Hypothesis testing in public health

Seth D. Temple
Statistics Ph.D. candidate
UW Biostatistics Teaching Demo
January 18, 2024

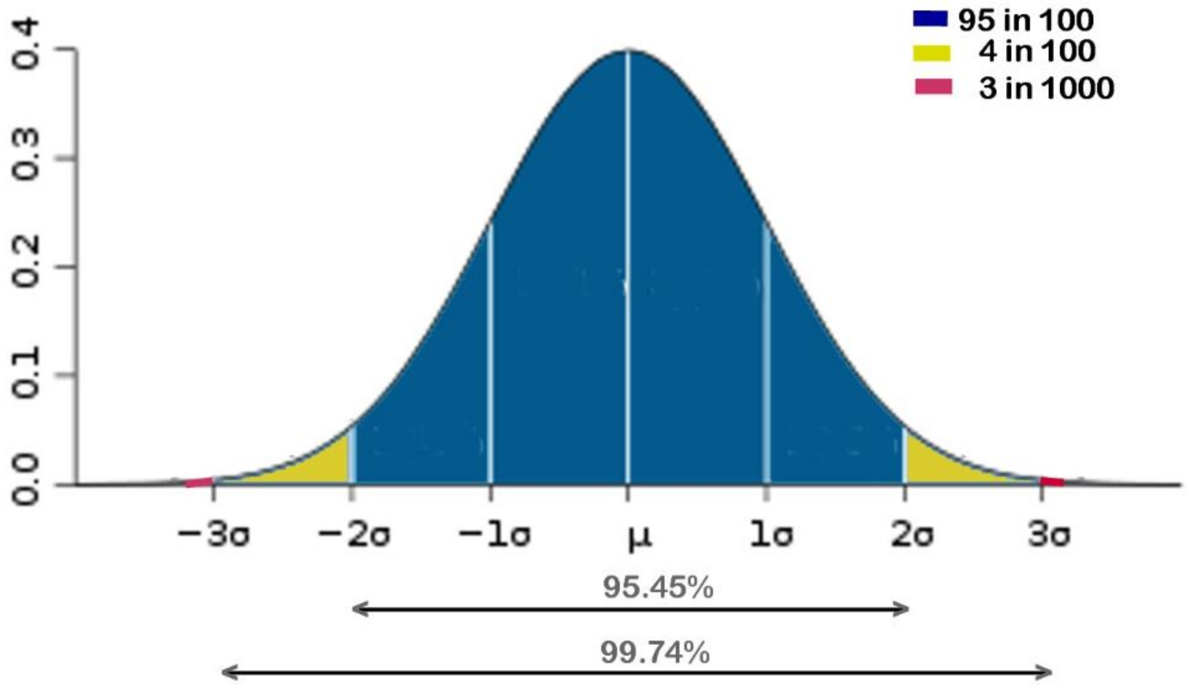
Last class

Parameter:

- Feature of a population
- Model for data generation

Statistic:

- Feature of a sample
- Estimate parameter from data



Learning objectives

- Frame a research question
  as statistical question
- Define Type 1 and Type 2 Error
- Discuss their importance
  in research study

* Not graded

Participation* expectations

- Raise your hand to ask me for clarification

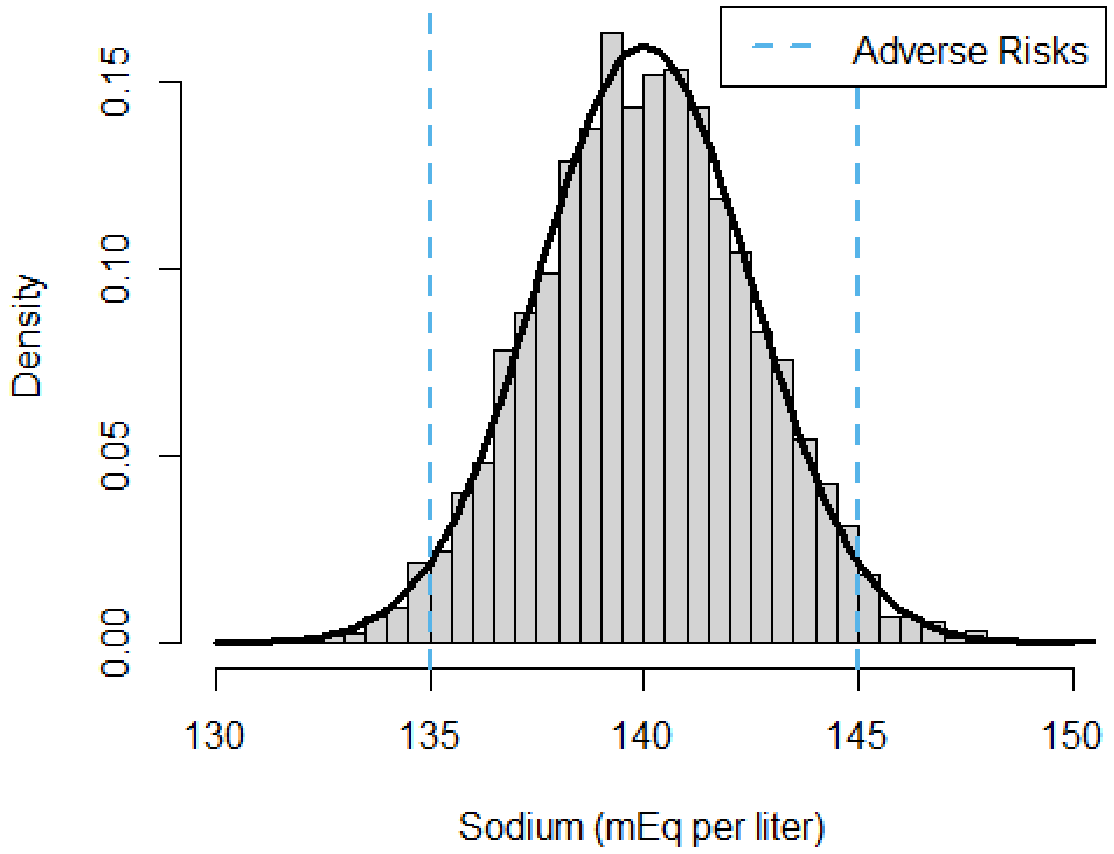
- Exercise 1
  1. Chatter in small group, or write response on notes (~30 sec.)
  2. Solicit responses for larger discussion

- Exercise 2
  - Think about your response privately (5 sec.)
  - Quick “Yes/No” poll after countdown

- Show respect; no judgment

Example 1: Sodium homeostasis

- Sodium is important to biological processes
- Too little may lead to health consequences
- Lamotrigine helps treat bipolar disorder. It may affect sodium levels.



Research Question

- Does **lamotrigine** affect the levels of **sodium metabolite levels**?

Statistical Question

- μ_0 is "average" parameter based on prior study
- μ_D is parameter for population taking lamotrigine
- Does the **sample mean** $\bar{X}_D$ sodium give enough evidence to reject that $\boldsymbol{\mu_D = \mu_0}$?

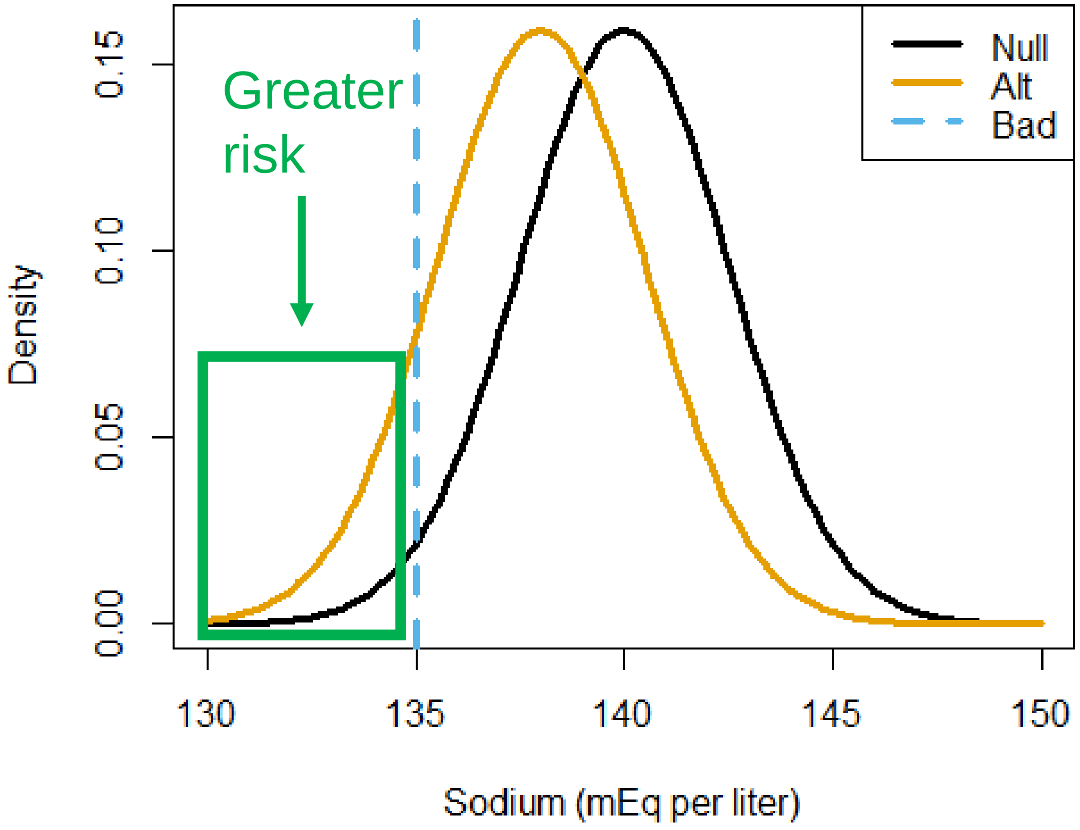
Why do we care about hypothesis tests?

Reward

- Prescribe patients an effective treatment

Risk

- Do harm to patients



Type 1 Error. We reject that $\mu_D = 140$ even though it is.

* This definition generalizes to other hypothesis tests.

Type 2 Error. We fail to reject $\mu_D = 140$ given that $\mu_D = 138$.

* This definition generalizes to other hypothesis tests, say $\mu_D \neq 140$.

(Have you taken a COVID test before?)

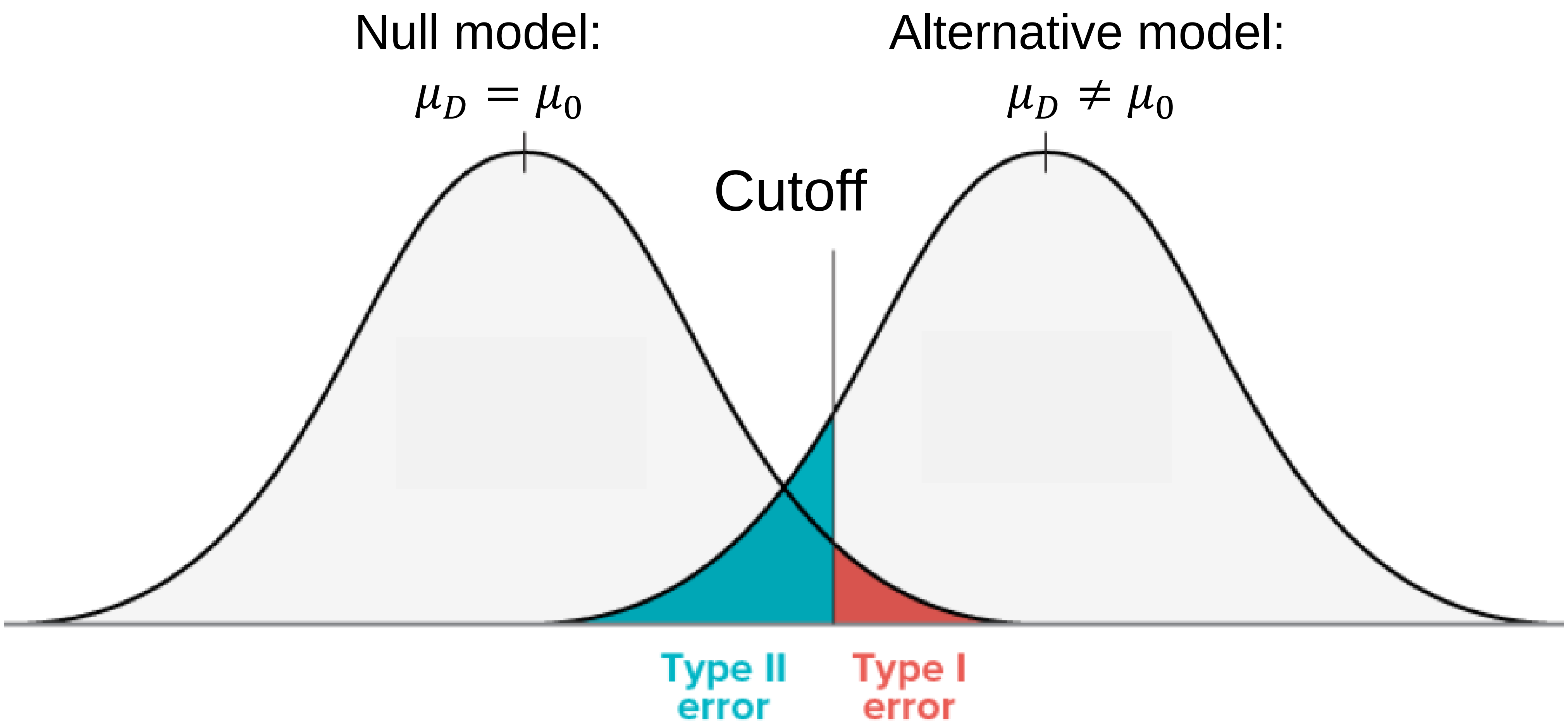
Warm-up viral tests

		Actual	
		+	-
Predicted / Concluded	+	True positive	False positive
	-	False negative	True negative

Type 1 error

Type 2 error

Visualizing Type I,II errors

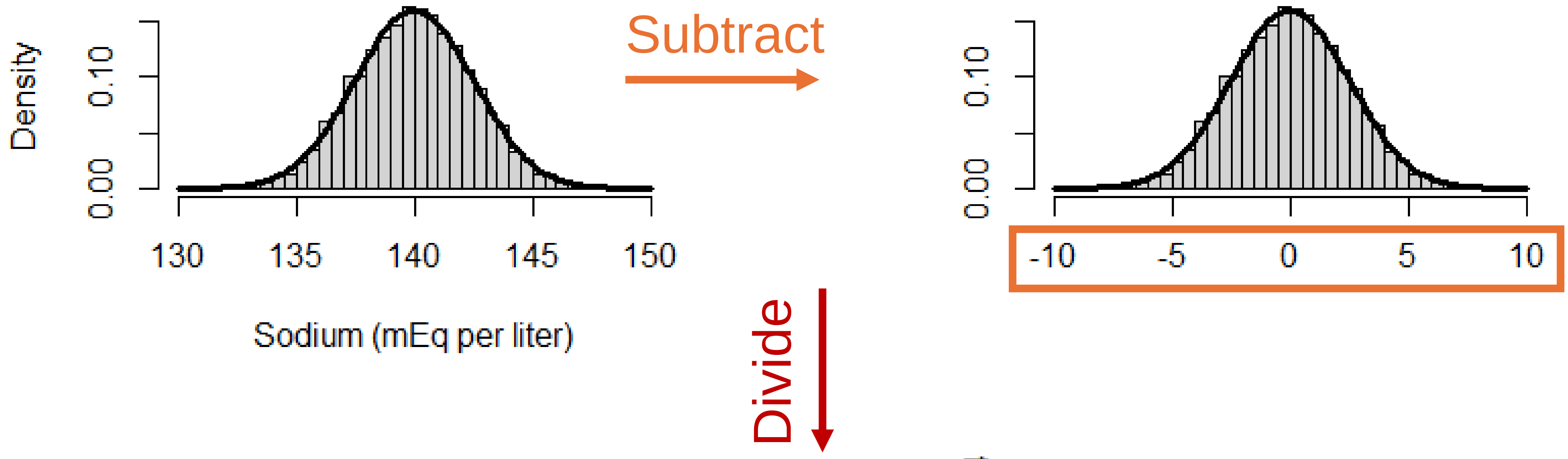


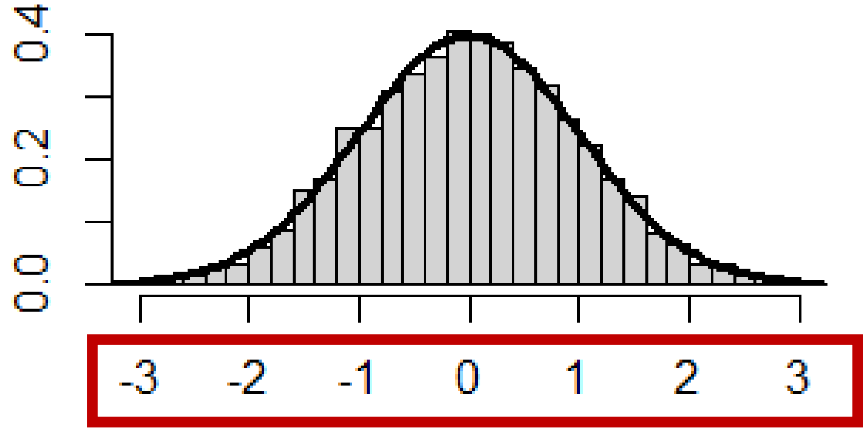
Hypothesis test calculation

- Collect blood work from 30 individuals
- $X_1, \ldots, X_{30}$ are their sodium levels
- Compute the mean

$$\bar{X}_D = \frac{X_1 + \cdots + X_{30}}{30}$$

$*** \; X_D$ is statistic estimating $\mu_0 \; ***$





Put the statistic on a common scale

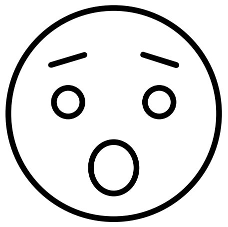
Hypothesis test calculation

Sodium levels $X_1, \ldots, X_{30}$ from patients taking lamotrigine

$$\bar{\boldsymbol{X}}_{\boldsymbol{D}} = \mathbf{138.42}$$

$$\boldsymbol{Z}_{\boldsymbol{D}} = \sqrt{\mathbf{30}} \times \frac{(\bar{\boldsymbol{X}}_{\boldsymbol{D}} - \mathbf{140})}{\mathbf{2.5}} = -\mathbf{3.46}$$

p-value: probability mean as extreme is 0.0005 !!!



p-value: the probability of observing a statistic as extreme as in our sample

Here our statistic is the mean

Under the null model

Type 1 error simulations

- Assume model $N(\mu_0 = 140,\, \sigma = 2.5)$
- Run 1000 experiments
    1. Simulate 30 samples from model
    2. Compute Z_D statistic
    3. If $|Z_D| > 1.96$, add 1 to running count

I found $|Z_D| > 1.96$ to occur $\frac{53}{1000} \approx 0.05$.

Discussion activity (60 sec.)

Study effect of lamotrigine on **# of mood swings**

- 50 participants in Drug D group
- 50 participants in Placebo P group

1. How would you phrase the statistical question in this study?
2. How would you phrase the Type 1 error?

My answer key

1. Does the difference of sample means $\boldsymbol{\bar{X}_D - \bar{X}_P}$ give enough evidence to reject that $\boldsymbol{\mu_D - \mu_P = 0}$?

2. We reject initial assumption $\mu_D = \mu_P$ despite $\mu_D = \mu_P$.

Back to sodium homeostasis example

Alternative Hypothesis Experiment

- Run an experiment 2000 times
  1. Simulate 30 samples from model Normal(μ_D, $\sigma = 2.5$)
  2. Compute $Z_D = \sqrt{30} \times (\bar{X}_D - 140)/2.5$
  3. If $|Z_D| > 1.96$, add 1 to a running count

True Parameter	Type 2 Error	Power = 1 - Type 2
139.75	0.91	0.09
139.50	0.81	0.19
139.00	0.42	0.58
138.00	0.01	0.99

Sample Size Experiment

Same as before, except # of people varies

The true drug effect is 139

Sample Size	Type 2 Error
10	0.77
20	0.57
30	0.42
50	0.28

Confidence level experiment

Same as before, except controlling Type 1 error varies

The true drug effect is 139

Confidence	Type 2 Error
0.10	0.29
0.05	0.41
0.01	0.65
0.001	0.86

Why would we want to have lower confidence level if it decreases Type 2 error?

My answer key

Type II error probability depends on specifying tolerance for Type I error.

We may want to make few Type 1 errors.

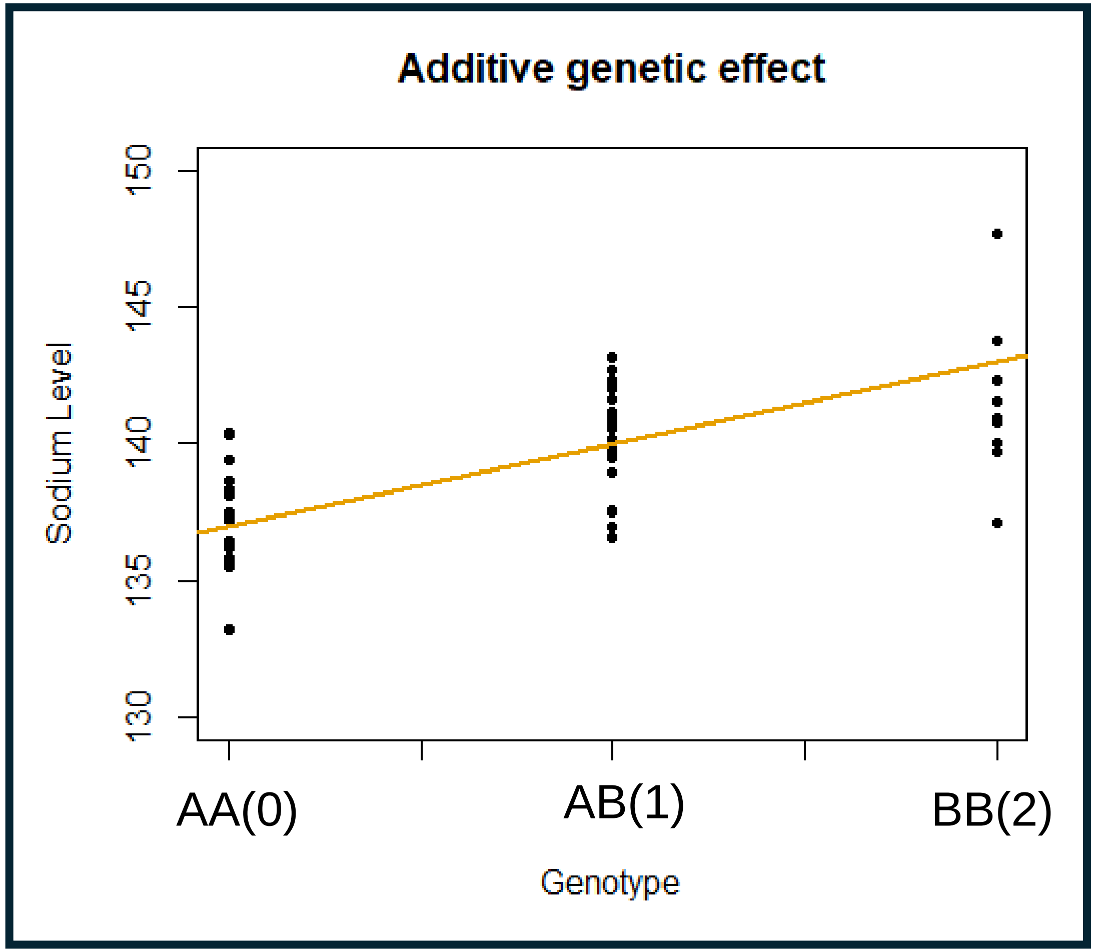
Beyond p-values

- Lamotrigine shows “significant” effect on mood!
  - How effective is significant? 😟
  - “Typical” is based on non-representative study. 😟
  - Prescription costs 😟
  - Studies had few participants ☹
  - Electrolytes are effective supplement ☺

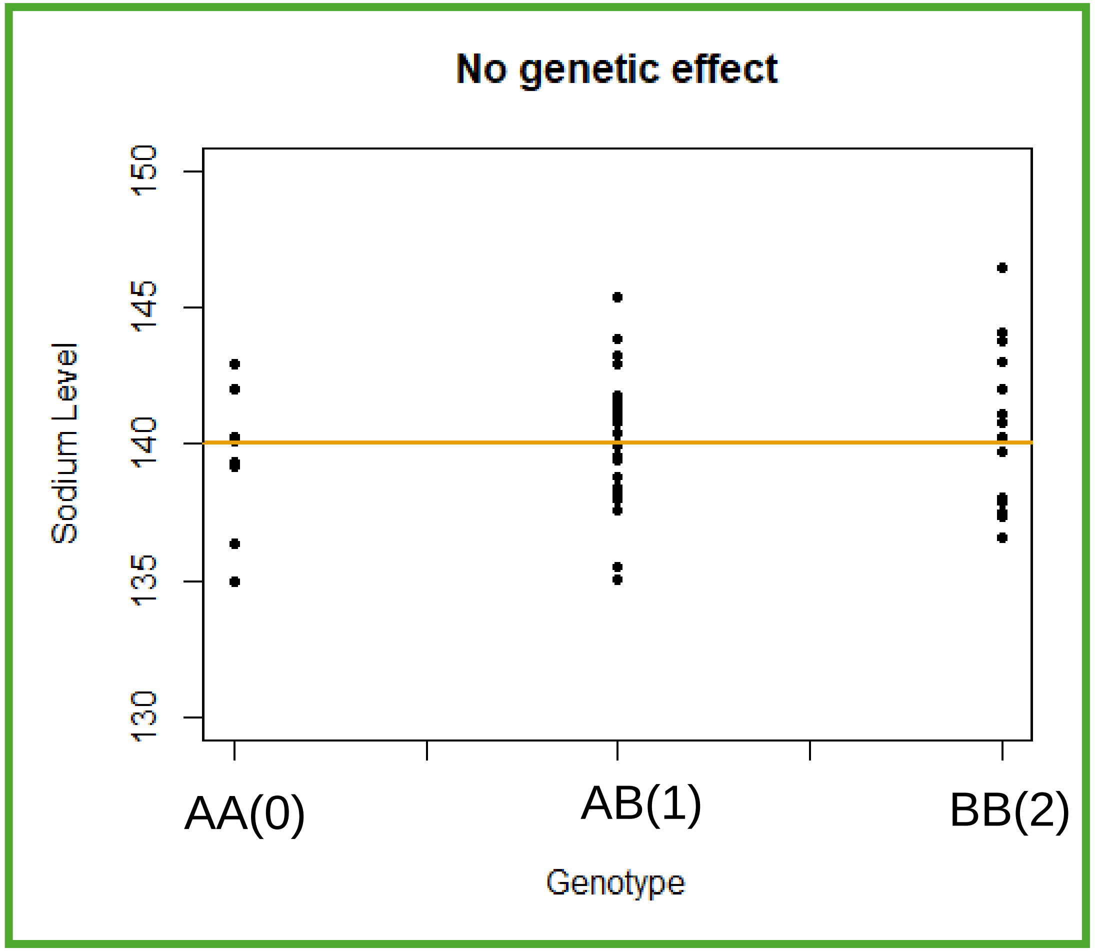
Example 2: Genetic associations

- Genotypes AA (0), AB (1), BB (2)
- $(Y_1, \dots, Y_n)$ sodium levels
- $(X_1, \dots, X_n)$ genotypes
- ϵ_i random noise

$$Y_i = \underbrace{\beta_0}_{\text{Intercept}} + \underbrace{\beta_1}_{\text{Slope}} \times X_i + \epsilon_i$$



Additive genetic effect
Sodium Level
130
135
140
145
150
AA(0)
AB(1)
BB(2)
Genotype



No genetic effect
Sodium Level
130
135
140
145
150
AA(0)
AB(1)
BB(2)
Genotype

Would you reject that slope is zero?

Studying one variant

Simulated no effect model:

$$Y = 140 + 0 \times X + \epsilon$$

Parameter	Estimate	p-value
Slope	-0.09	0.67

Would you reject that slope is zero?

Studying one variant

Simulated additive effect model:

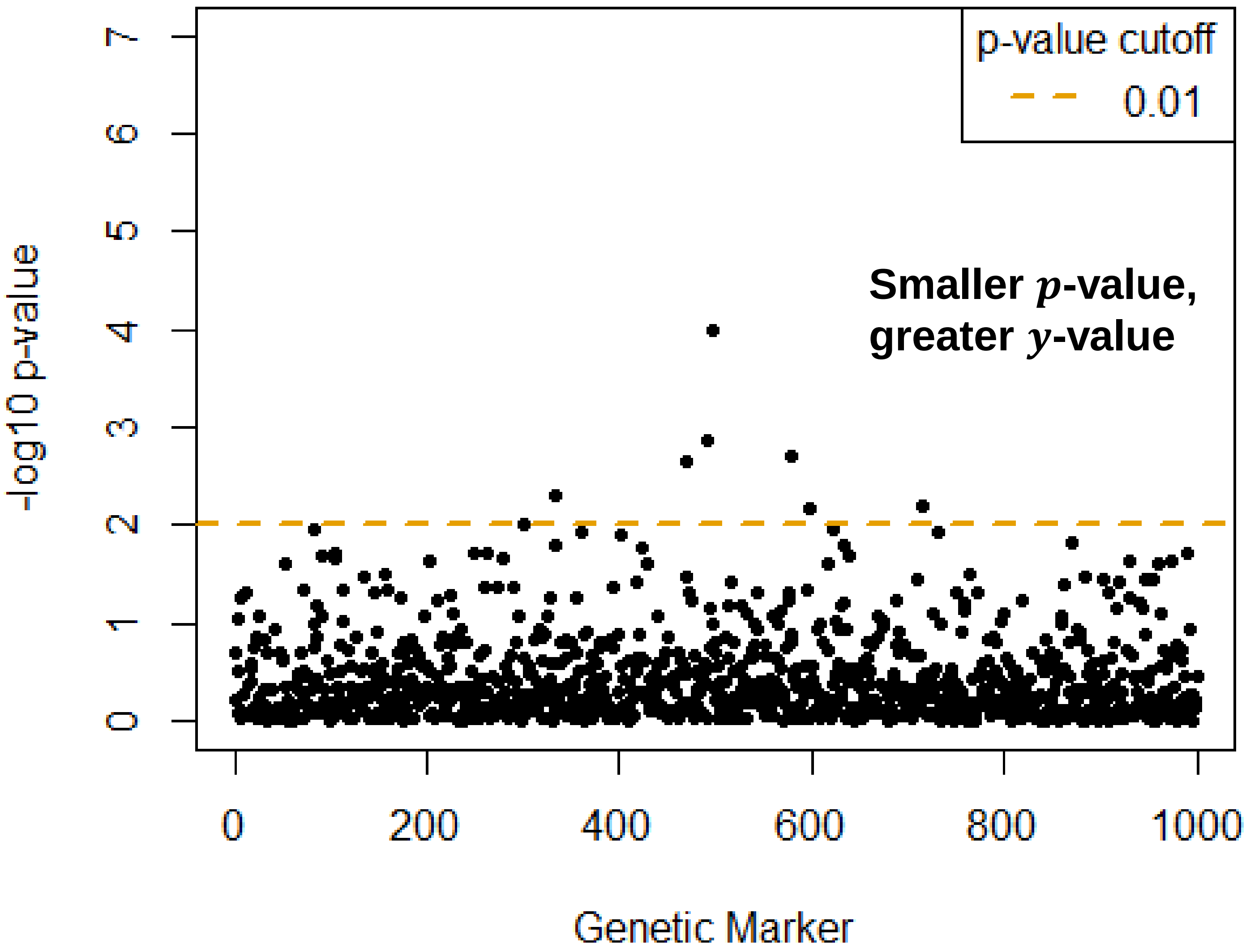
$$Y = 137 + 3 \times X + \epsilon$$

Parameter	Estimate	p-value
Slope	2.91	2e-16

Simulate no genetic effect

1. Draw 16, 17, 17 genotypes AA, AB, BB
2. Draw 50 errors $\epsilon \sim N(\mu = 0, \sigma = 2.5)$
3. Compute Y
    - Intercept $\beta_0 = 140$
    - Slope $\beta_1 = 0$
4. Compute p-value from average $\bar{Y}$

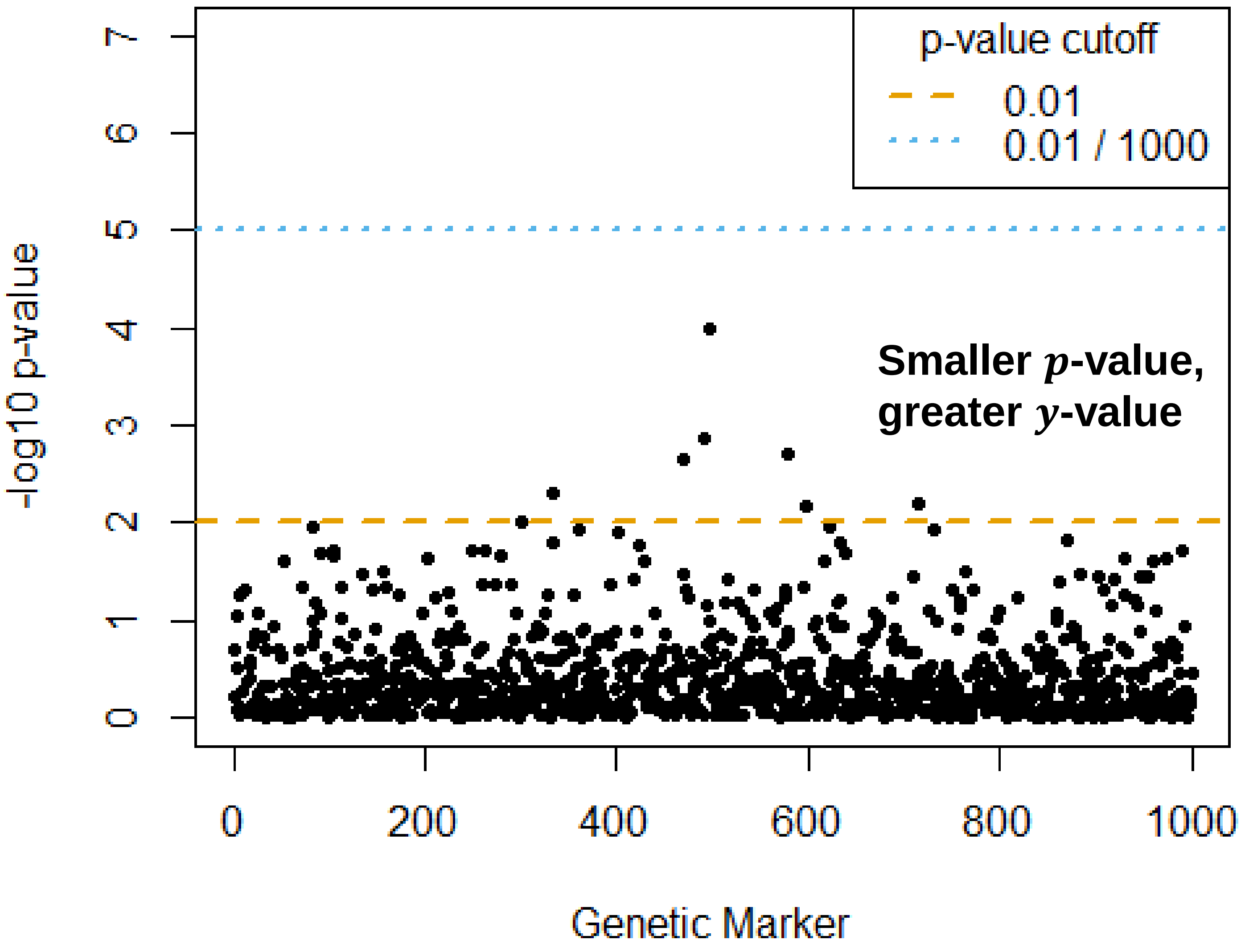
Is sodium metabolite associated w/ genotypes?



Why do we observe **~ 10 / 1000** “significant” associations?

- p-value threshold should be smaller
- Divide by # tests

Is sodium metabolite associated w/ genotypes?



Today's review

- Two examples of statistical analysis
  - Average sodium level (continuous)
  - Genetic association (linear relationship)

- **Type 1 error** is about controlling for false positives
  - Testing too many things
- **Power** is about identifying true positives

Next class: more types of hypothesis tests

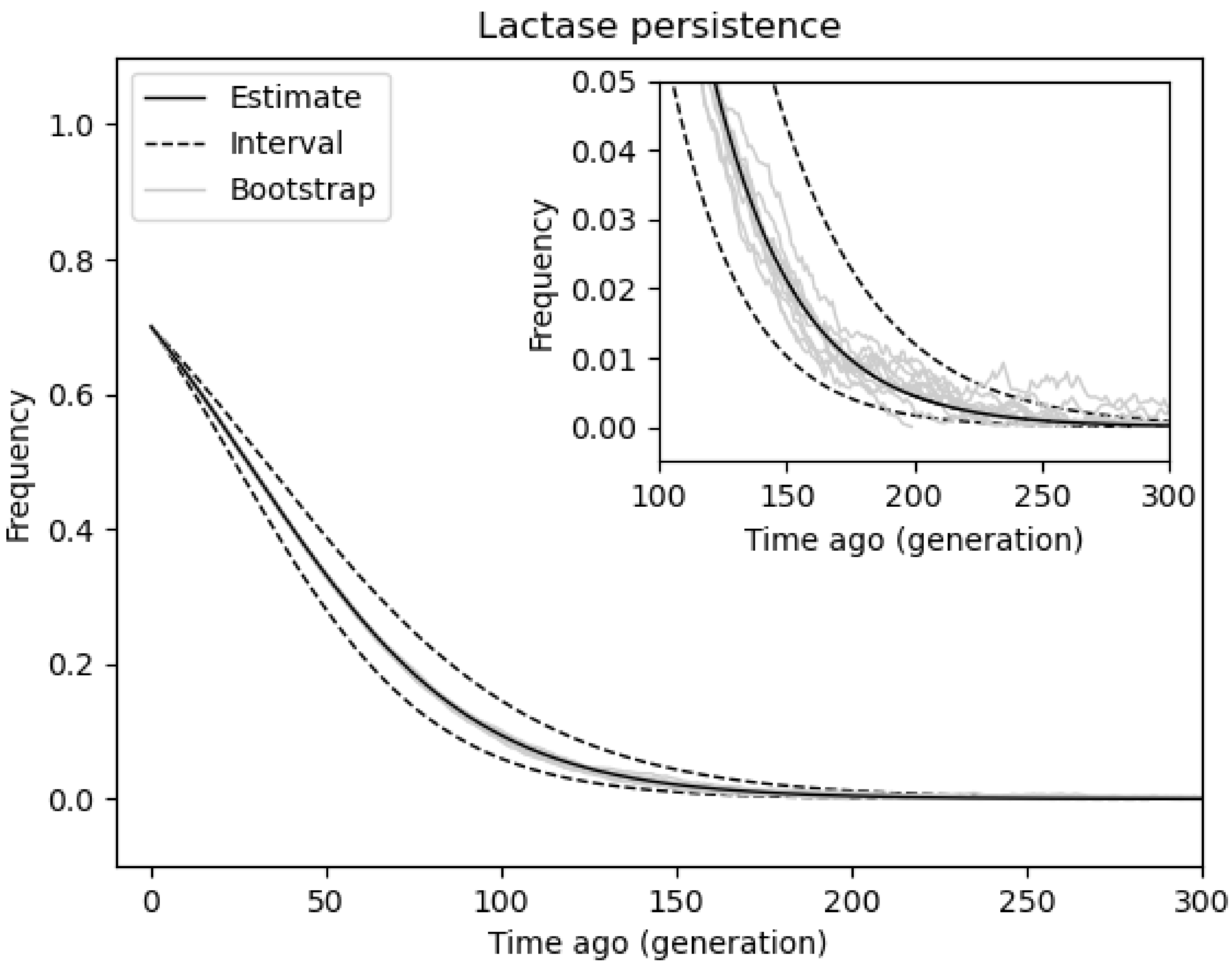
My vision in this role

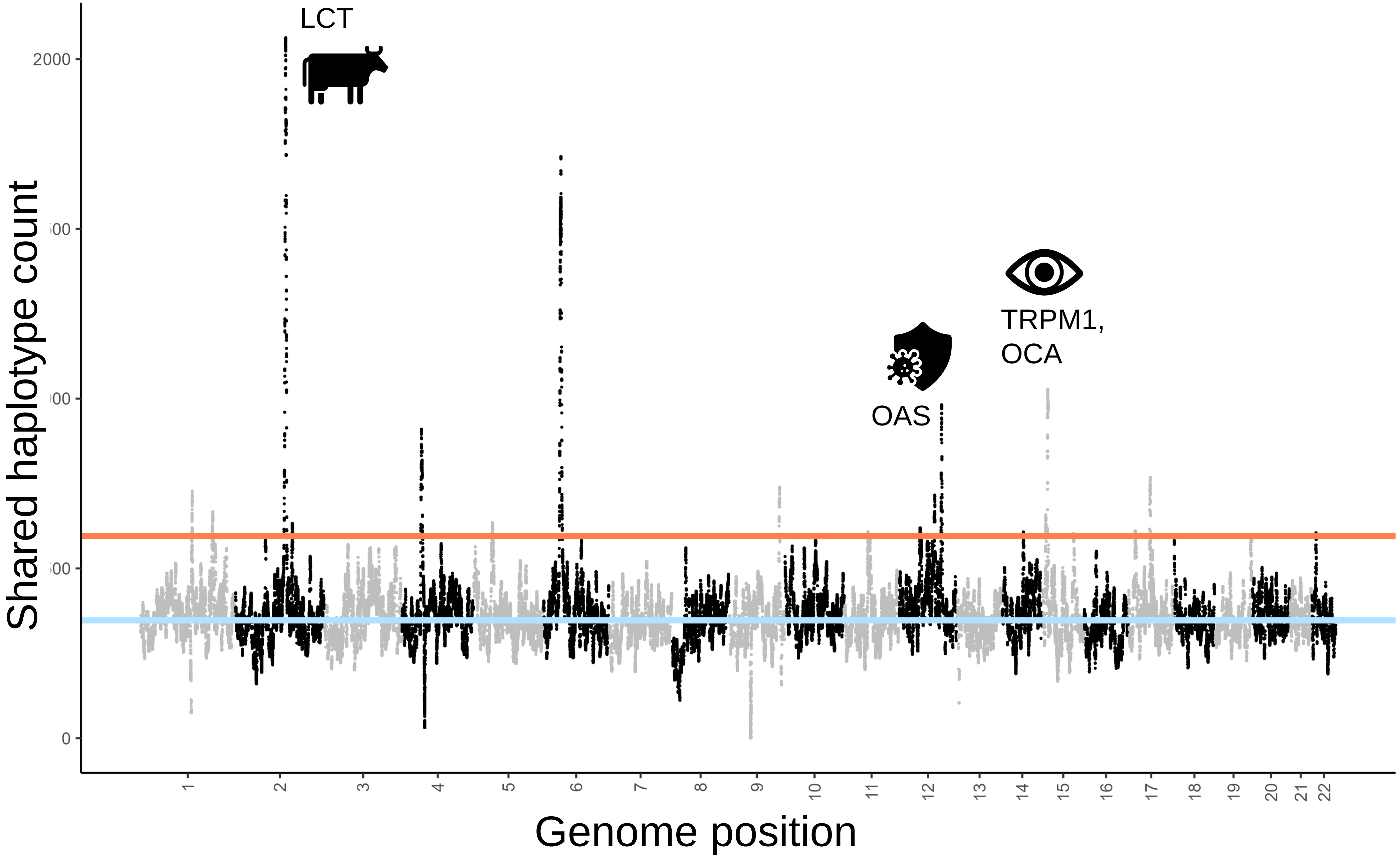
- Today: teaching undergrads in public health
- Able to teach grad courses (512, 513, 533, 570-1, …)
- Passionate about statistical genetics
  - Broadening curriculum beyond human genomics
  - Outreach (workshops, online courses)
  - An undergraduate course
  - Mentoring student research
  - Promoting DEI

sdtemple.github.io/files/teaching-demo-sdtemple.pdf
sdtemple.github.io/files/teaching-demo-sdtemple.R

Thank you for the invite

Appendix: my research


Lactase persistence
Estimate
Interval
Bootstrap
1.0
0.8
0.6
0.4
0.2
0.0
Frequency
0
50
100
150
200
250
300
Time ago (generation)
0.05
0.04
0.03
0.02
0.01
0.00
Frequency
100
150
200
250
300
Time ago (generation)



LCT
OAS
TRPM1,
OCA
Shared haplotype count
2000
0
Genome position
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

Current & future research

- Correlated binary random vector
  in light of unobserved process
- Recent selection in non-European or non-human studies
- Deep learning in population genetics