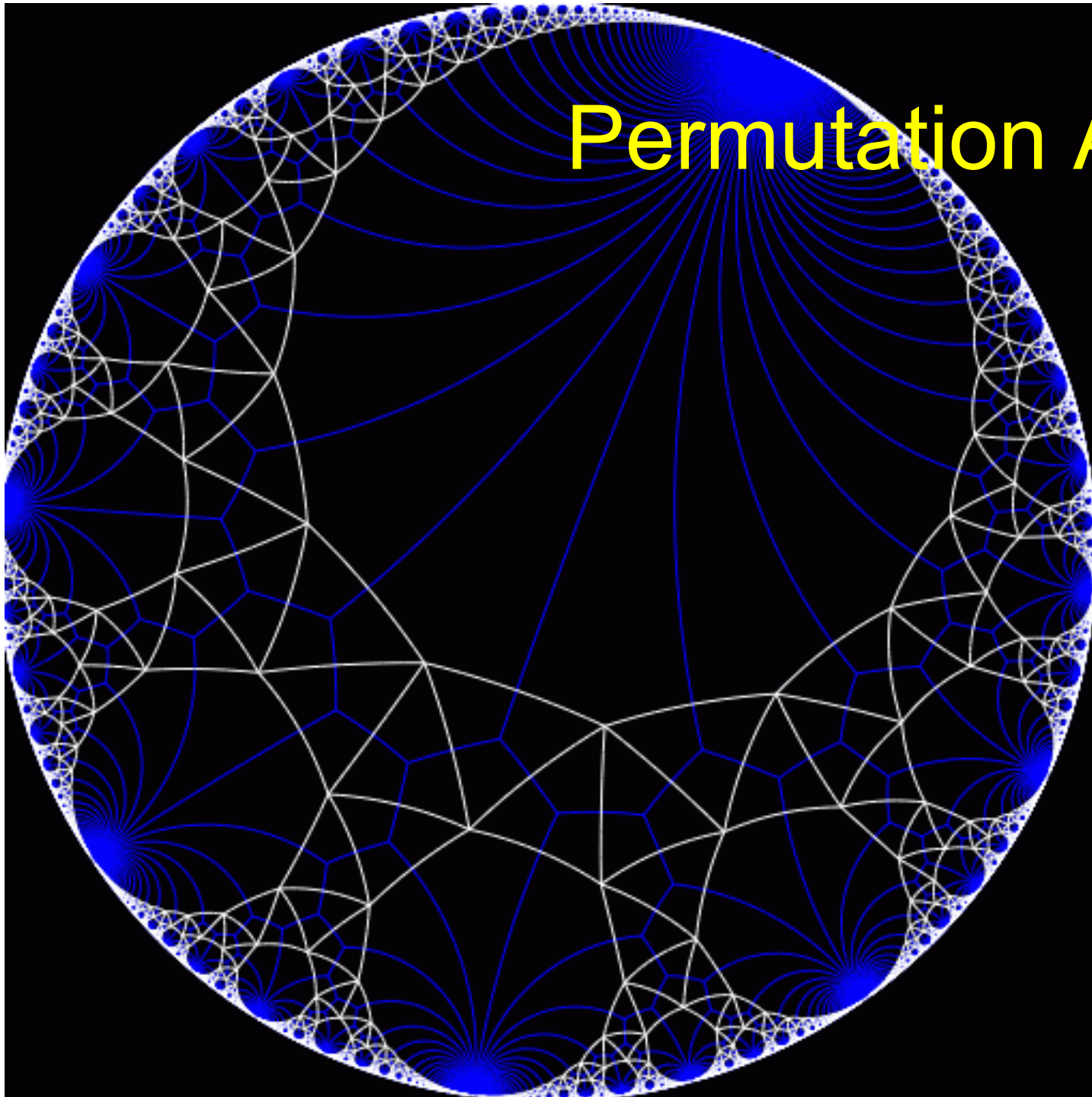


Permutation Analysis

Benjamin
Neale,
Michael
Neale,
Manuel
Ferreira



Who came up with permutation?

- Hint: it's a statistical tool
- R. A. Fisher
- Proposed as validation for Student's t-test in 1935 in Fisher's *The Design of Experiments*

Basic Principle

1. Under the null, all data comes from the same distribution
2. We calculate our statistic, such as mean difference
3. We then shuffle the data with respect to group and recalculate the statistic (mean difference)
4. Repeat step 3 multiple times
5. Find out where our statistic lies in comparison to the null distribution

Real Example

- Case-Control data, and we want to find out if there is a mean difference

	case	control
1	-0.49274	10 1.471227
2	-0.30228	11 0.612679
3	0.093007	12 -0.47886
4	0.715722	13 0.746045
5	1.272872	14 0.871994
6	-1.37599	15 0.985237
7	-0.14798	16 -0.44421
8	-1.22195	17 0.246393
Mean difference .541	9 1.2812	18 0.68246
	Mean -0.01979	0.52144

Permutation One

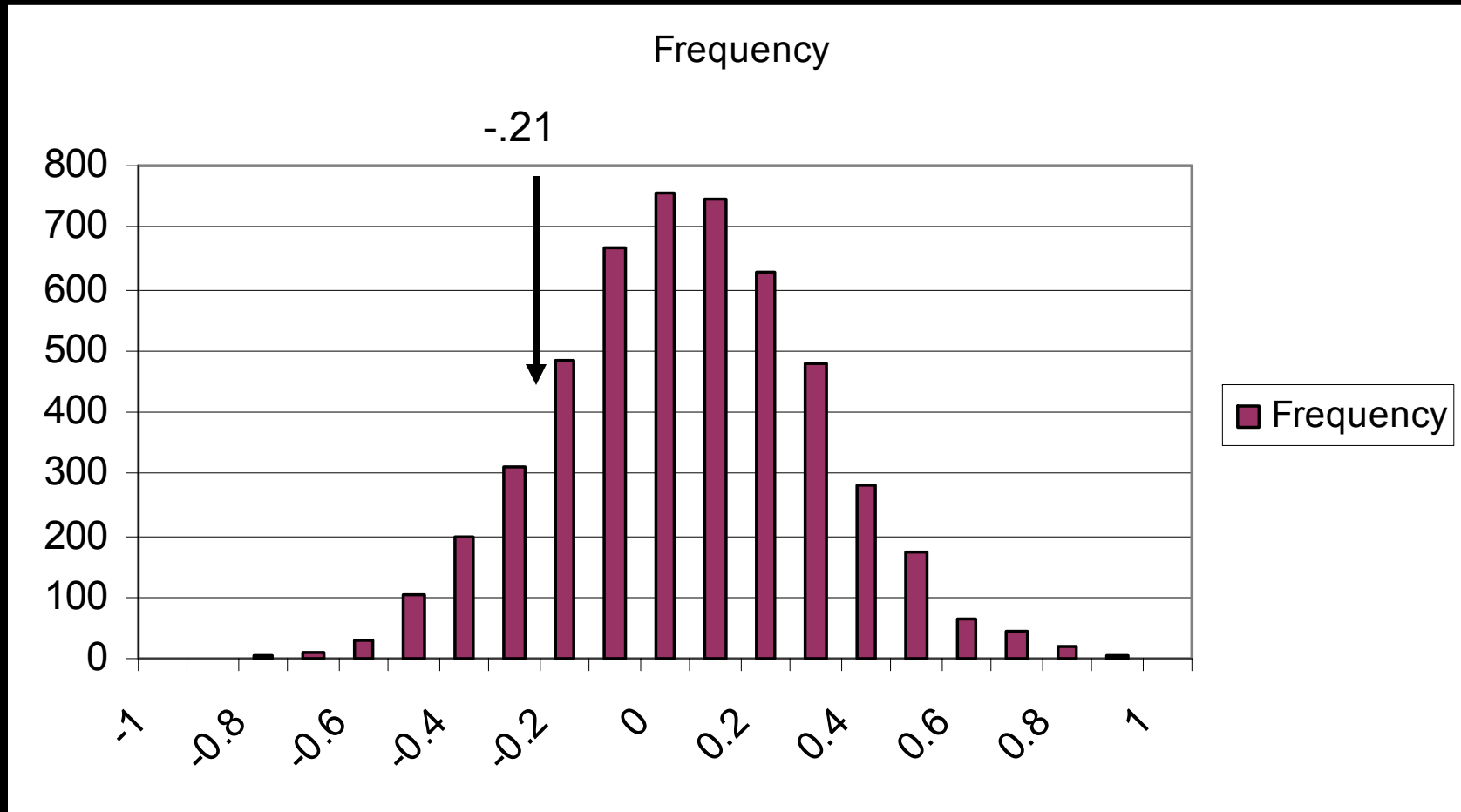
	case		control
9	1.2812	11	0.612679
3	0.093007	18	0.68246
17	0.246393	14	0.871994
15	0.985237	4	0.715722
16	-0.44421	6	-1.37599
1	-0.49274	2	-0.30228
7	-0.14798	5	1.272872
10	1.471227	12	-0.47886
13	0.746045	8	-1.22195
Mean	0.415354		0.086295

Mean difference = .329

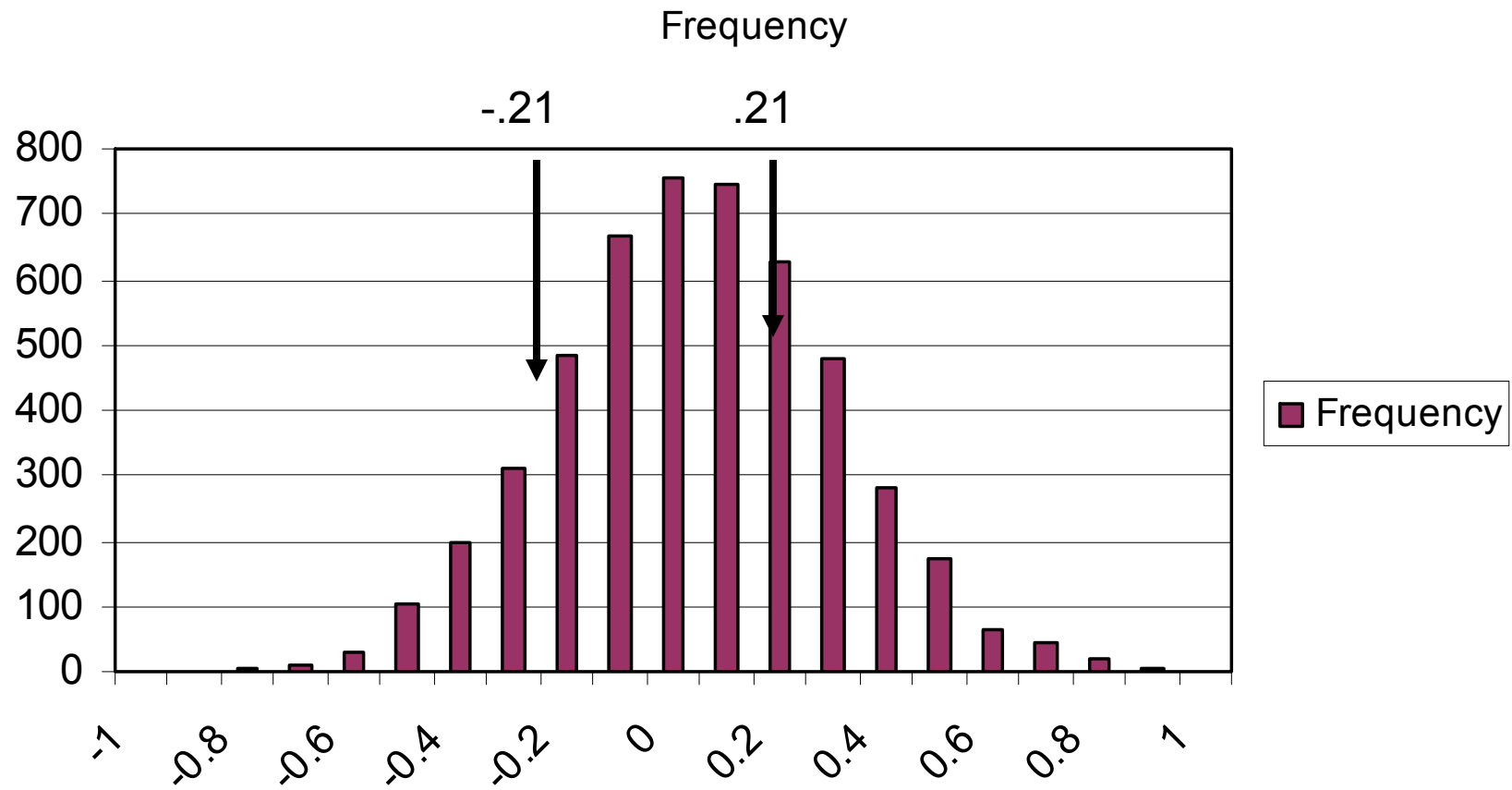
Simulation example

- I simulated 70 data points from a single distribution—35 cases and 35 controls
- Mean difference of -.21
- I then permuted randomly assigning case or control status
- Empirical significance= $\frac{\#hits}{\#permutations}$

Distribution of mean differences from permutations



Distribution of mean differences from permutations



Empirical Significance

- #hits is any permuted dataset that had a mean difference $>.21$ or $<-.21$
- #permutations is the trials permuted datasets we generate
- $\text{Result}(\#hits/\#permutations) = 2024/5000 = .4048$
- T test results = $.3672$

General principles of permutation

- Disrupt the relationship being tested
 - Mean difference between group: switch groups
 - Test for linkage in siblings: randomly reassign the ibd sharing
 - If matched control then within pair permute
 - Shaun will describe tests for association

General advantages

- Does not rely on distributional assumptions
- Corrects for hidden selection
- Corrects for hidden correlation

How would we do QTL permutation in Mx?

1. We analyze our real data and record χ^2
2. For sibpairs we shuffle the ibd probabilities for each sibpair
3. We reanalyze the data and record the new χ^2
4. We generate a distribution of χ^2 for the permuted sets
5. Place our statistic on the distribution
6. Repeat for all locations in genome

Some caveats

- Computational time can be a challenge
- Determining what to maintain and what to permute
- Variable pedigrees also pose some difficulties
- Need sufficient data for combinations
- Unnecessary when no bias, but no cost to doing it

Some Exercises

- What would we permute to determine if the MZ correlation is equal to the DZ correlation in a univariate analysis?
- What would we permute to determine if QTL dominance is significant in linkage analysis