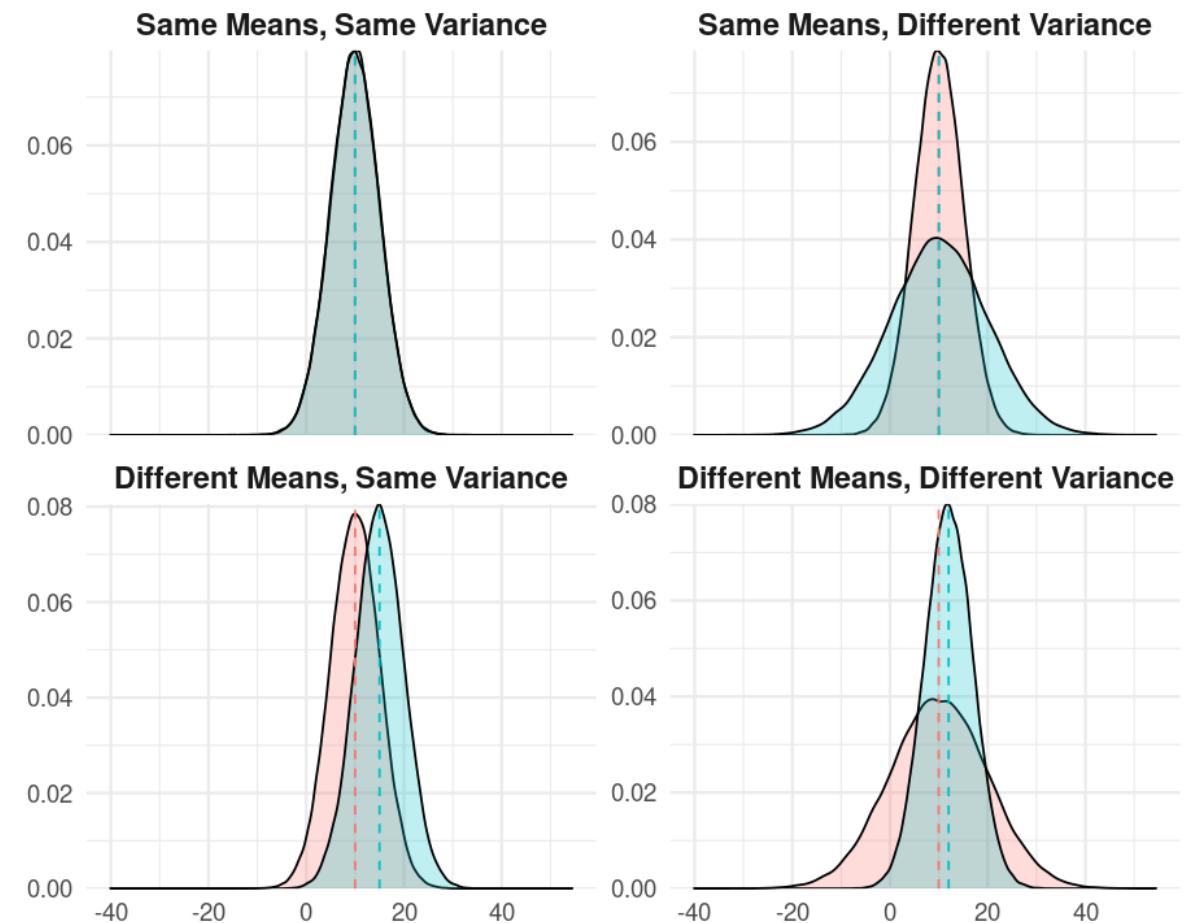
$$t = \frac{(\bar{X}-\bar{Y}) - (\mu_1 - \mu_2)}{s_{\bar{X}-\bar{Y}}}$$
$$s_{\bar{X}-\bar{Y}}^2 = s_X^2 + s_Y^2$$
$$\approx s_p^2/n + s_p^2/m$$

Paired samples



# Statistical tests

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



# Which test to use?

Because flow-charts are fun!

Parametric test

Non-Parametric test

## TESTING HYPOTHESES ABOUT THE MEAN

**One sample Z-test**  
 $H_0: \mu = \mu_0$   
Test statistic:  
 $\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$

**One sample t-test**  
 $H_0: \mu = \mu_0$   
Test statistic:  
 $\frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$

**Paired t-test**  
collapse to one-sample test for differences  
 $H_0: \mu_d = 0$   
Test statistic:  
 $\frac{D}{s_d/\sqrt{n}} \sim t_{n-1}$

**Wilcoxon Signed Rank test**  
 $H_0:$  Differences have symmetric distribution  
Test statistic:  
 $\frac{W_+ - E(W_+)}{\sqrt{Var(W_+)}} \sim N(0,1)$

**How many samples?**

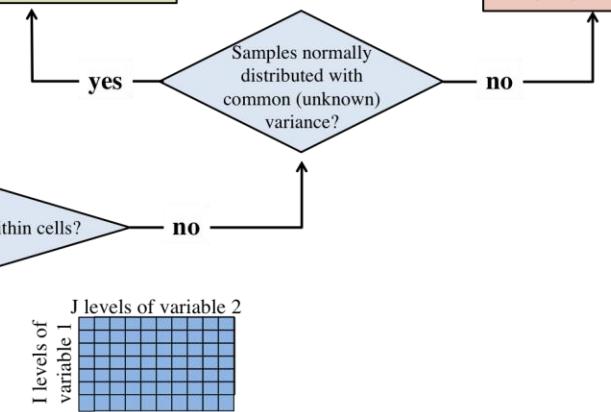
**2 samples**

**Paired samples?**

**Two-sample t-test**  
 $H_0: \mu_X = \mu_Y$   
Test statistic:  
 $\frac{\bar{X} - \bar{Y}}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$

**Two-way ANOVA**  
 $H_{0A}: \mu_{1..} = \dots = \mu_I$   
 $H_{0B}: \mu_{1..} = \dots = \mu_J$   
 $H_{0int}: \text{no interaction}$

**Randomized Block Design**  
 $H_{0A}: \mu_1 = \dots = \mu_I$   
 $\frac{SS_A / (I-1)}{SS_{AB} / (I-1)(J-1)} \sim F_{I-1, (I-1)(J-1)}$

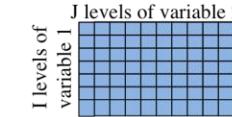


Number of variables (samples correspond to different levels of a variable)

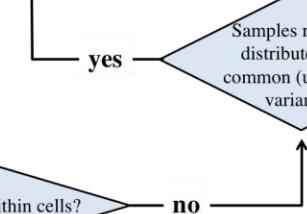
**1 variable**

**One-way ANOVA**  
 $H_0: \mu_1 = \dots = \mu_I$   
Test statistic:  
 $\frac{SS_B / (I-1)}{SS_W / (I(J-1))} \sim F_{I-1, J(J-1)}$

**2 variables**



**Kruskal Wallis test**  
 $H_0: F_1 = \dots = F_I$   
Test statistic:  
 $\frac{12}{IJ(IJ+1)} SS_B \sim \chi^2_{I-1}$



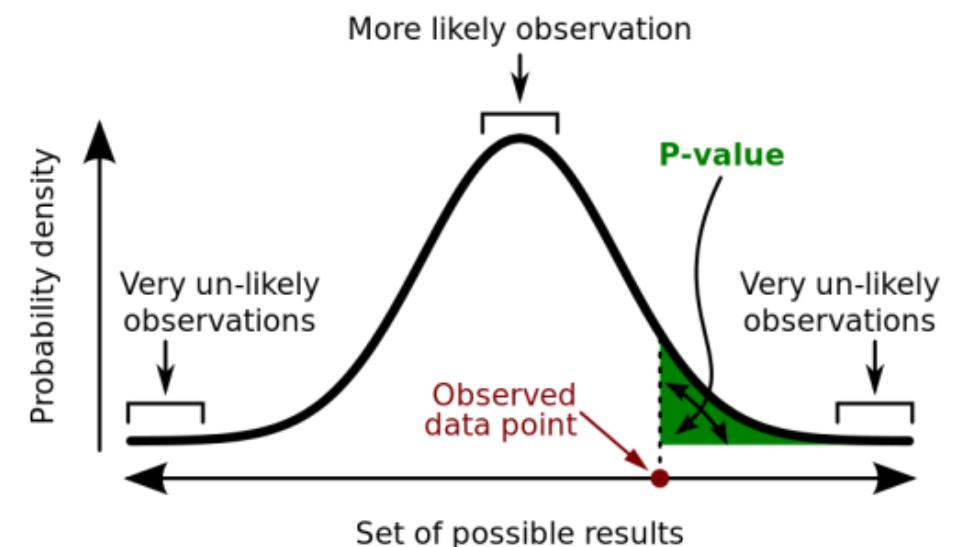
# *p*-value

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,

<i>p</i> -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?



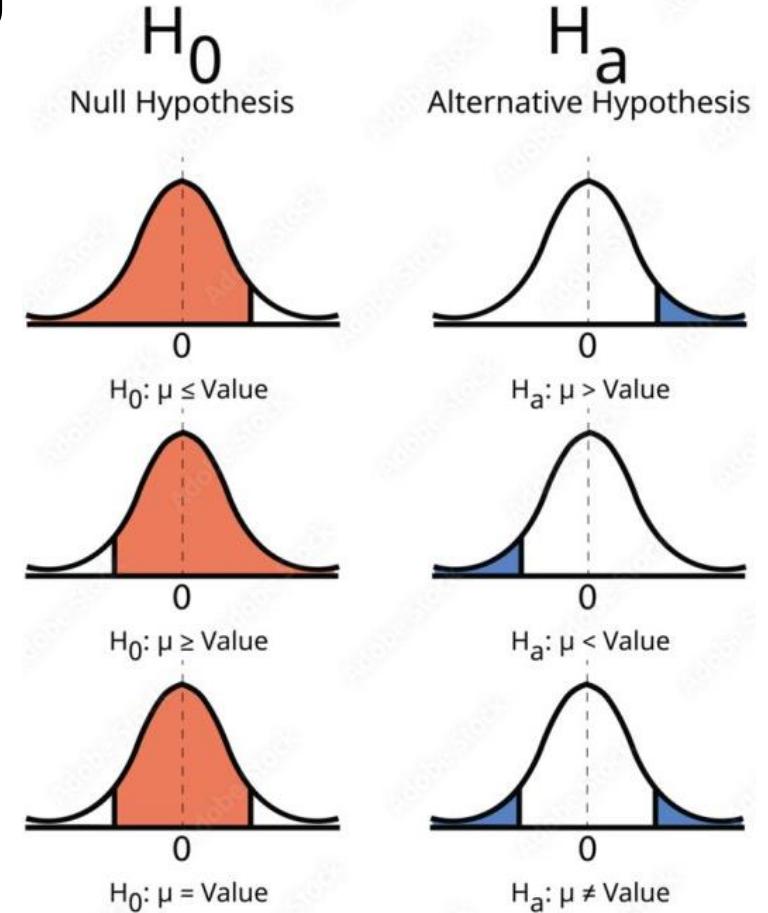
- large  $p$ -value is strong evidence in favor of  $H_0$ .
  - False
- The  $p$ -value is the probability that the null hypothesis is true
  - False
- $p = 0.05$  means there's a 95% chance  $H_0$  is true
  - False

# One-side/two-side tests

Remember that  $p$ -value is the probability of observing the same as or more extreme event. (under  $H_0$ )

Depending on  $H_0$

- Test "equal": two-side
- Test "smaller than": one-side
- Test "greater than": one-side



# $p$ -value distribution when $H_0$ is true

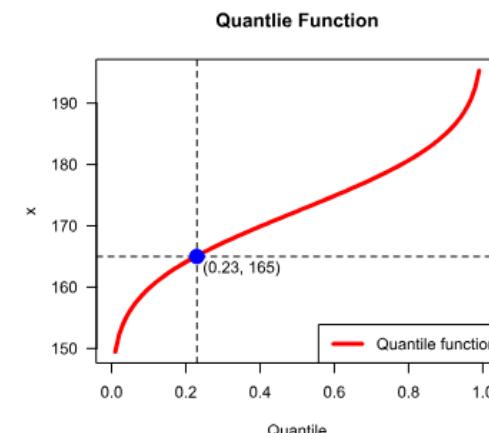
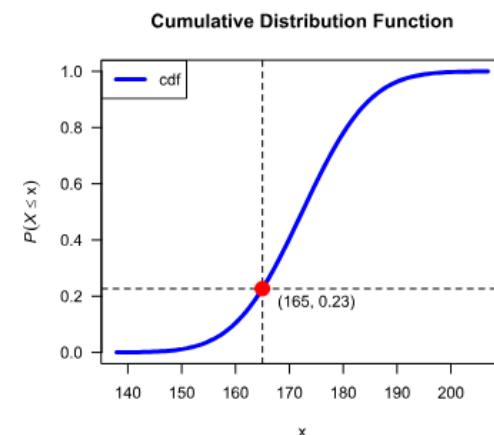
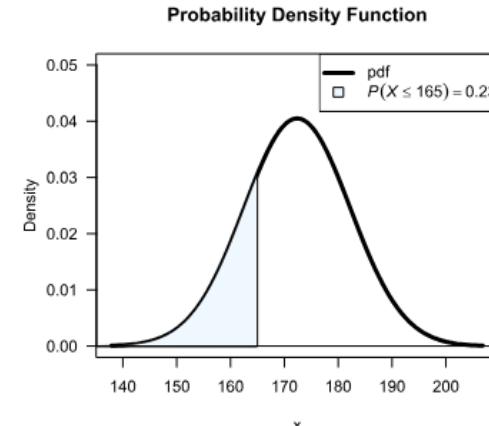
Let a random value  $P$  correspond to Test statistics  $T$   
( $t$ ,  $p$ : some fixed threshold)

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

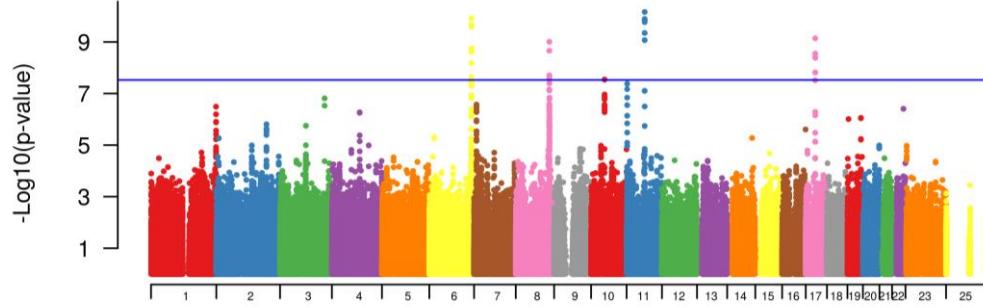
$p$ -value follows **uniform** distribution under the Null

We can use this rationale for diagnostics!

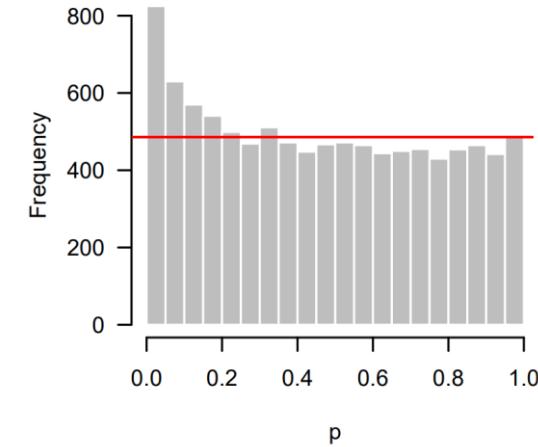
- Histogram of  $p$ -values
- Quantile-Quantile plot (QQ plot)



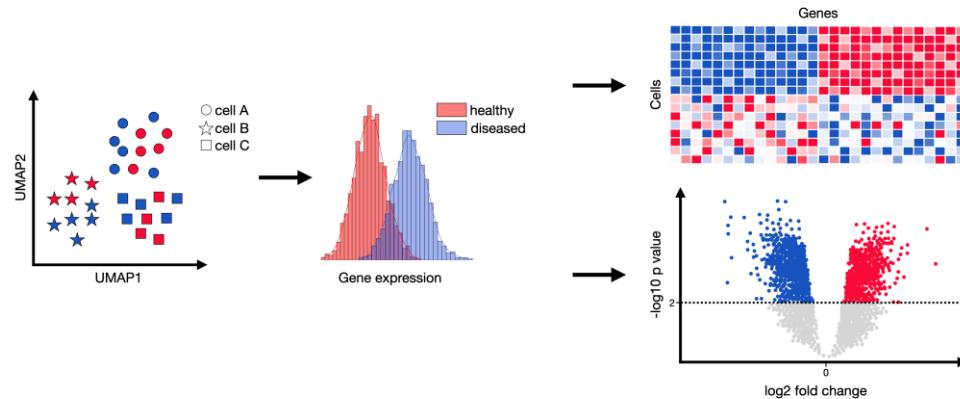
# Examine $p$ -value distribution



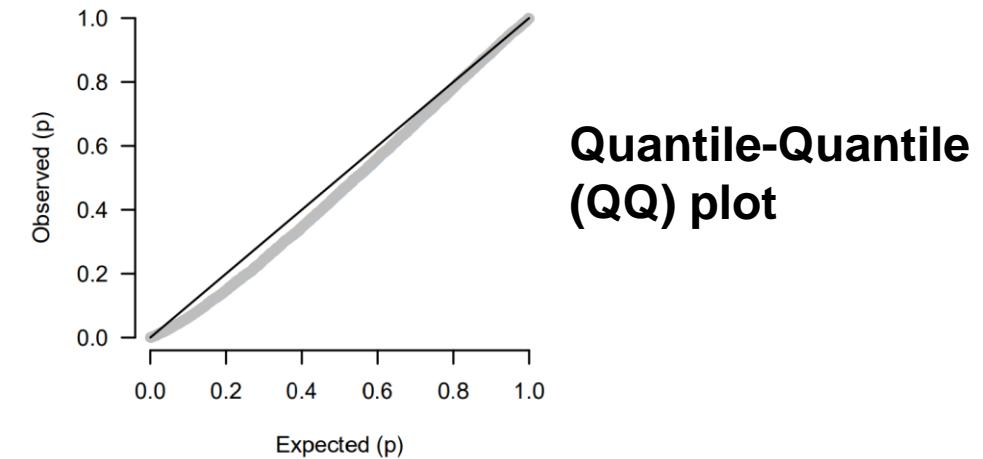
**Genetics**



**Histogram**



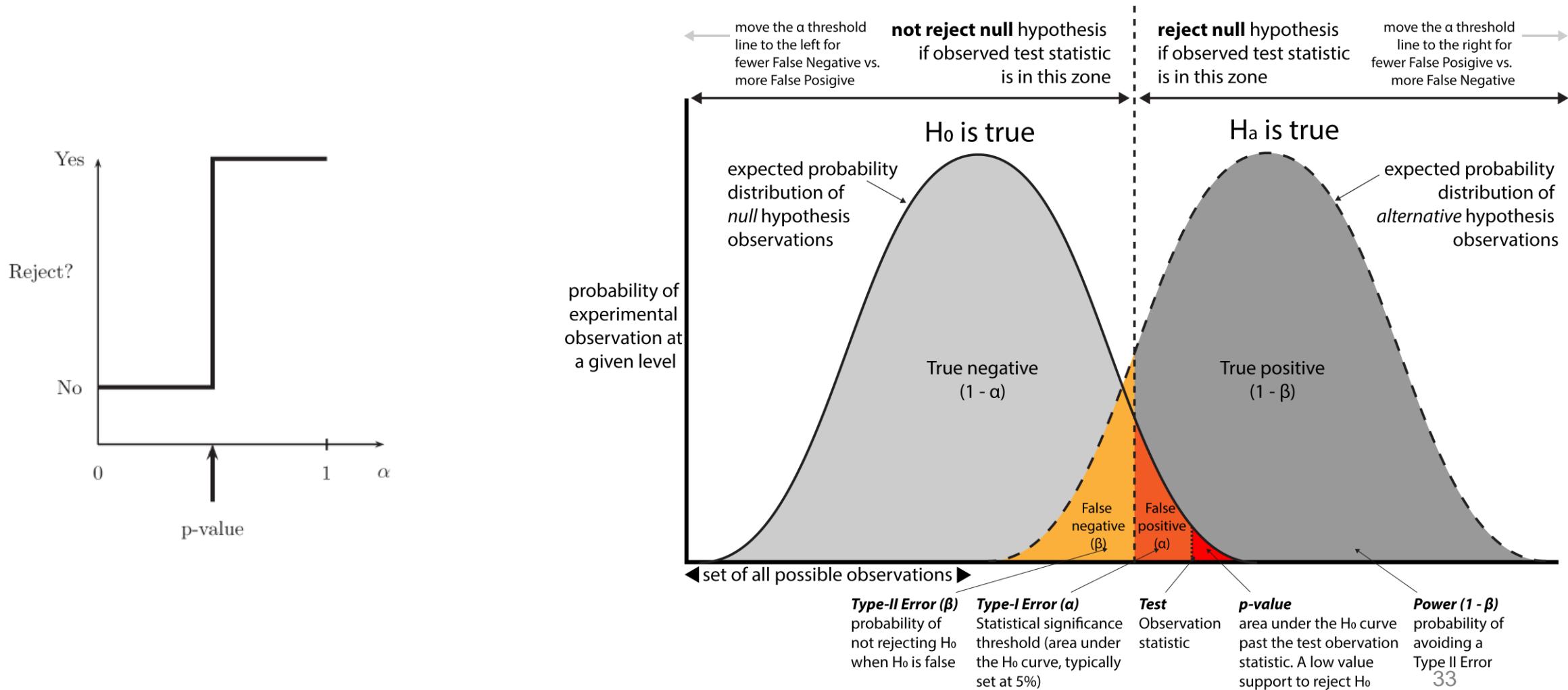
**Transcriptomics**



**Quantile-Quantile (QQ) plot**

# Decision errors

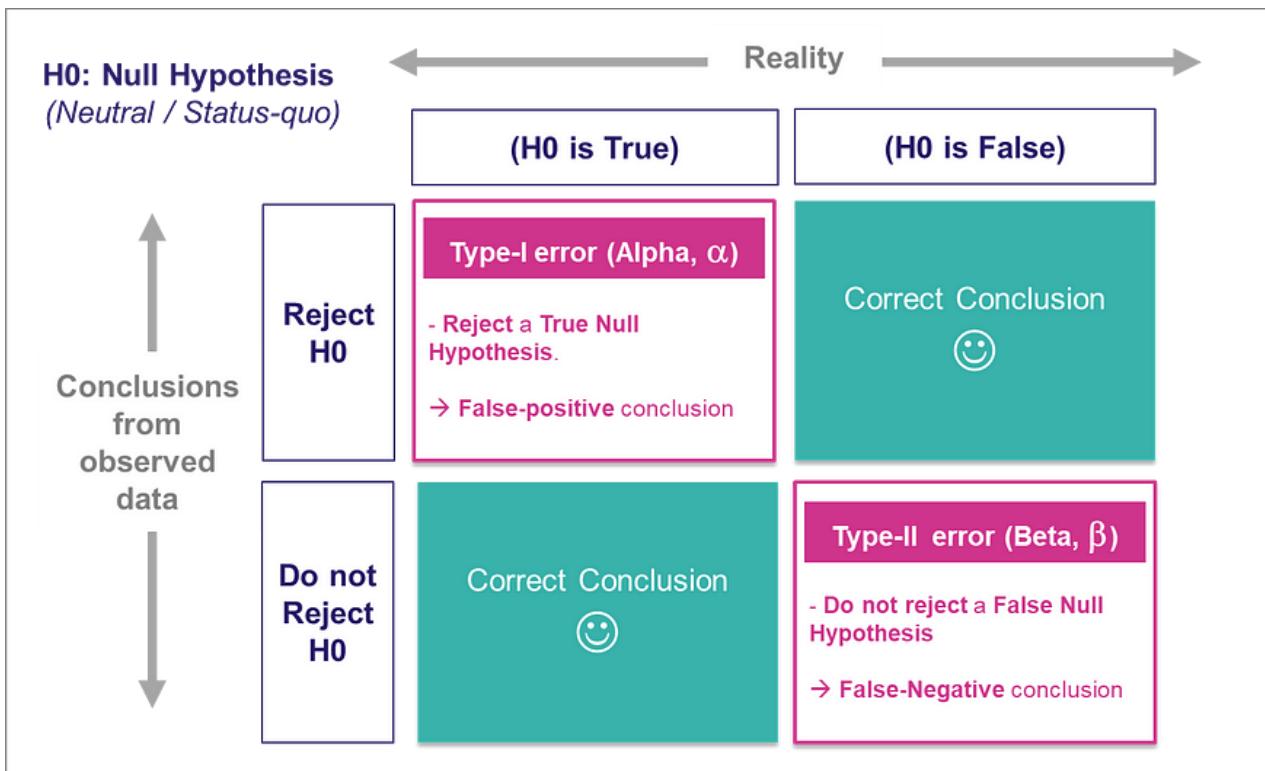
We use statistics to make decisions



# 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



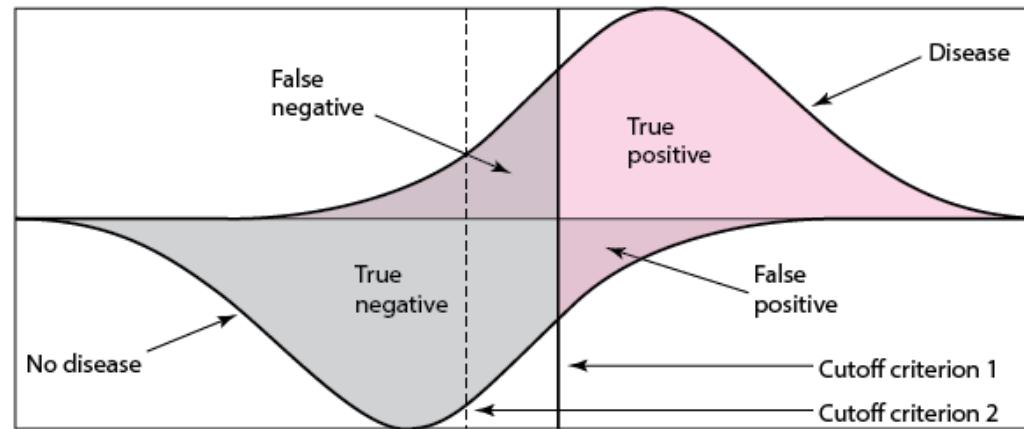
Type I Error



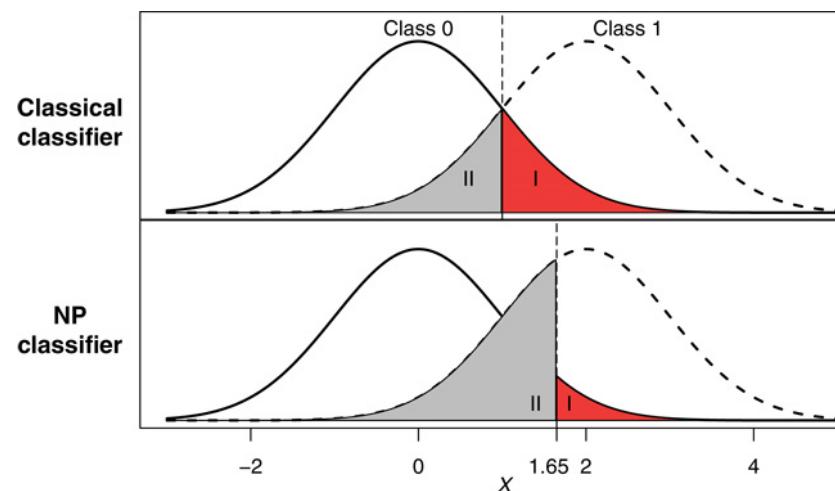
Type II Error



# Decision errors and asymmetry classification



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



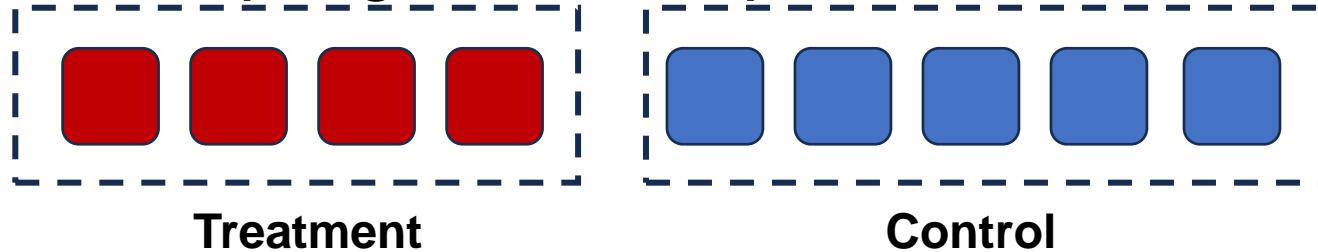


# Permutation and Bootstrap

Computer Age Statistical Inference

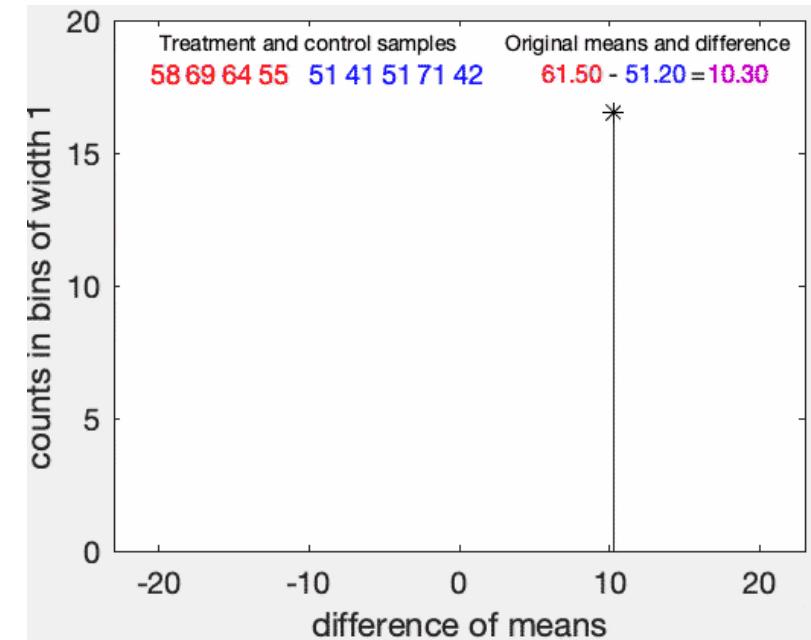
# Permutation test

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

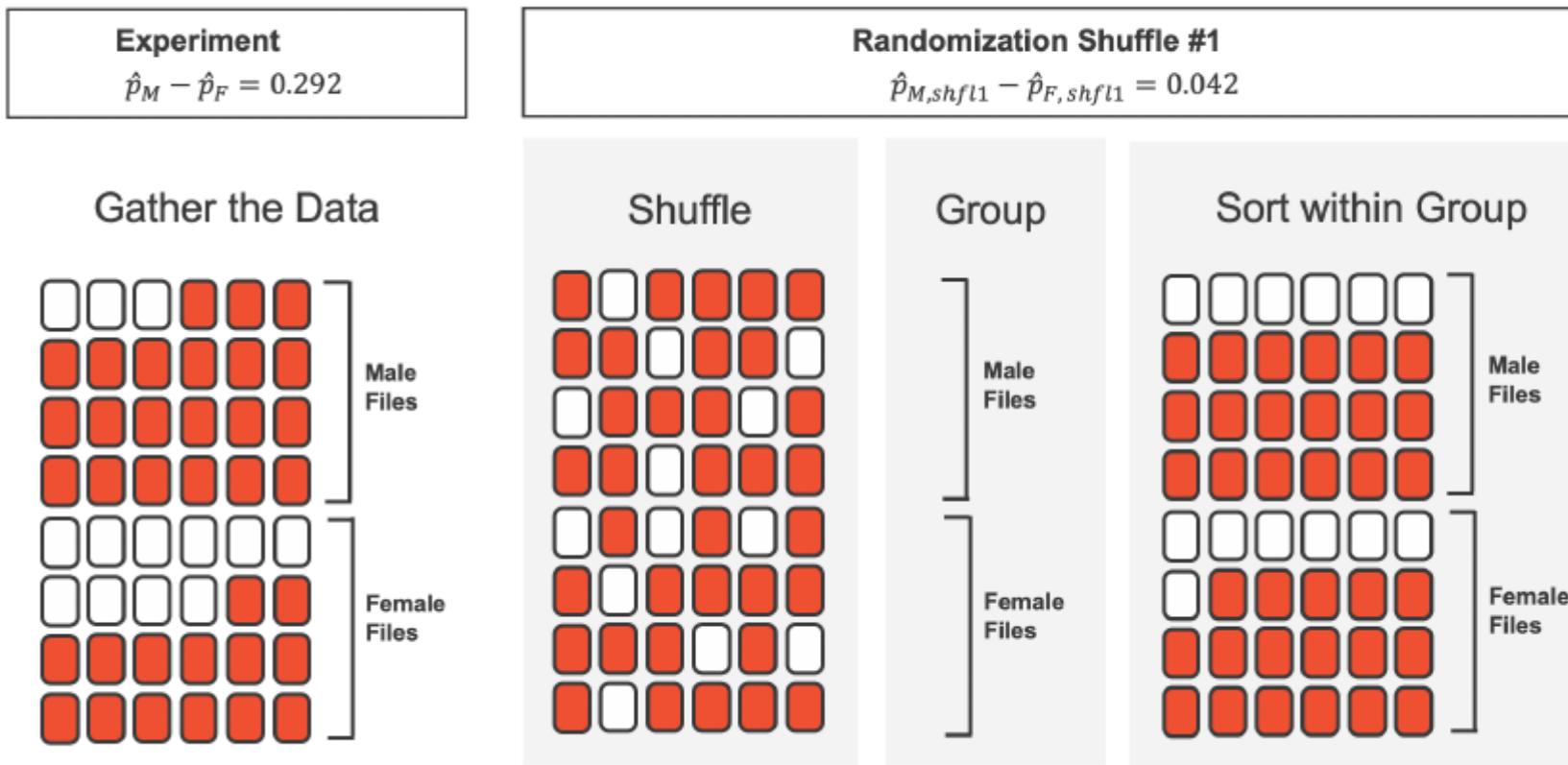
Table 11.1: Summary results for the sex discrimination study.

sex	decision		Total
	promoted	not promoted	
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

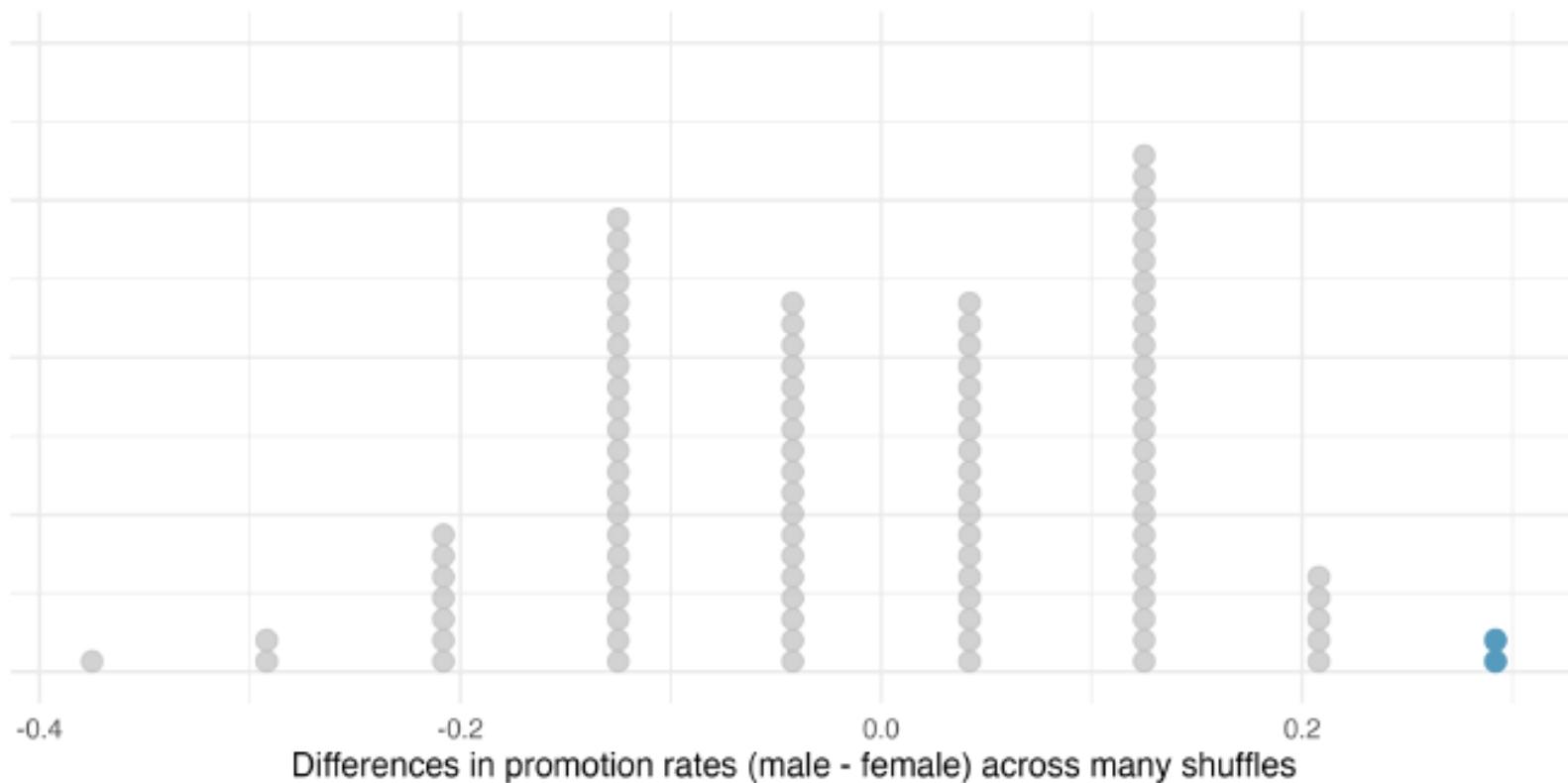
Take sex\_discrimination data as an example



# Permutation test

Take sex\_discrimination data as an example

Experiment	Randomization Shuffle #1
$\hat{p}_M - \hat{p}_F = 0.292$	$\hat{p}_{M,shfl1} - \hat{p}_{F,shfl1} = 0.042$



Observed statistic  
vs.  
null statistics

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

# Permutation test

## Pros:

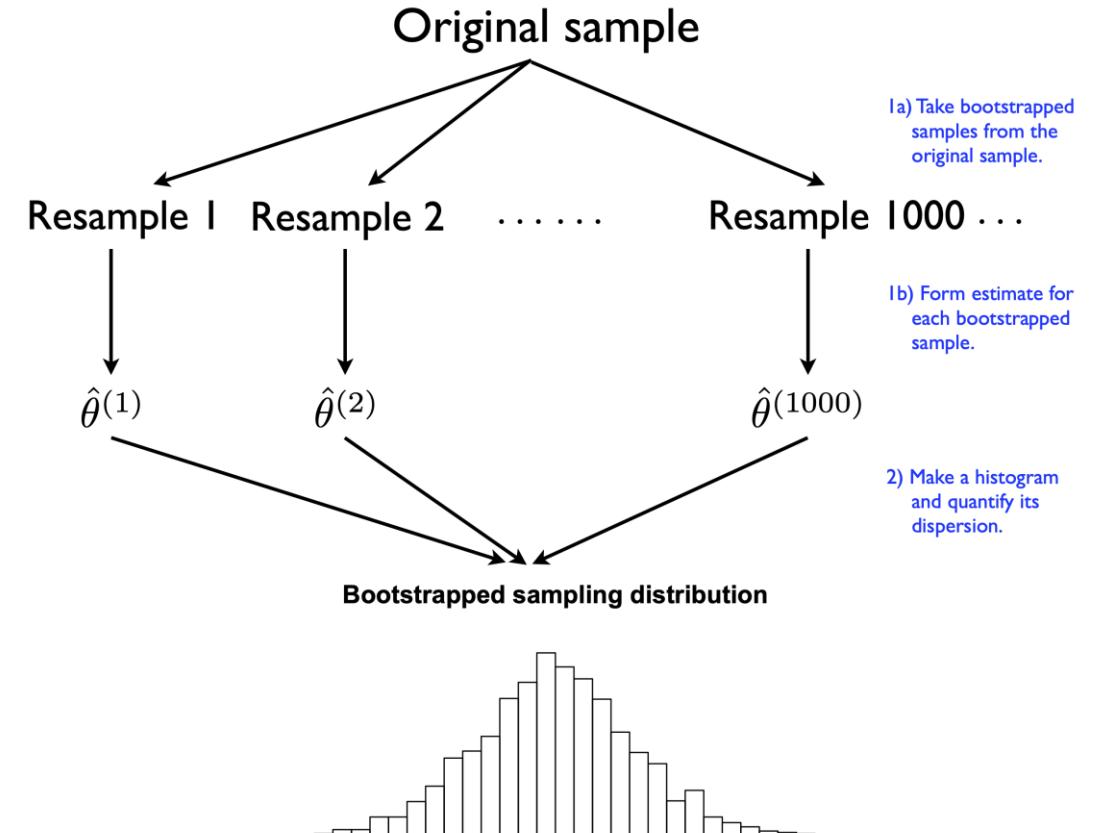
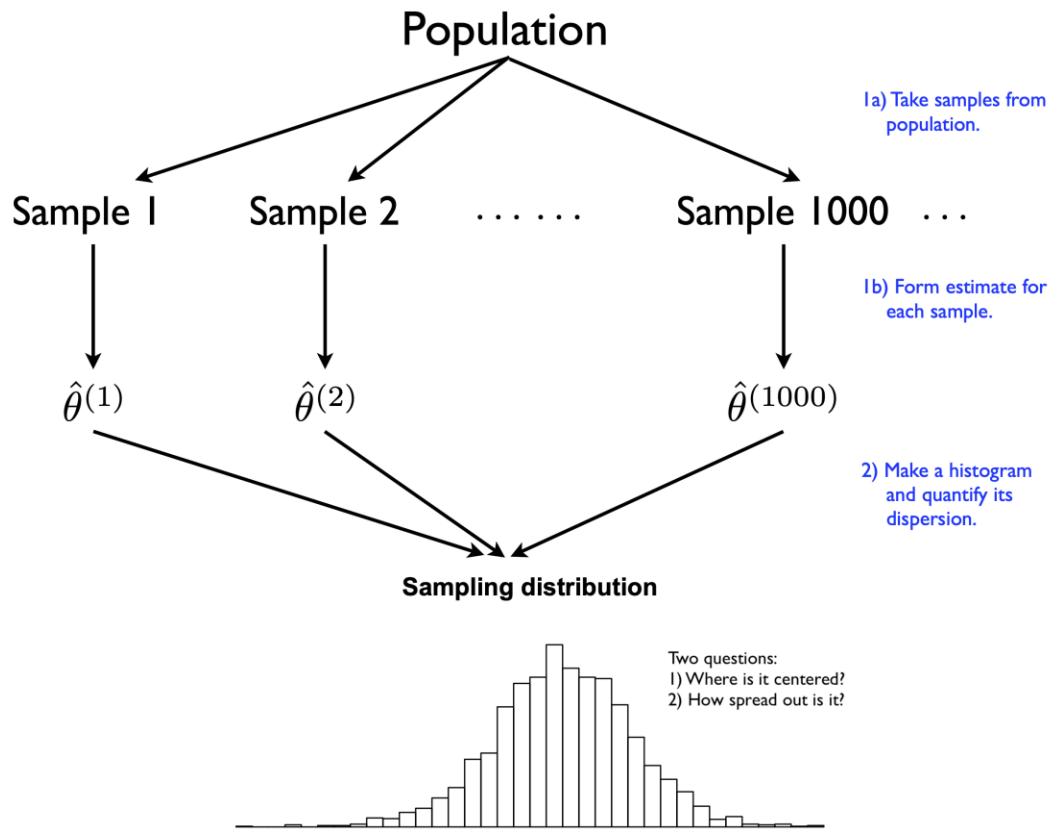
- ❑ **Distribution-free:** No assumptions about data distribution (non-parametric); Works well with skewed or non-normal data.
- ❑ **Flexible:** Applicable to a wide range of test statistics (mean, median, correlation, etc.).
- ❑ **Easy to implement:** Simple concept based on resampling.

## 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

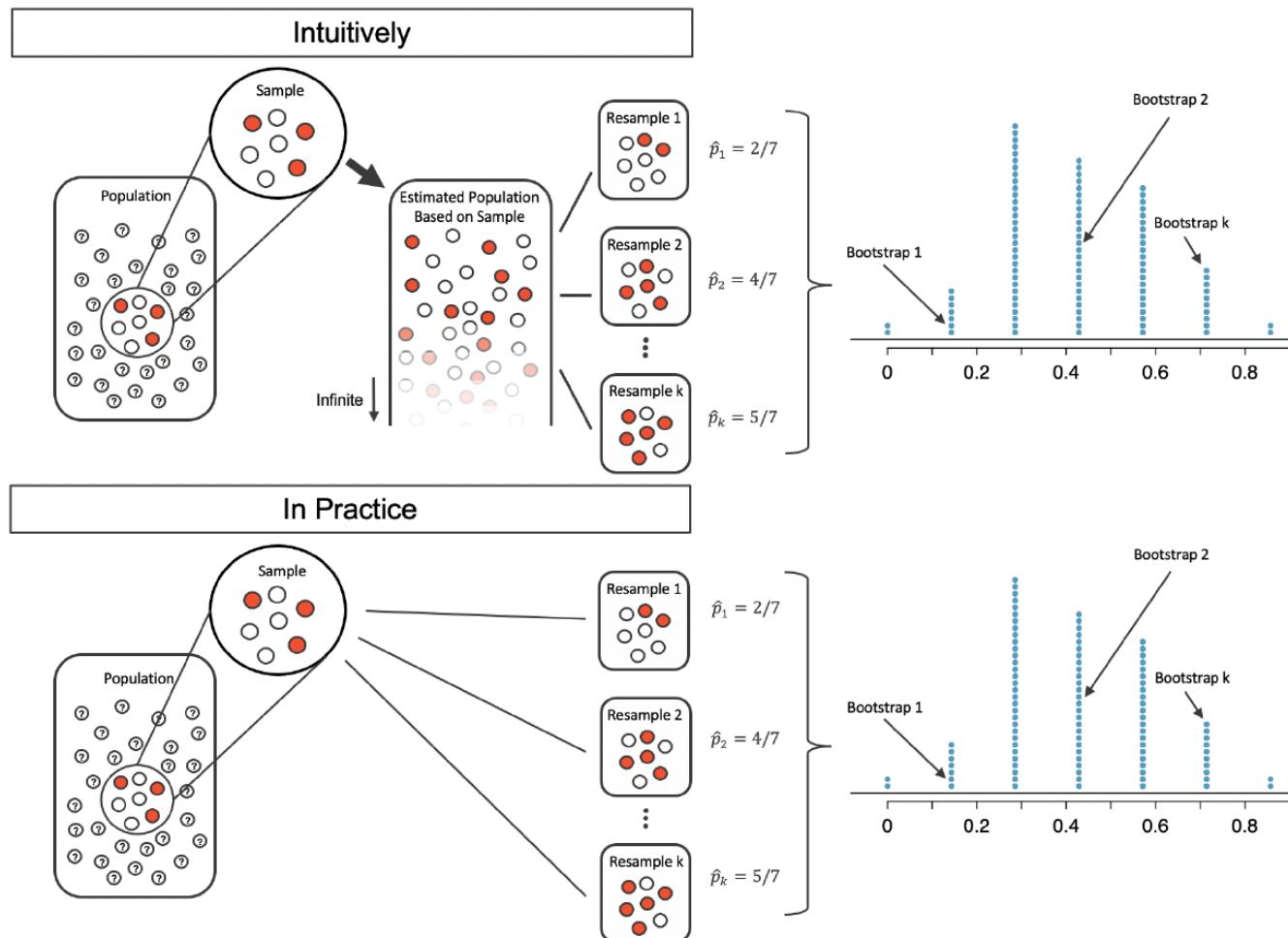
# Bootstrap

## Resampling with replacement



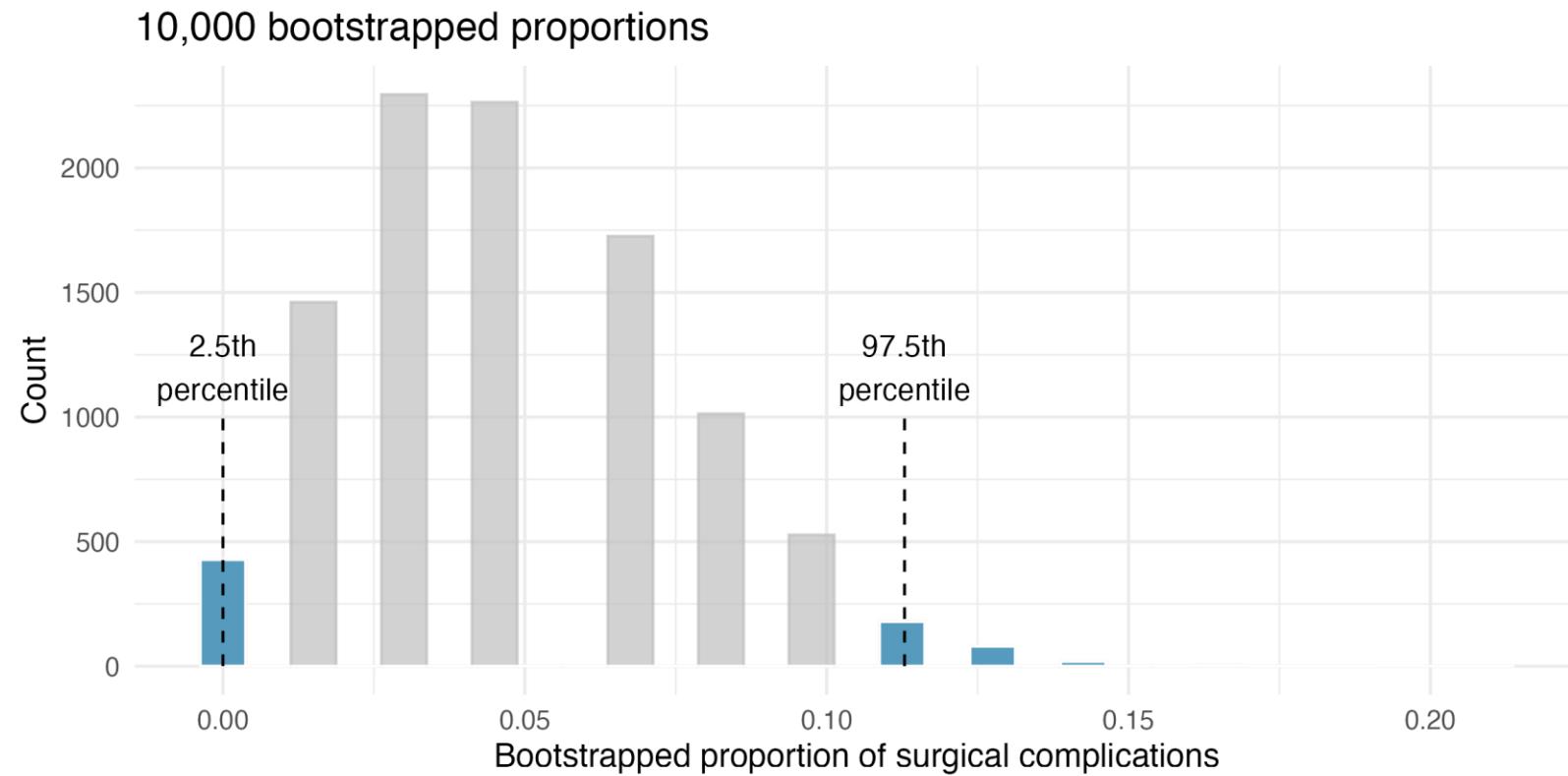
# Bootstrap

## One-sample bootstrap



# Bootstrap

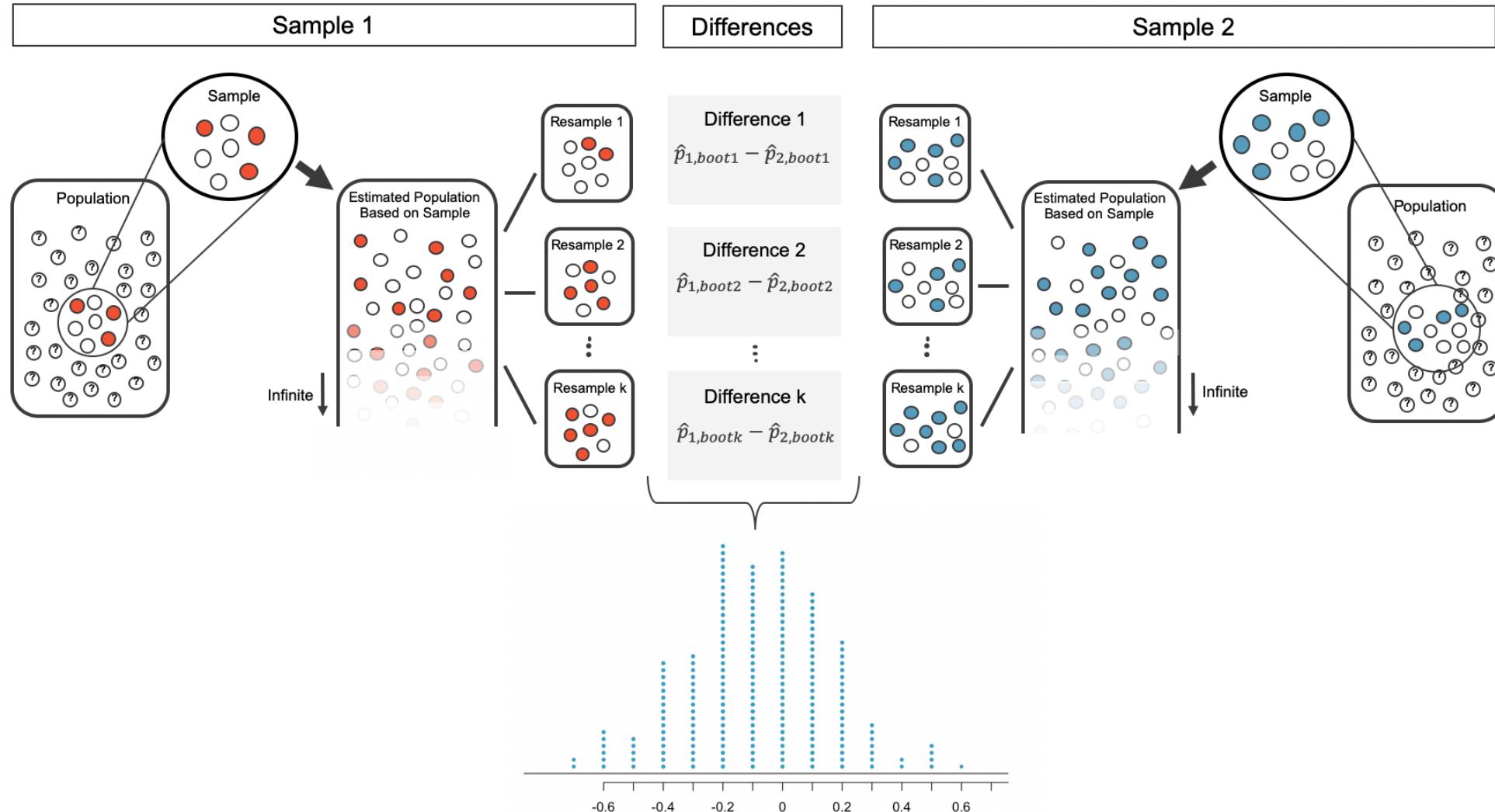
## Confidence interval with bootstrapping



# Bootstrap



## Two-sample bootstrap





FALSE DISCOVERIES



imgflip.com

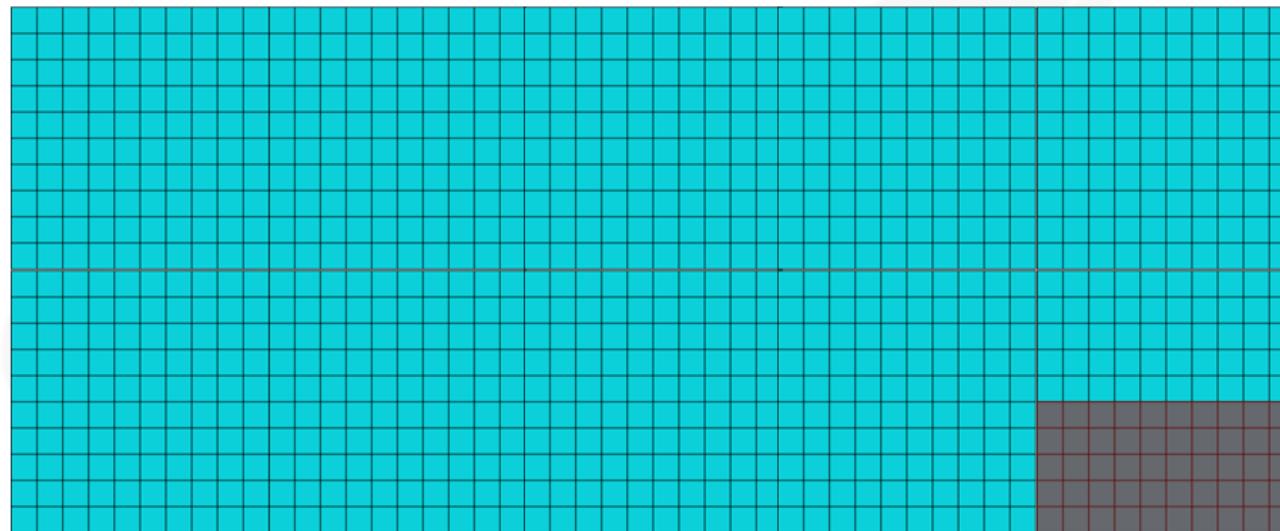
# Multiple testing correction

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 uniformly distributed under the null,  
small  $p$ -values can arise due to chance

# False discovery rate

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

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

All significant genes  
( $p\text{-value} < 0.05$ )

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

# 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

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



Significance threshold:

p-value < 0.05

Association with the beach

- Significant
- Not significant
- False positive

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

# *p*-value adjustment method

- Bonferroni correction:      reject  $H_0$  if  $p_i \leq \frac{\alpha}{m}$
- Benjamini-Hochberg procedure (one of the most cited stats papers, ~114k)

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

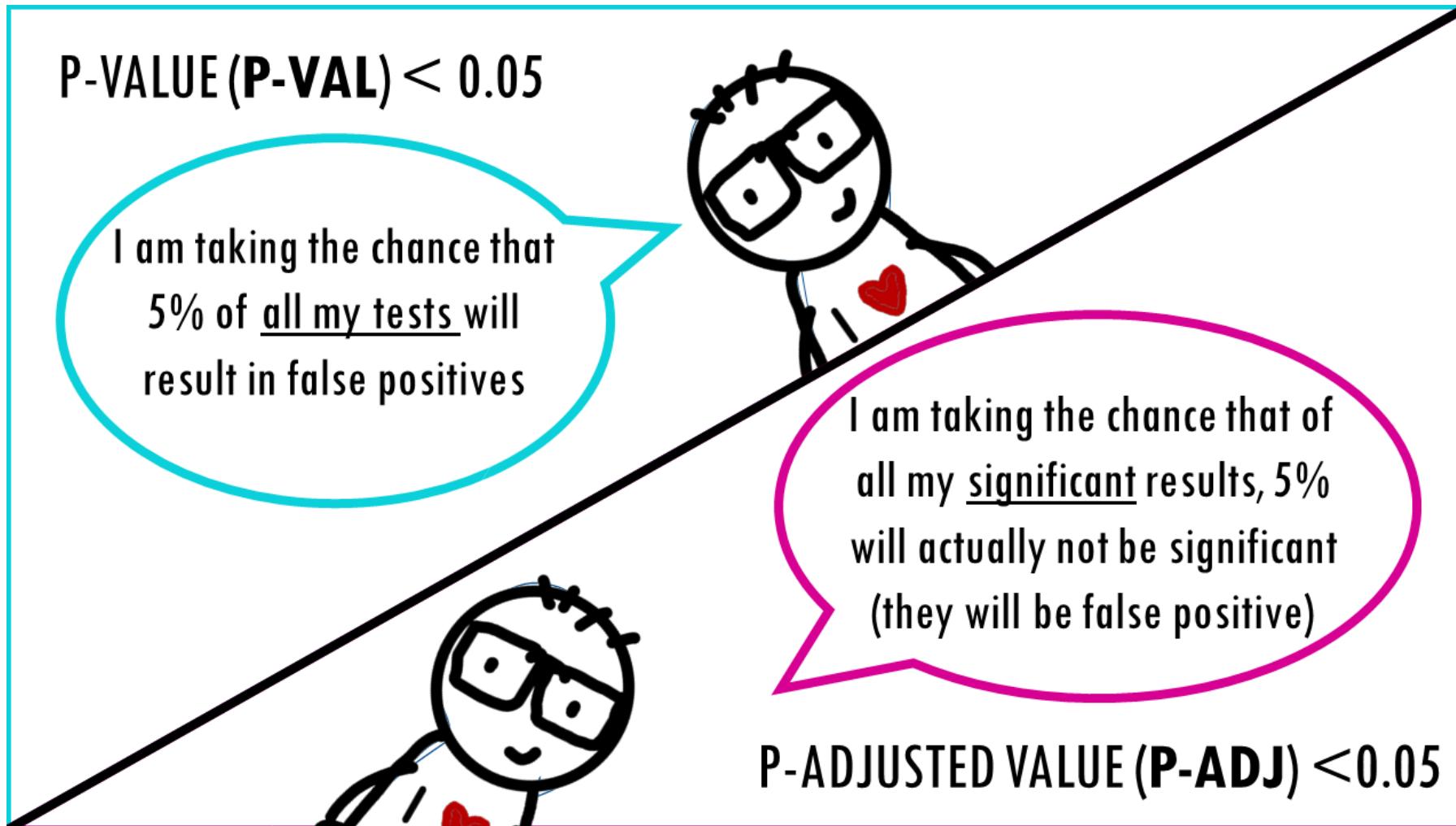
$$p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(m-1)} \leq p_{(m)}$$

and denote their corresponding null hypotheses as:

$$H_{(1)}, H_{(2)}, \dots, H_{(m-1)}, H_{(m)}$$

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 between $p < 0.05$ and $p.\text{adj} < 0.05$



# 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