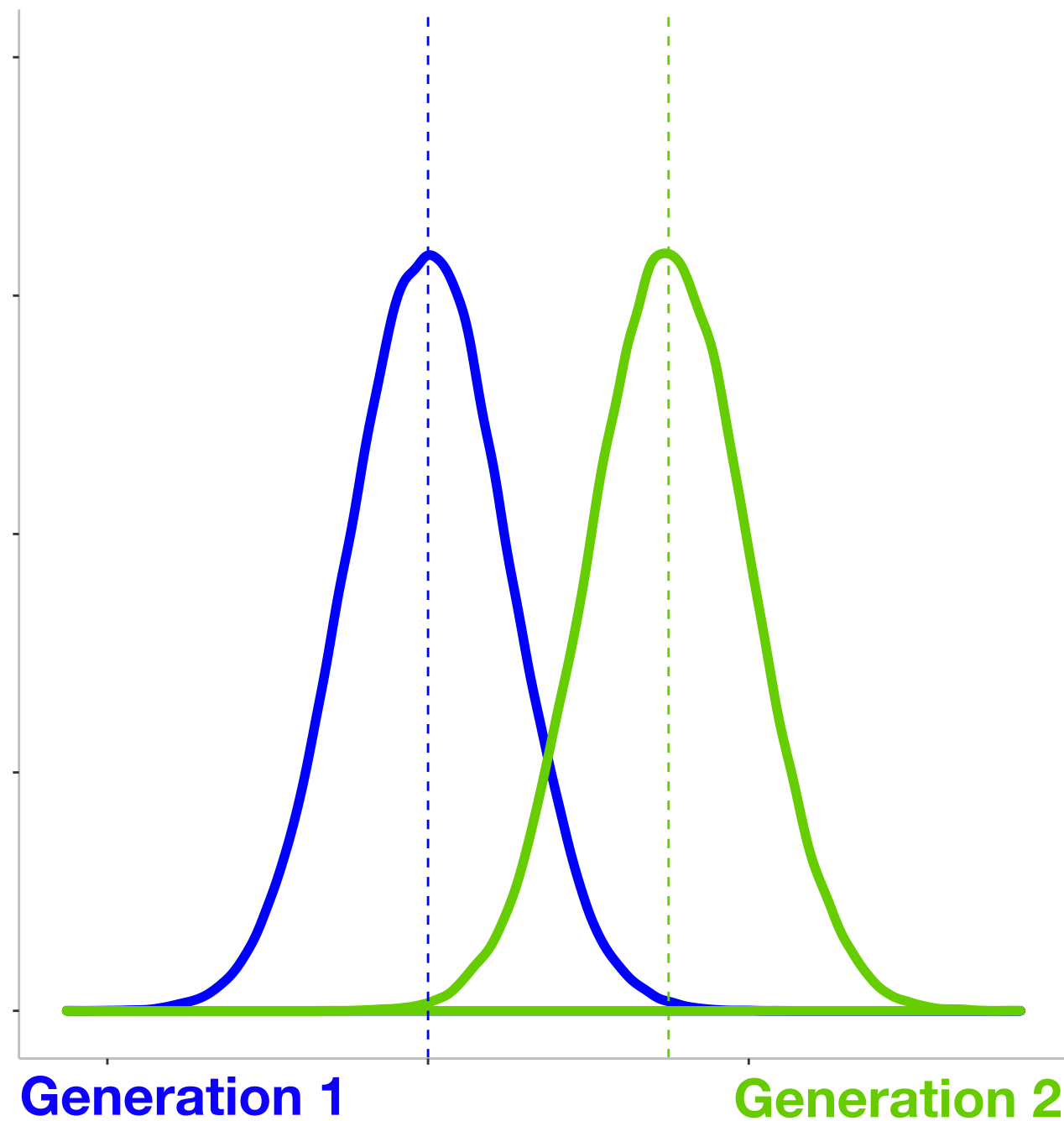
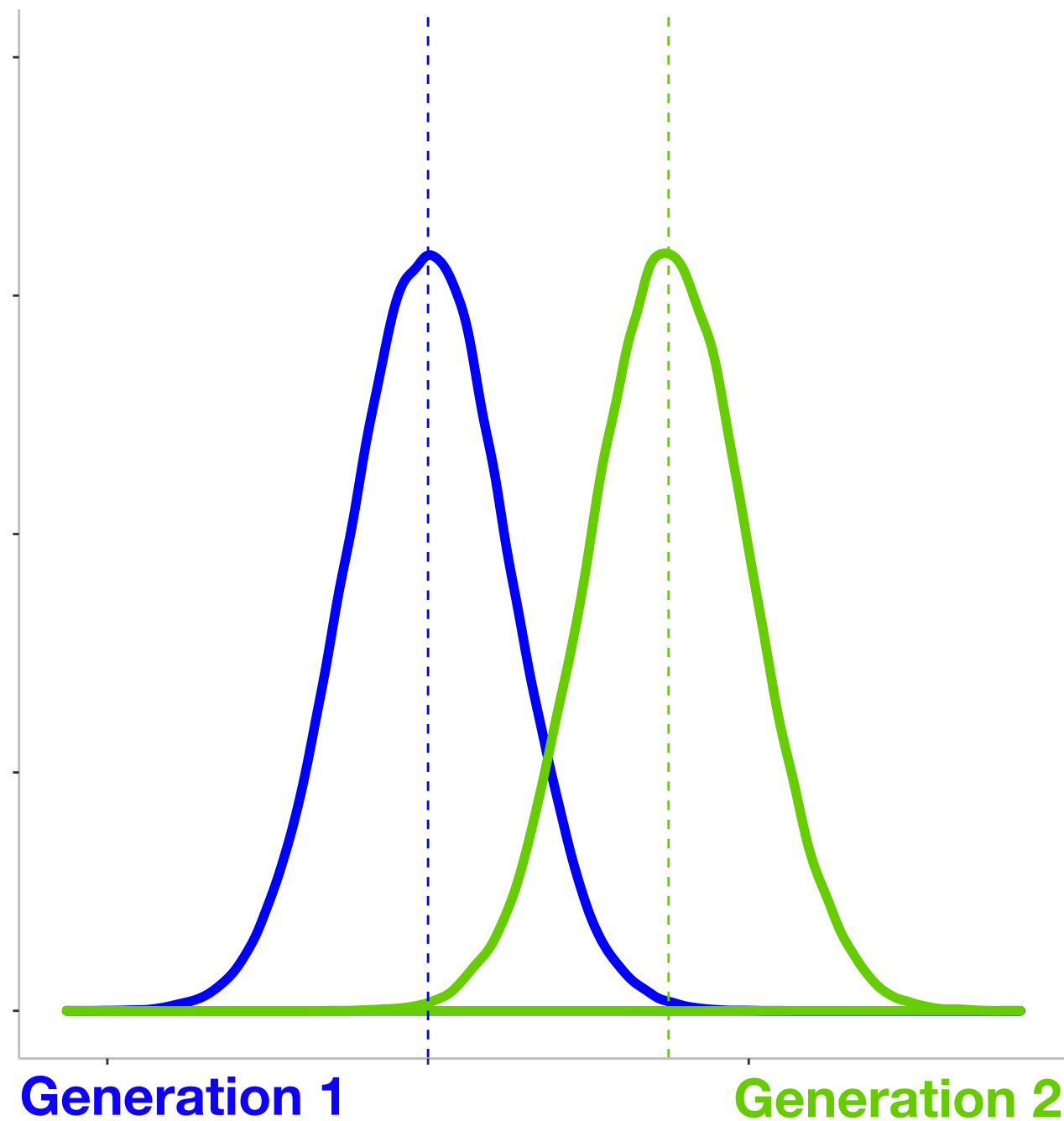


Directional selection



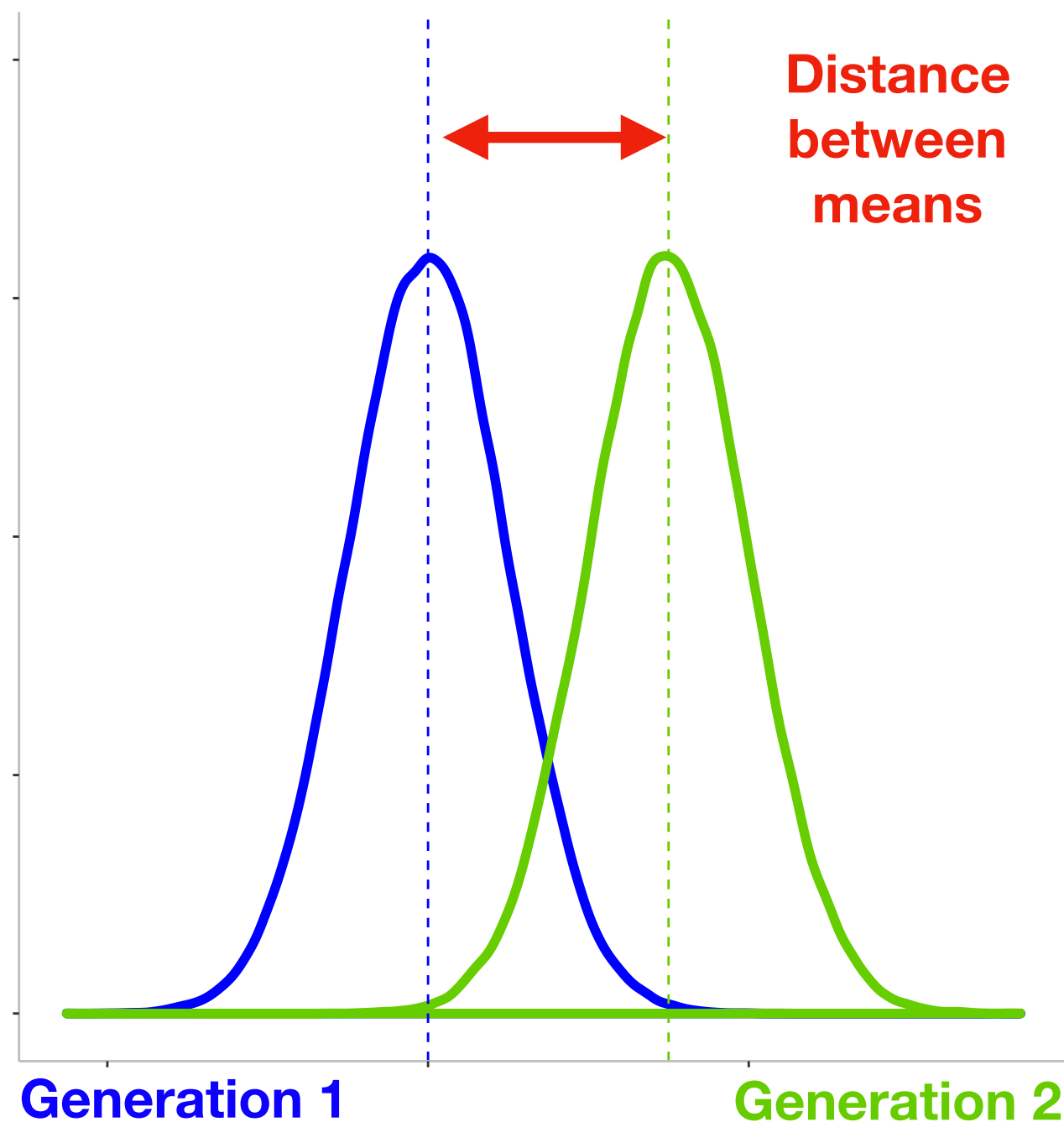
Directional selection



What is our evidence?



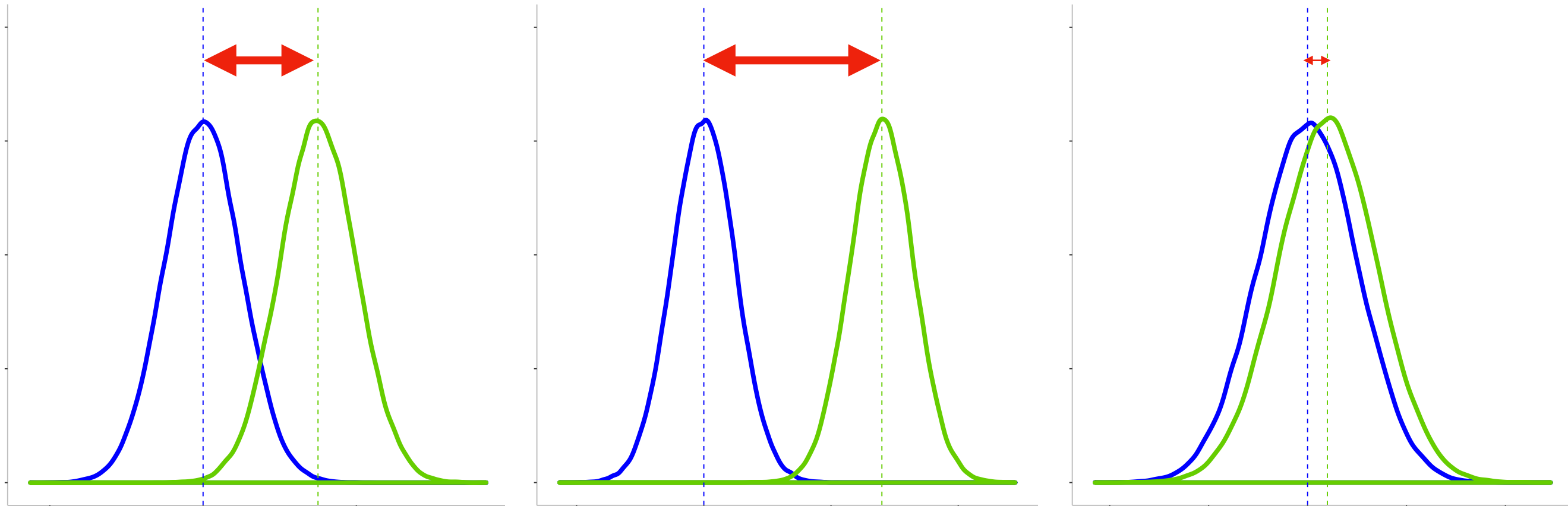
Directional selection



What is our evidence?



Directional selection

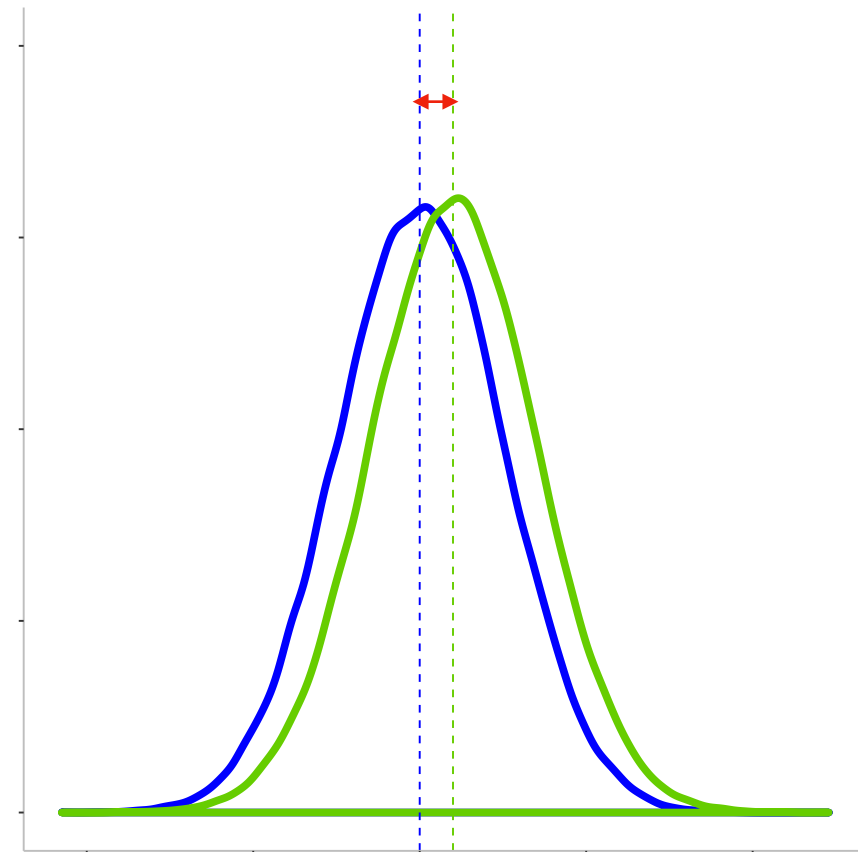
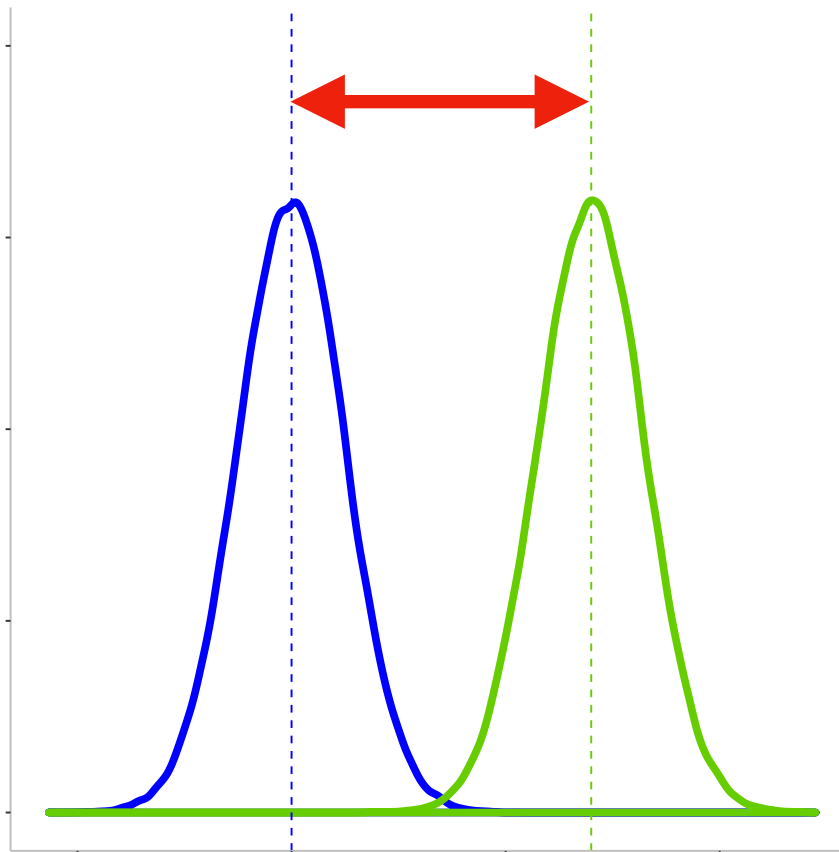
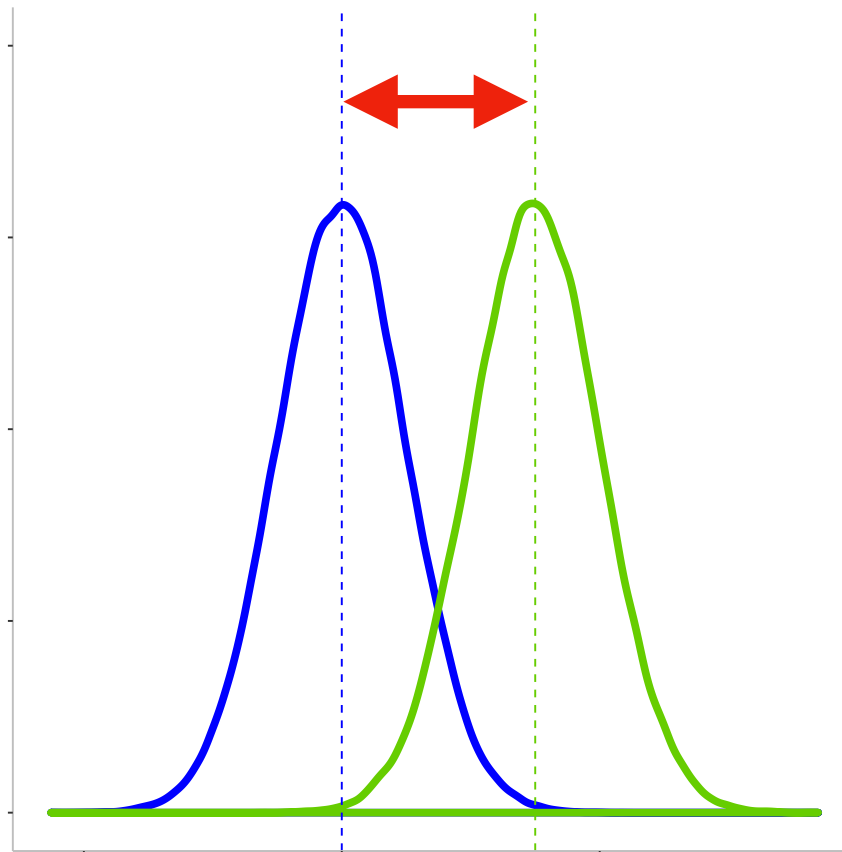


What is our evidence?

**Distance
between
means**



Directional selection



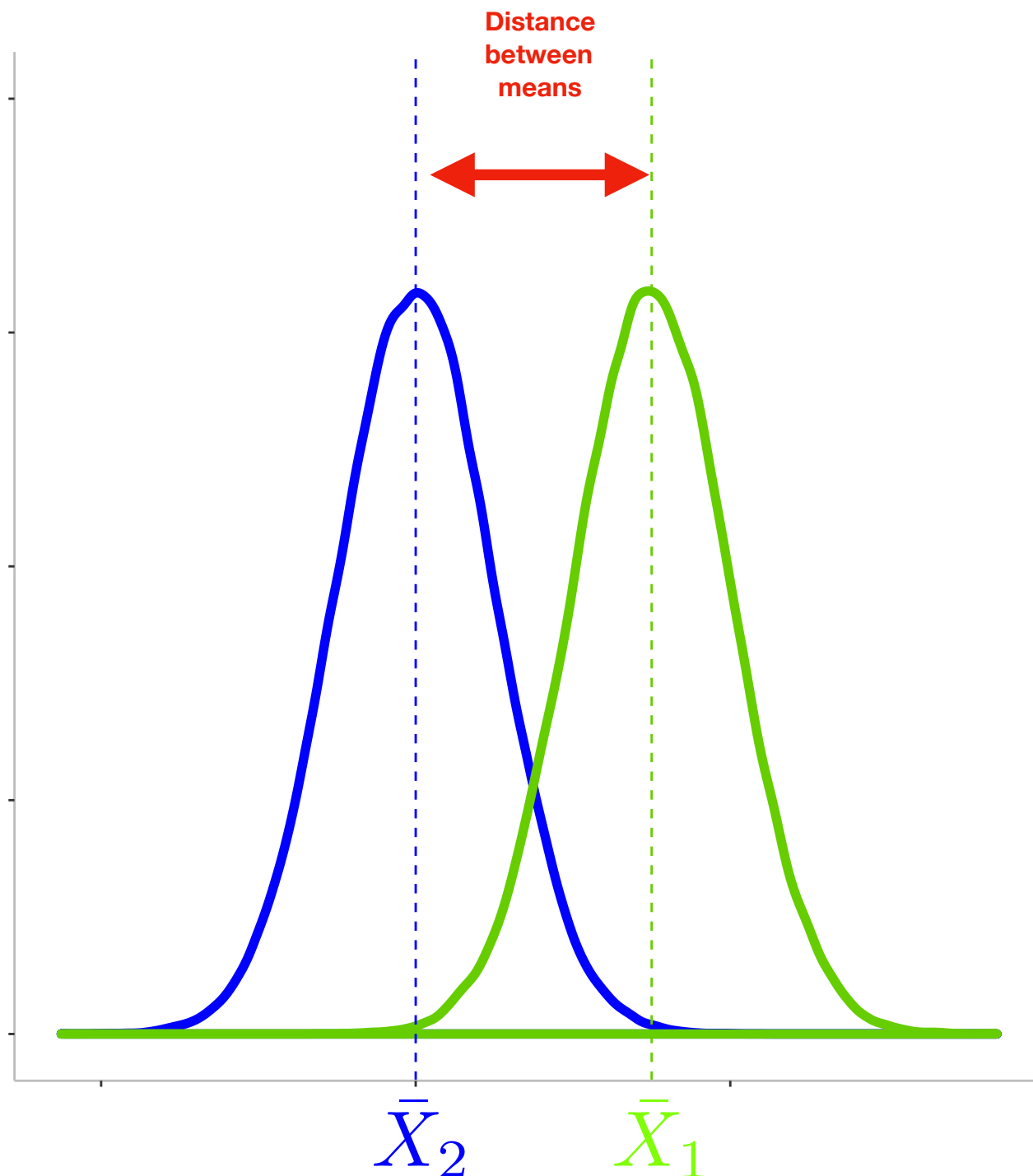
What is our evidence?

**Distance
between
means**

Is this enough distance to
claim directional selection?



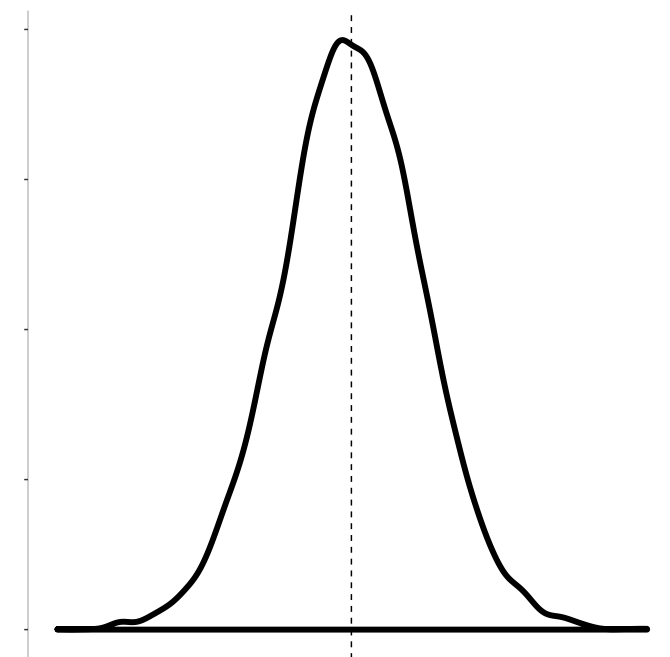
Directional selection



What is our evidence?

t test

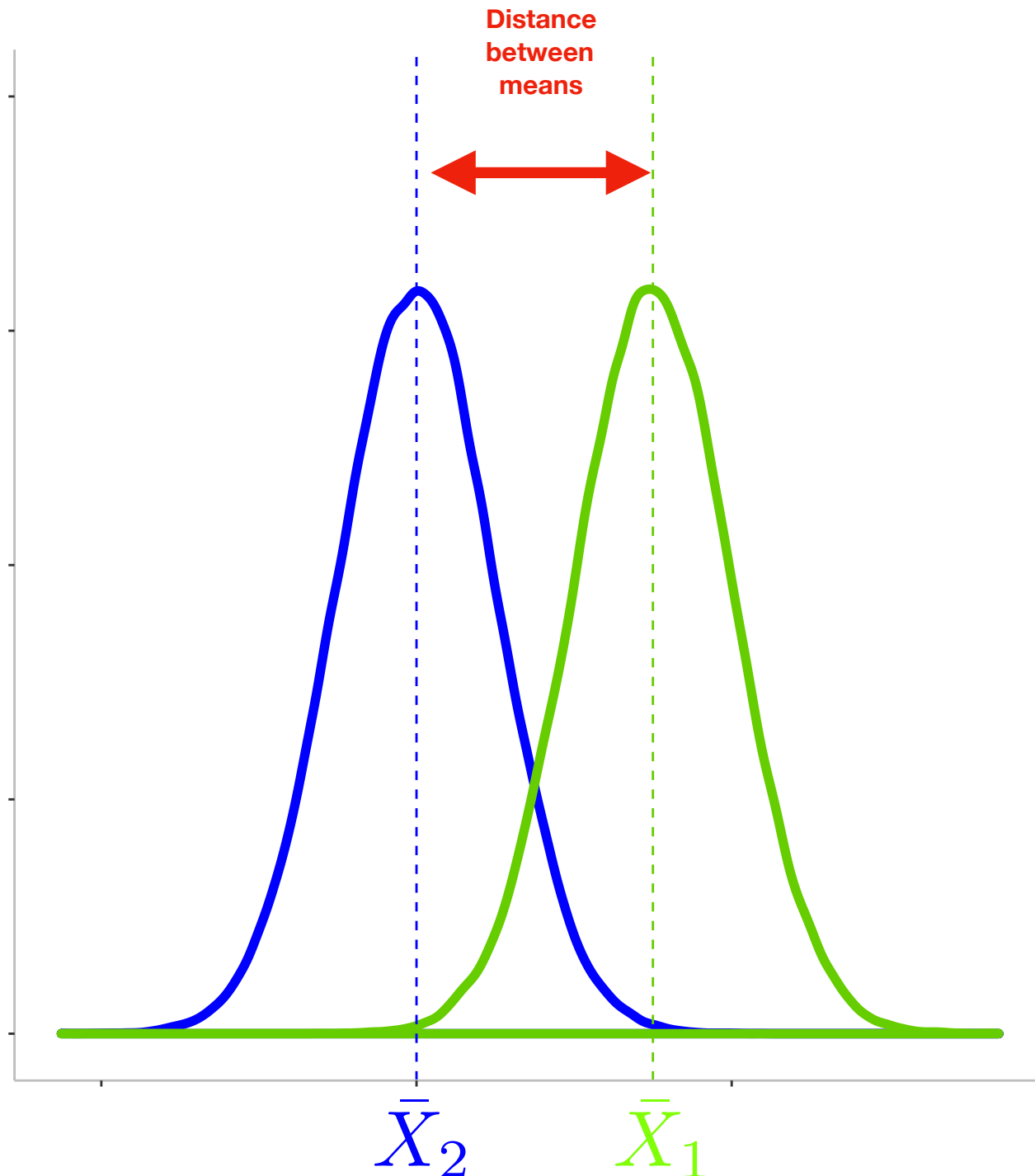
$$t = \frac{\bar{X}_1 - \bar{X}_2}{s/\sqrt{n}}$$



Expected distribution of differences if there is NO directional selection



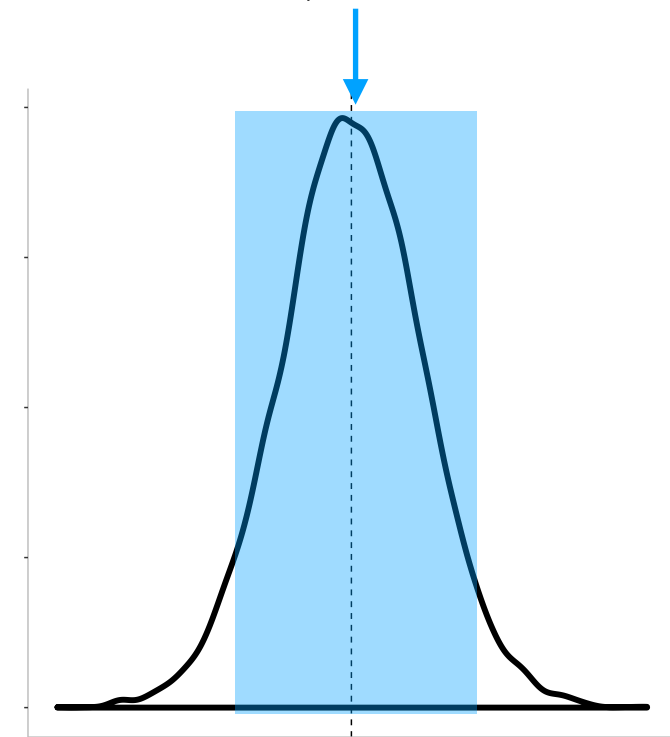
Directional selection



What is our evidence?

t test

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s / \sqrt{n}}$$

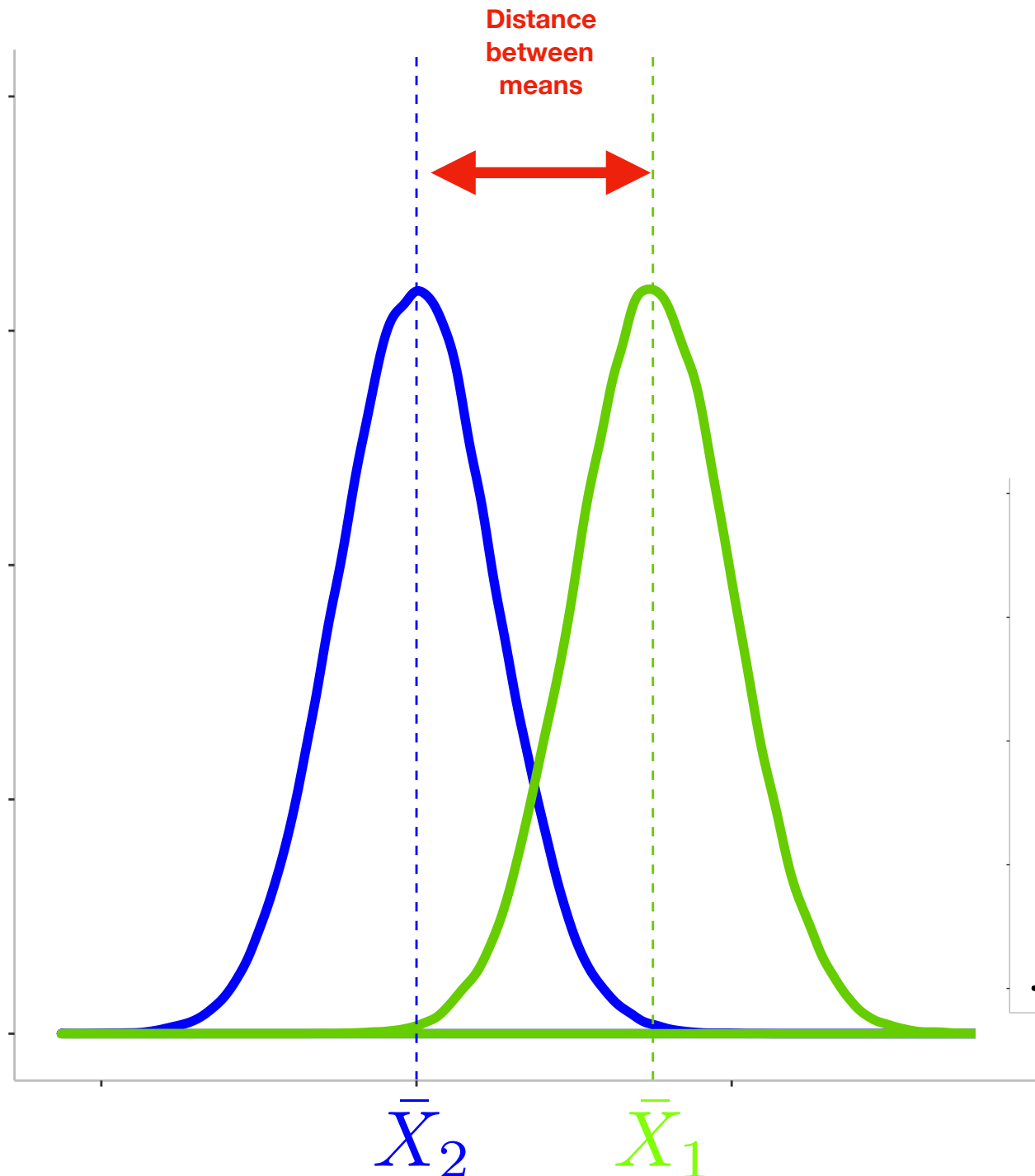


Expected distribution of differences if there is NO directional selection

Agrees with statistical null hypothesis: no difference in means (no directional selection)



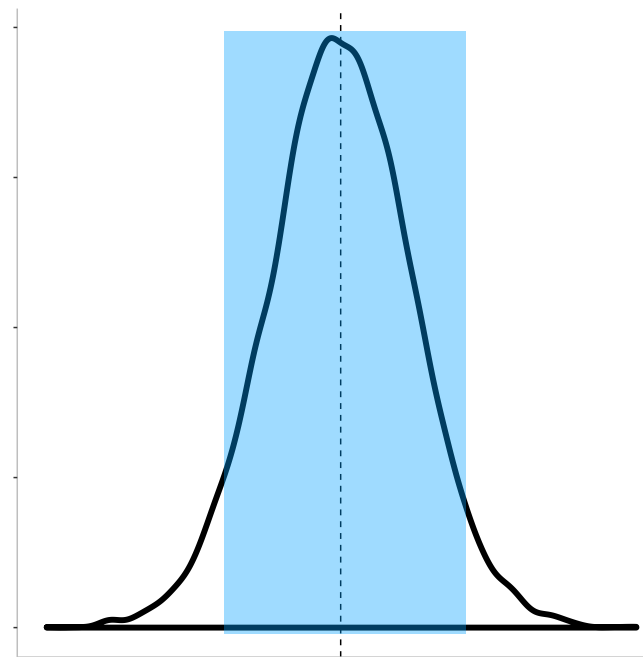
Directional selection



What is our evidence?

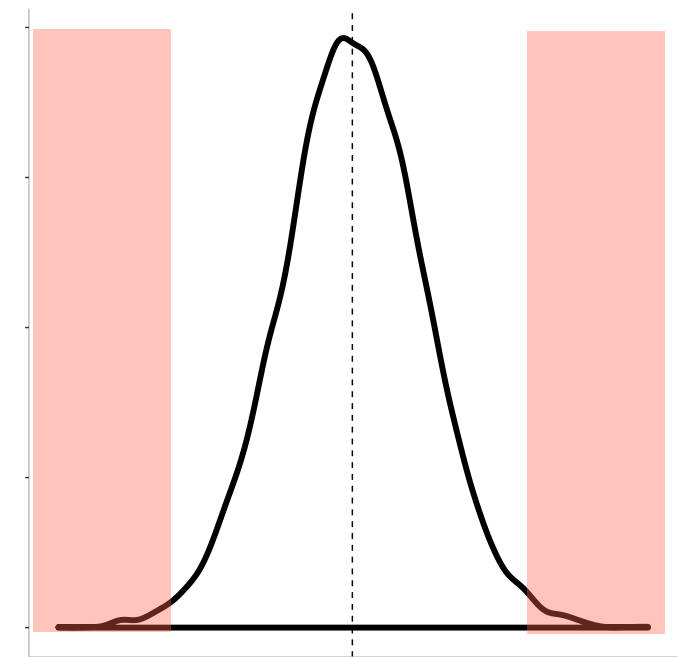
t test

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s/\sqrt{n}}$$



Expected distribution of differences if there is NO directional selection

Agrees with statistical null hypothesis: no difference in means (no directional selection)

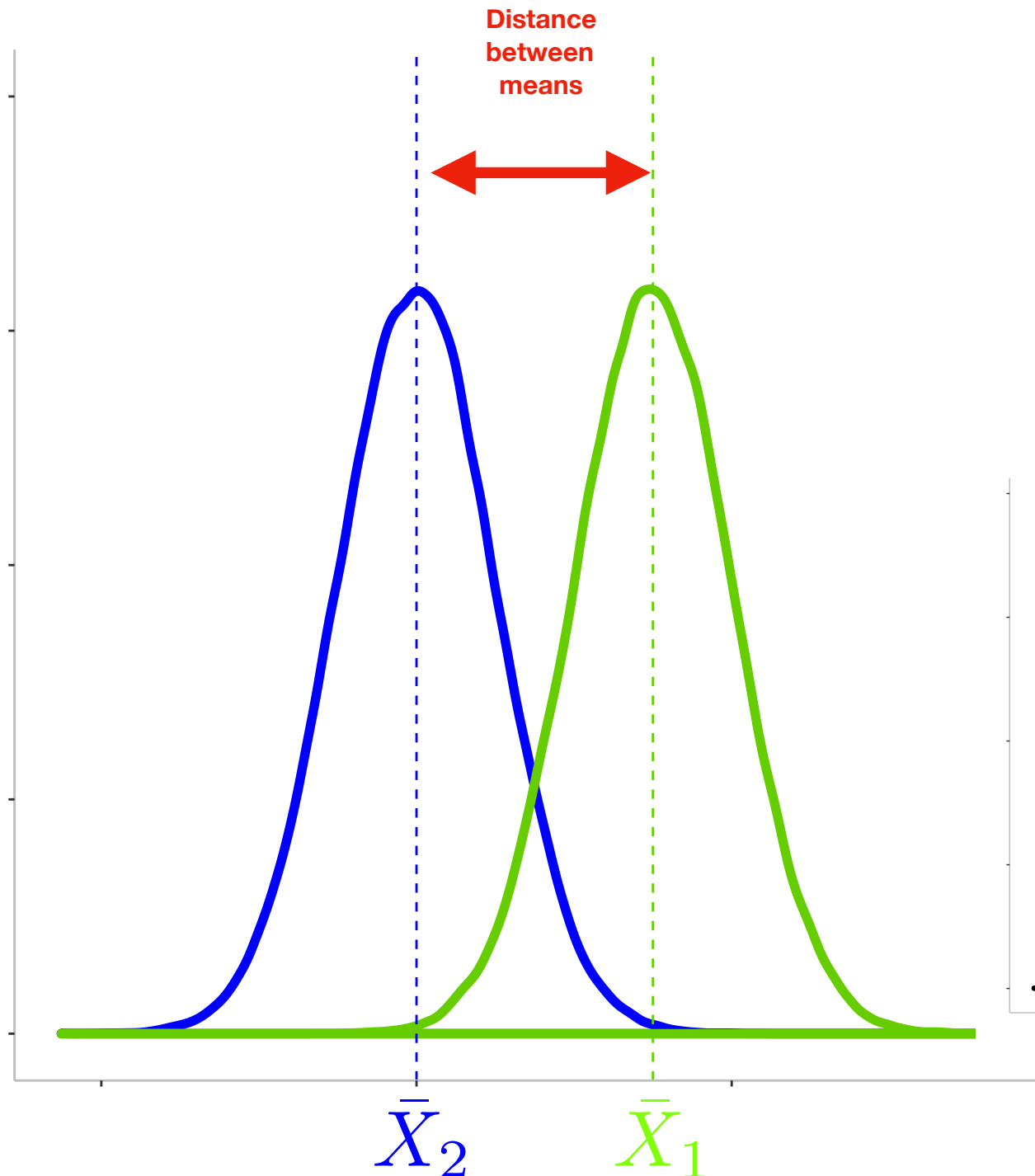


Expected distribution of differences if there is NO directional selection

Evidence of directional selection



Directional selection



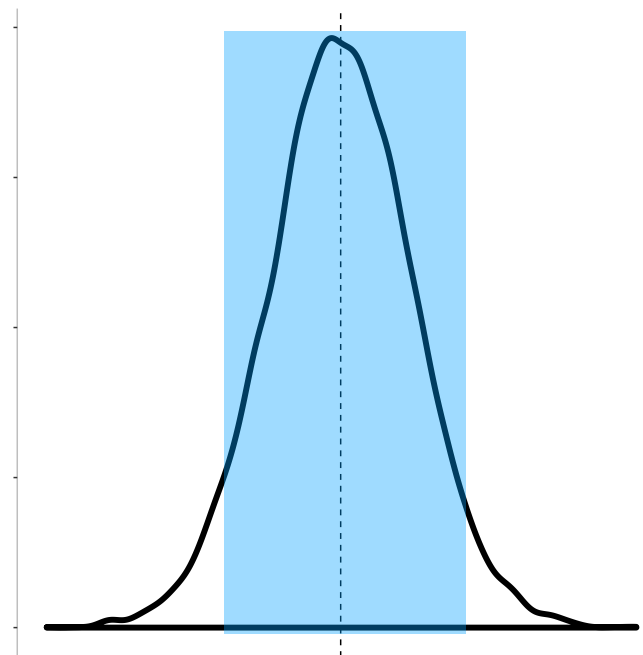
What is our evidence?

t test

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s/\sqrt{n}}$$

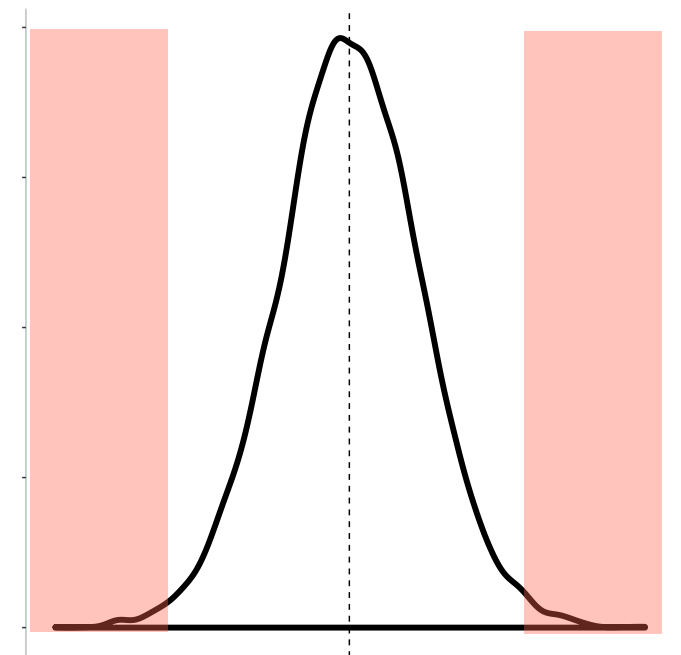
Assumptions

- Normality
- Equal variance



Expected distribution of differences if there is NO directional selection

Agrees with statistical null hypothesis: no difference in means (no directional selection)



Expected distribution of differences if there is NO directional selection

Evidence of directional selection



Our data

	A	B
1	Cotyledons	Generation
2546	3	P
2547	2	P
2548	3	P
2549	3	P
2550	2	P
2551	2	P
2552	3	P
2553	3	P
2554	2	P
2555	2	P
2556	2	P
2557	3	P
2558	3	O
2559	2	O
2560	3	O
2561	2	O
2562	3	O
2563	3	O
2564	3	O
2565	3	O
2566	4	O
2567	3	O

Good practices when storing data in spreadsheets:

1. Save file as csv
2. Name file without spaces or special characters
3. Meaningful column names without special characters
4. No empty cells
5. One thing per cell
6. Create a data dictionary
7. No calculations in the raw data files
8. Do not use font/color to highlight data

THE AMERICAN STATISTICIAN
2018, VOL. 72, NO. 1, 2–10
<https://doi.org/10.1080/00031305.2017.1375989>

Data Organization in Spreadsheets

Karl W. Broman^a and Kara H. Woo^b



<https://solislemuslab.github.io/>

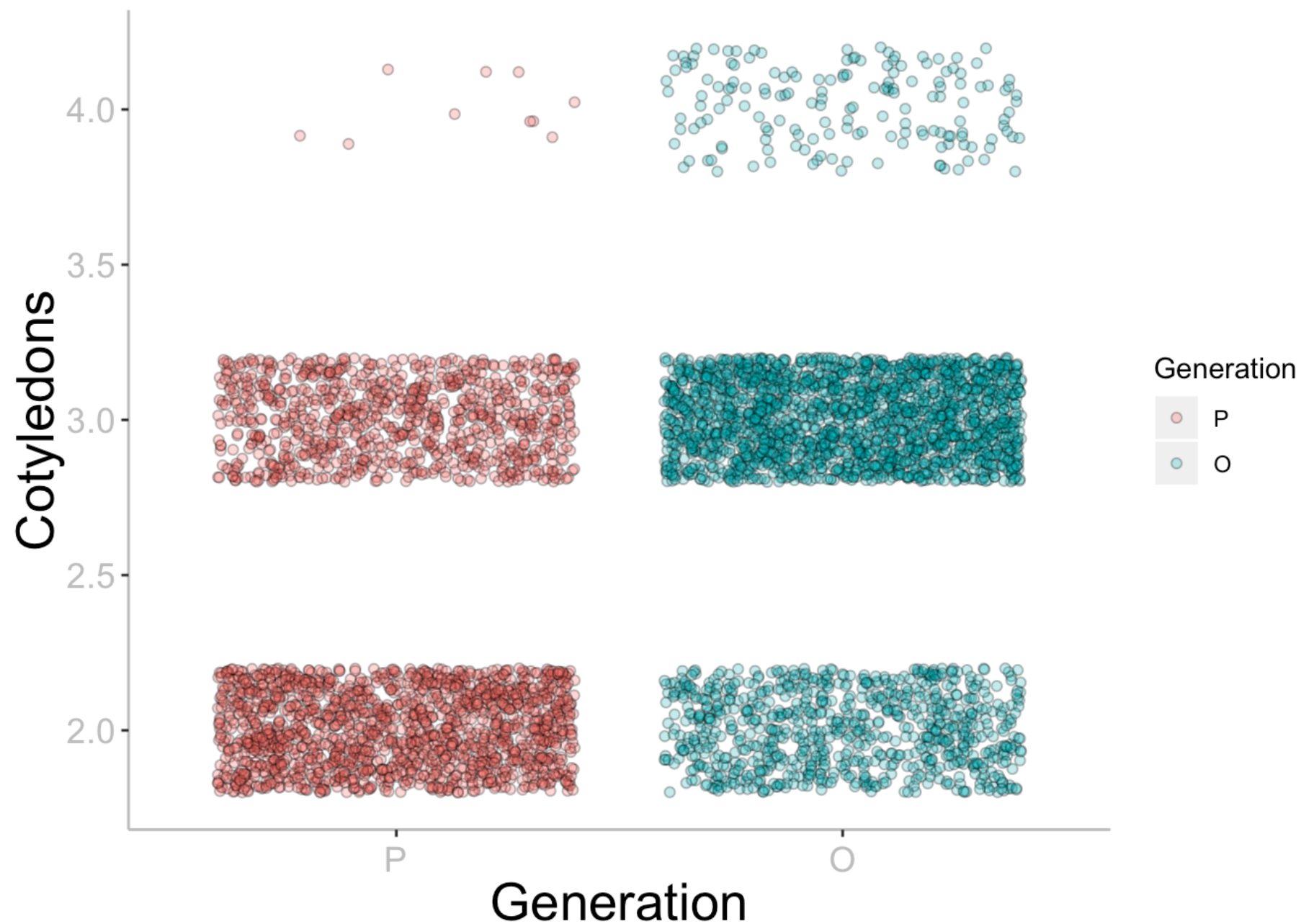


@solislemuslab



crsl4

Data Visualization



<https://solislemuslab.github.io/>



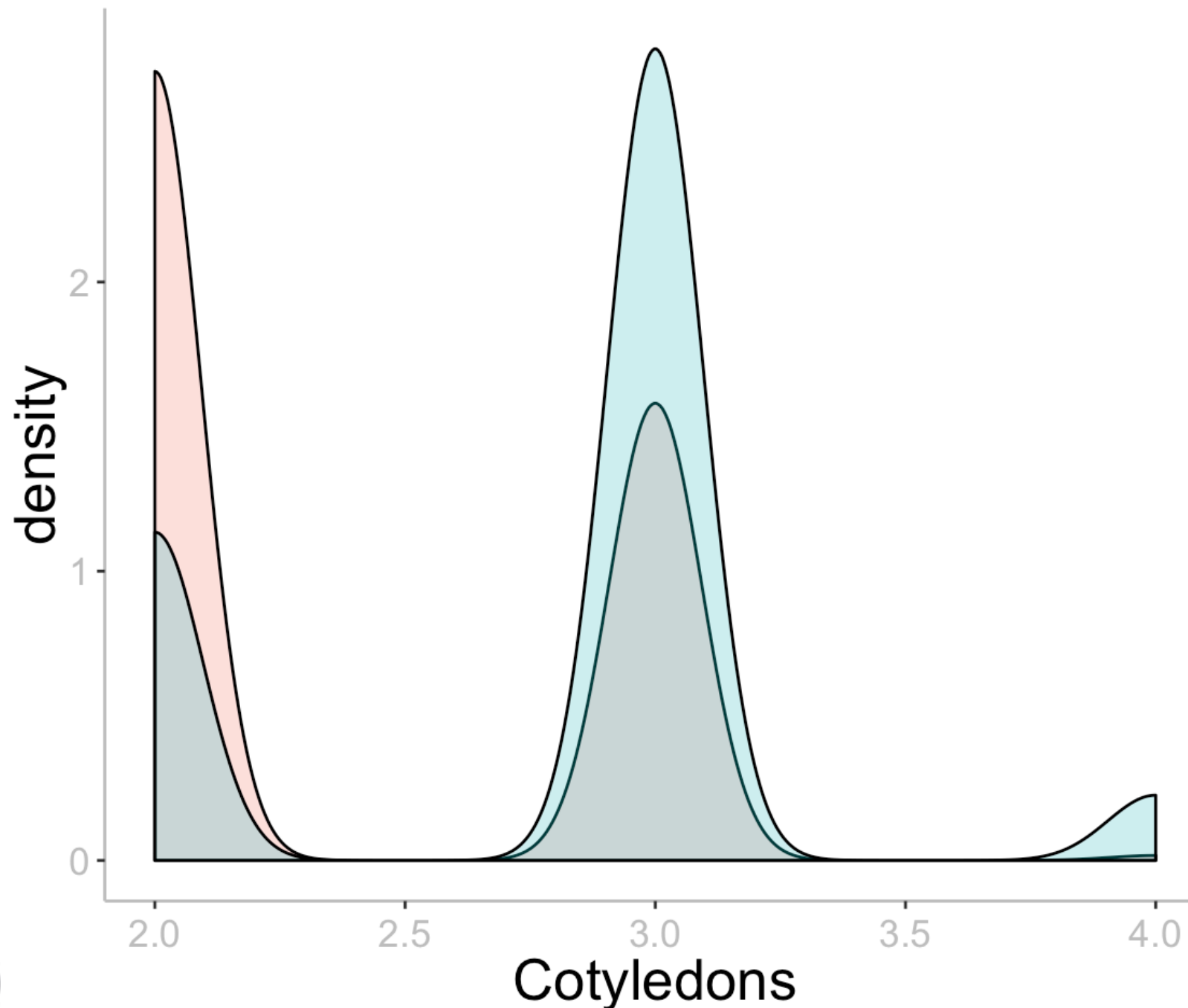
@solislemuslab



crsl4



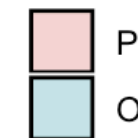
Data Analysis: t test



Assumptions

- Normality
- Equal variance

Generation



<https://solislemuslab.github.io/>



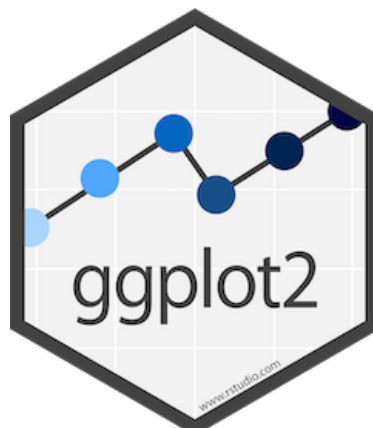
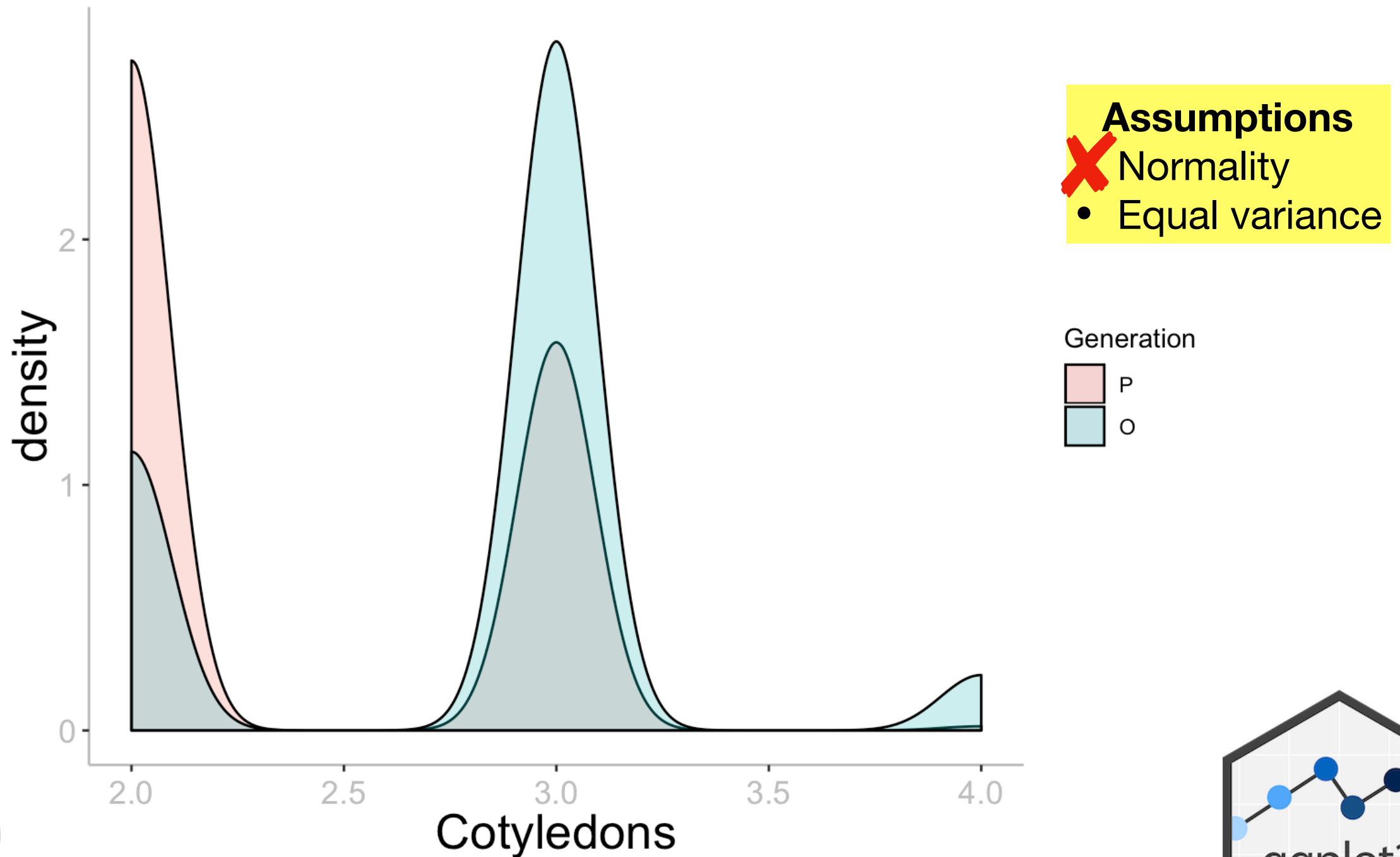
@solislemuslab



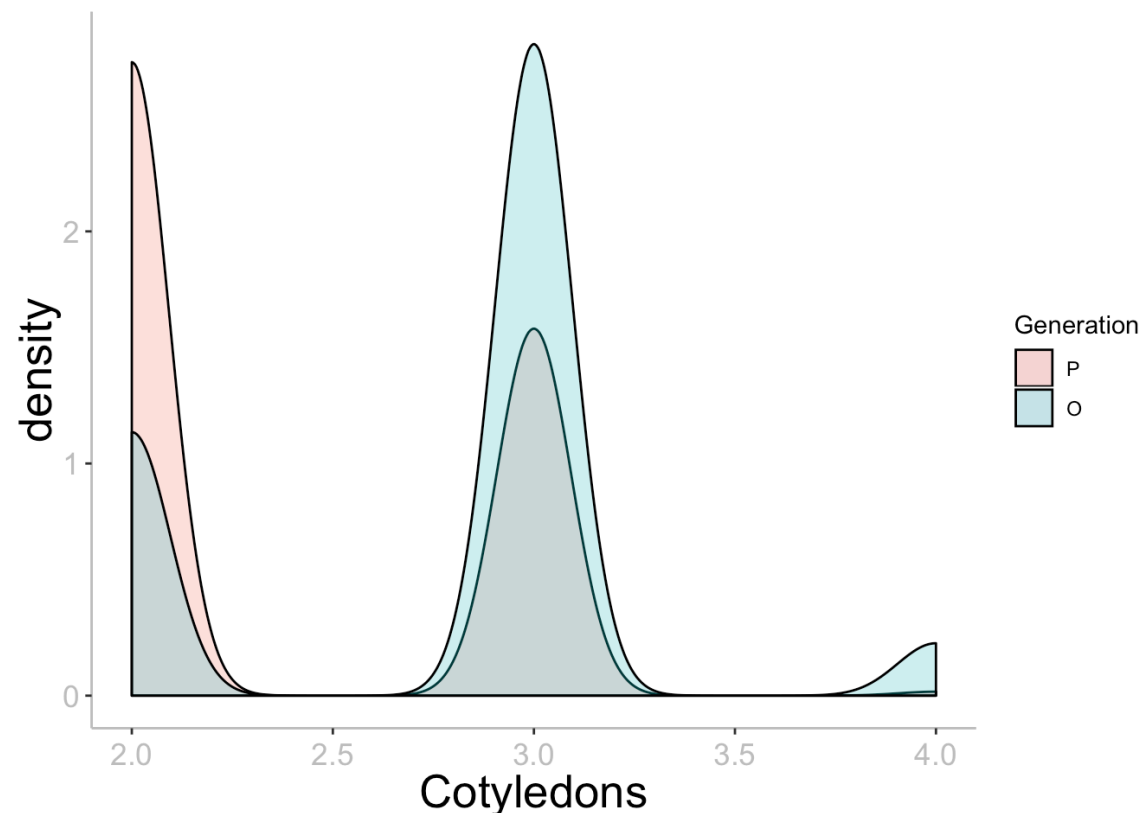
crsl4



Data Analysis: t test



Data Analysis

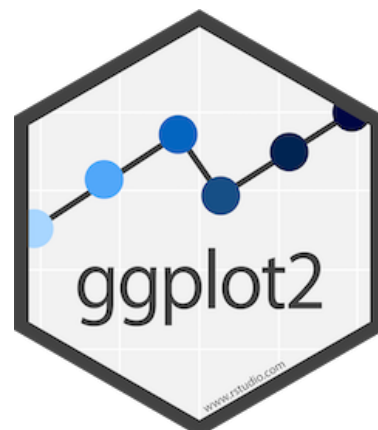


What is the problem?

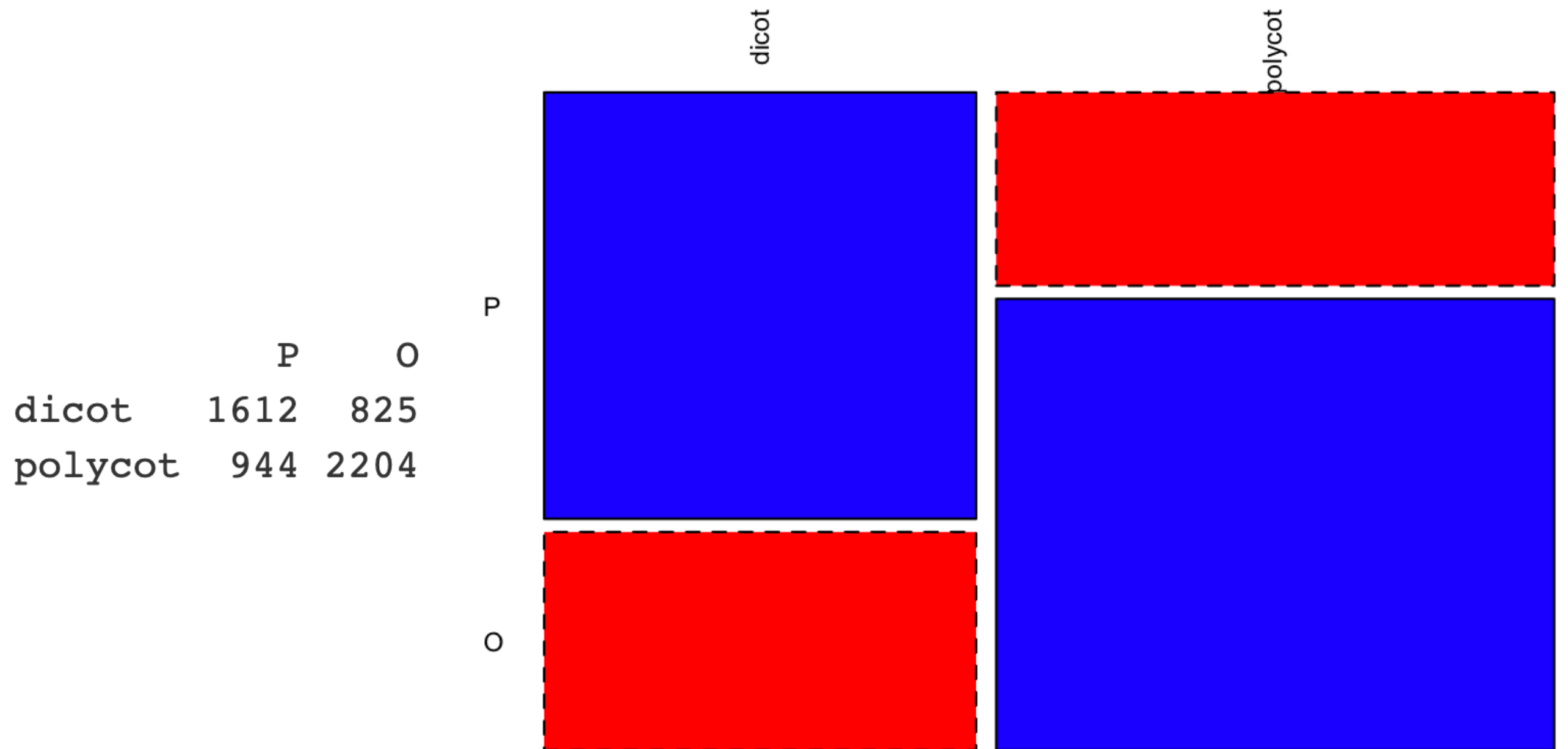
Variable is not truly continuous.
These are counts

What can we do?

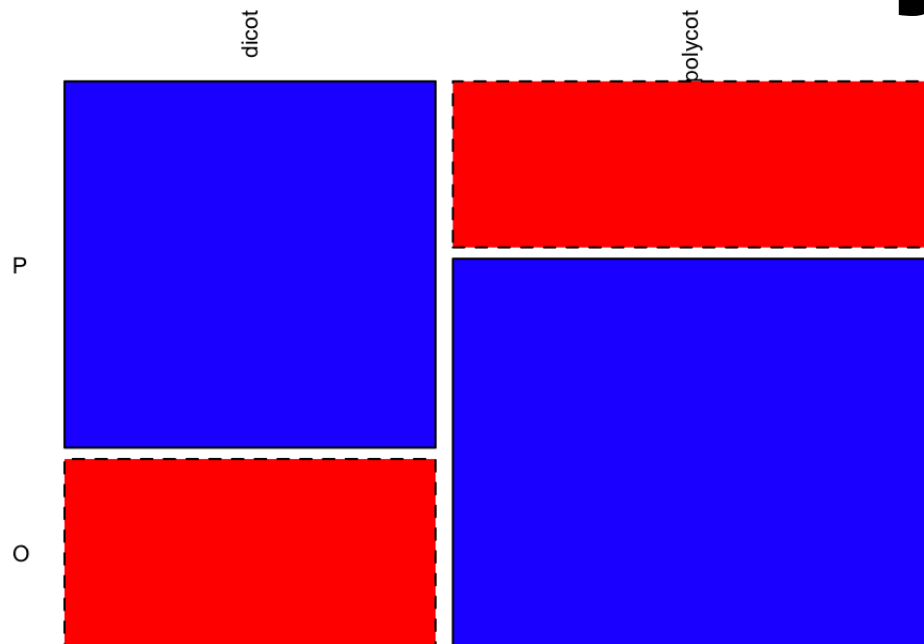
- Use a fancier t test that does not require normality (Wilcoxon-Mann-Whitney test)
- Create categories (dicots/polycots) and do a chi-square test



Data Analysis: chi-square test



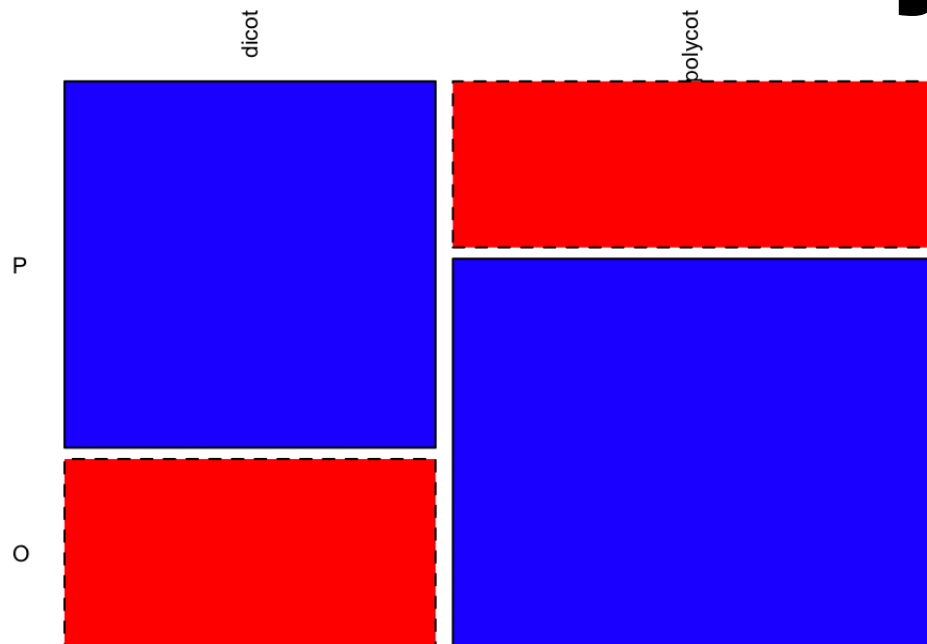
Data Analysis: chi-square test



$$\chi^2 = \sum \frac{(o - e)^2}{e}$$

	P	O
dicot	1612	825
polycot	944	2204

Data Analysis: chi-square test

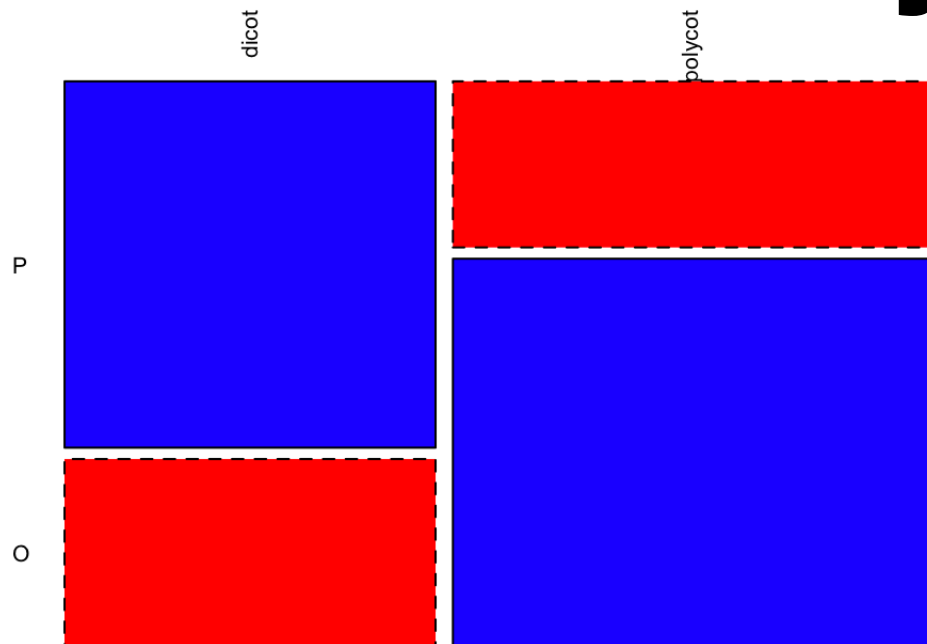


$$\chi^2 = \sum \frac{(o - e)^2}{e}$$

	P	O
dicot	1612	825
polycot	944	2204

Statistical null hypothesis: No difference in the proportion of dycots/polycots in parents and offspring populations

Data Analysis: chi-square test

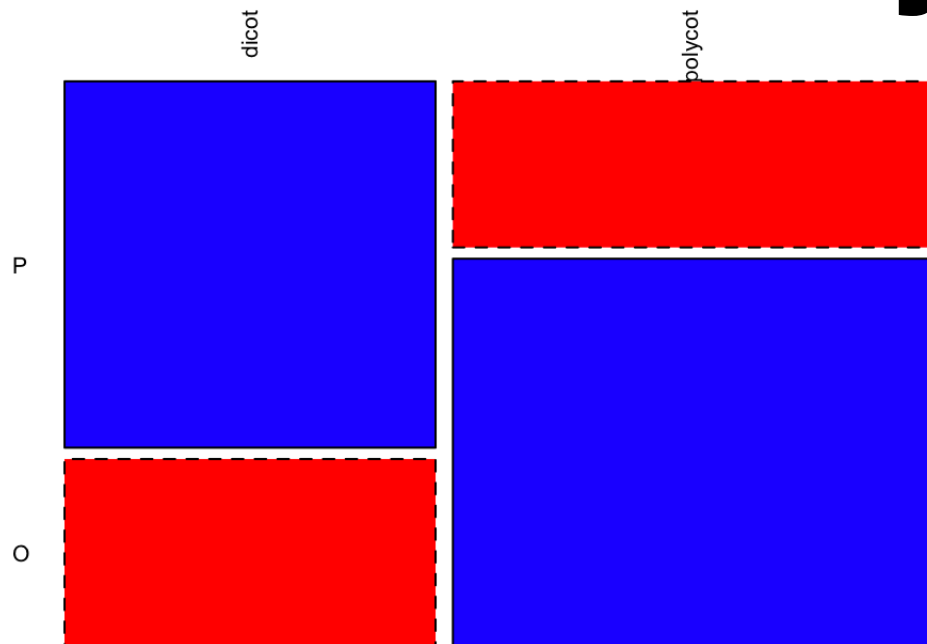


$$\chi^2 = \sum \frac{(o - e)^2}{e}$$

	P	O	
dicot	1612	825	2437
polycot	944	2204	3148
	2556	3029	5585

Statistical null hypothesis: No difference in the proportion of dycots/polycots in parents and offspring populations

Data Analysis: chi-square test

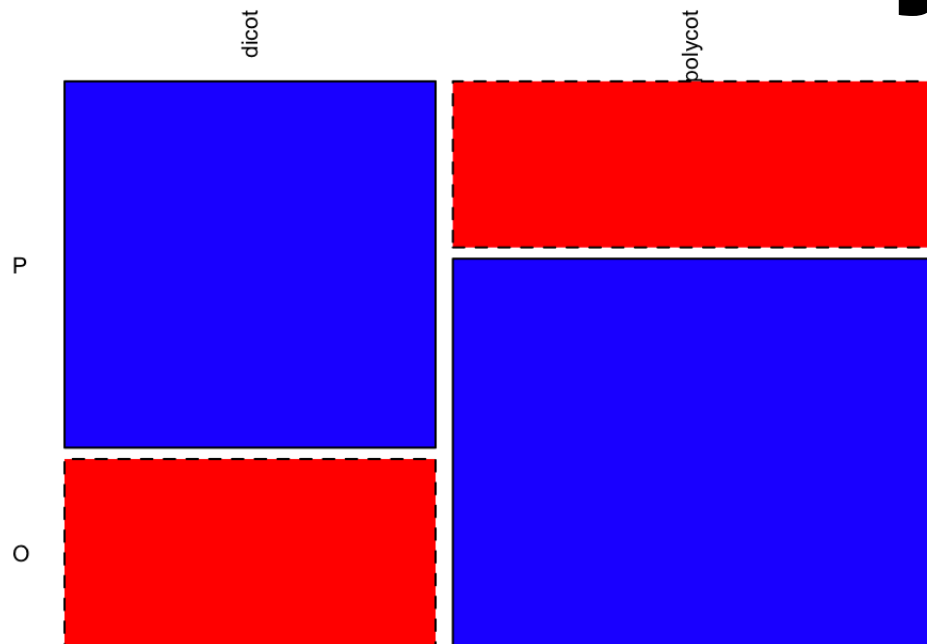


$$\chi^2 = \sum \frac{(o - e)^2}{e}$$

	P	O	
dicot	1612	825	2437
polycot	944	2204	3148
	2556	3029	5585

Statistical null hypothesis: No difference in the proportion of dycots/polycots in parents and offspring populations

Data Analysis: chi-square test



$$\chi^2 = \sum \frac{(o - e)^2}{e}$$

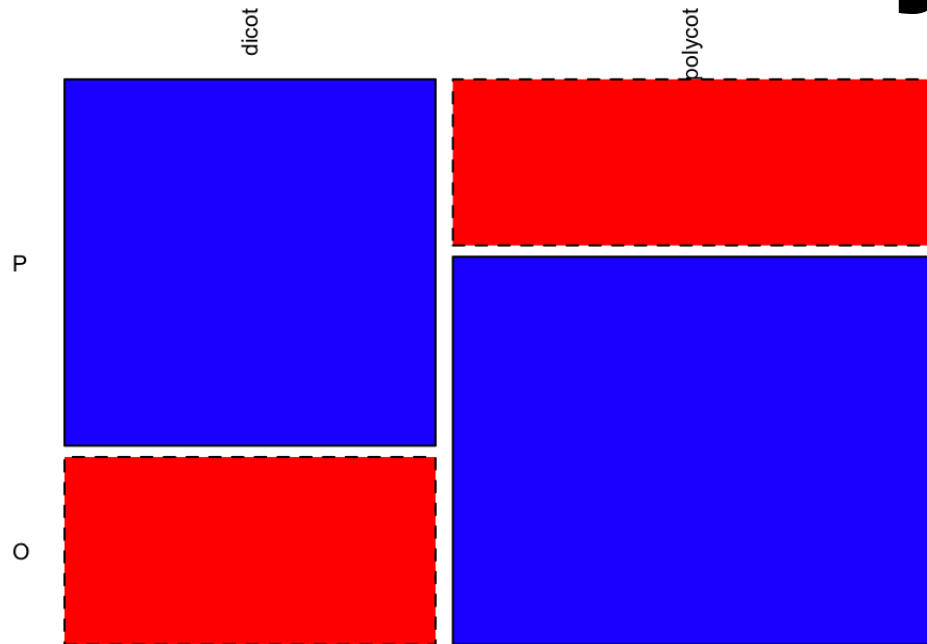
	P	O	
dicot	1612	825	2437
polycot	944	2204	3148
	2556	3029	5585

$\frac{2437}{5585} = 0.436$

$\frac{3148}{5585} = 0.564$

Statistical null hypothesis: No difference in the proportion of dycots/polycots in parents and offspring populations

Data Analysis: chi-square test



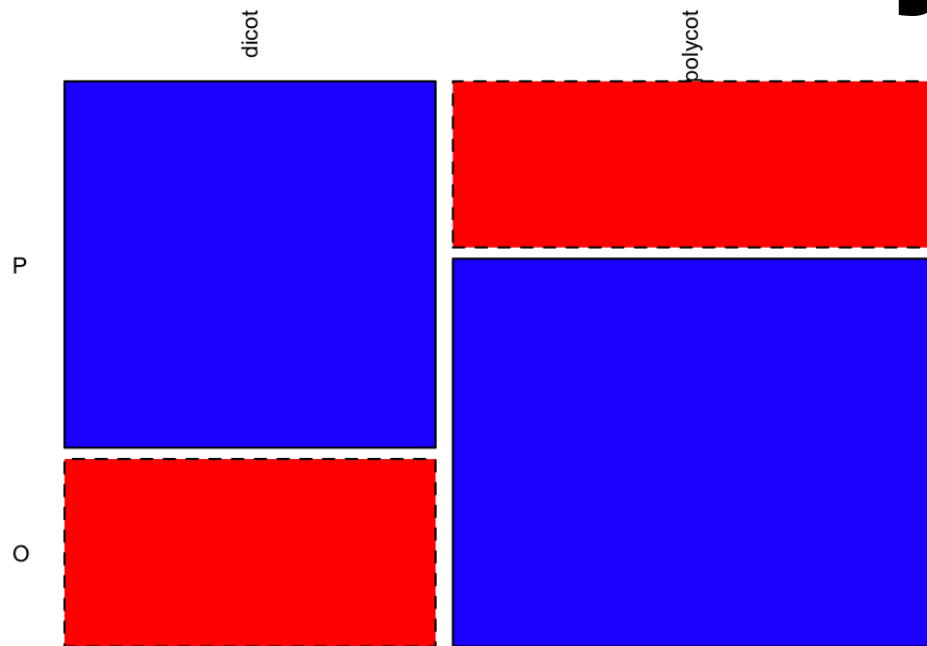
$$\chi^2 = \sum \frac{(o - e)^2}{e}$$

	P	O	
dicot	1612	825	$\frac{2437}{5585} = 0.436$
polycot	944	2204	$\frac{3148}{5585} = 0.564$
	2556	3029	5585

Statistical null hypothesis: No difference in the proportion of dycots/polycots in parents and offspring populations



Data Analysis: chi-square test



$$\chi^2 = \sum \frac{(o - e)^2}{e}$$

	P	O	
dicot	1612	825	2437
polycot	944	2204	3148
	2556	3029	5585

$2556 * 0.436 = 2556 * \frac{2437}{5585} \rightarrow 1115.3$

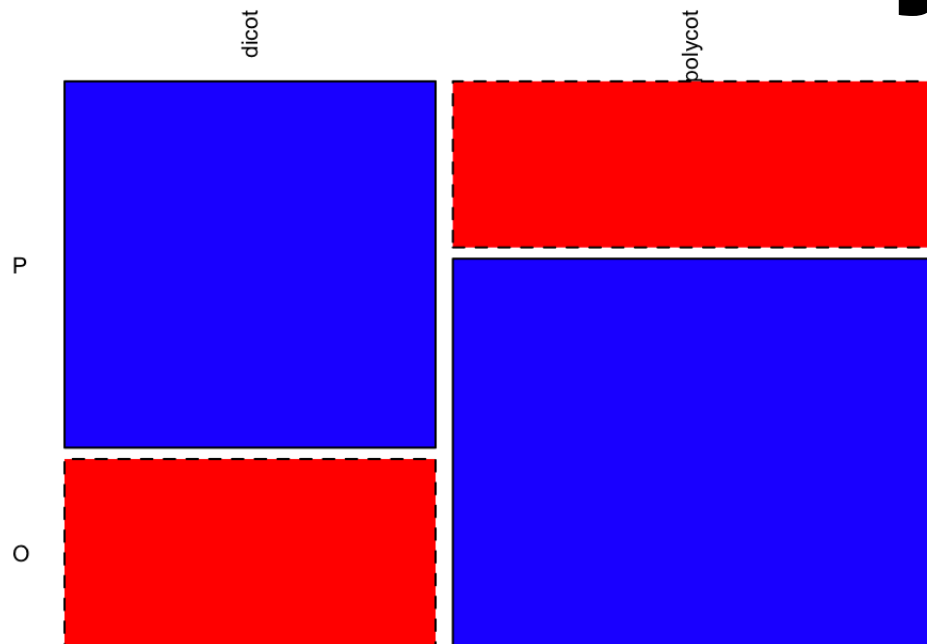
$\frac{2437}{5585} = 0.436$

$\frac{3148}{5585} = 0.564$

Statistical null hypothesis: No difference in the proportion of dycots/polycots in parents and offspring populations



Data Analysis: chi-square test



$$\chi^2 = \sum \frac{(o - e)^2}{e}$$

		P	O	
	dicot	1612	825	2437
	polycot	944	2204	3148
		2556	3029	5585

$2556 * 0.436 = 2556 * \frac{2437}{5585} \rightarrow 1115.3$

$2556 * 0.564 = 2556 * \frac{3148}{5585} \rightarrow 1440.7$

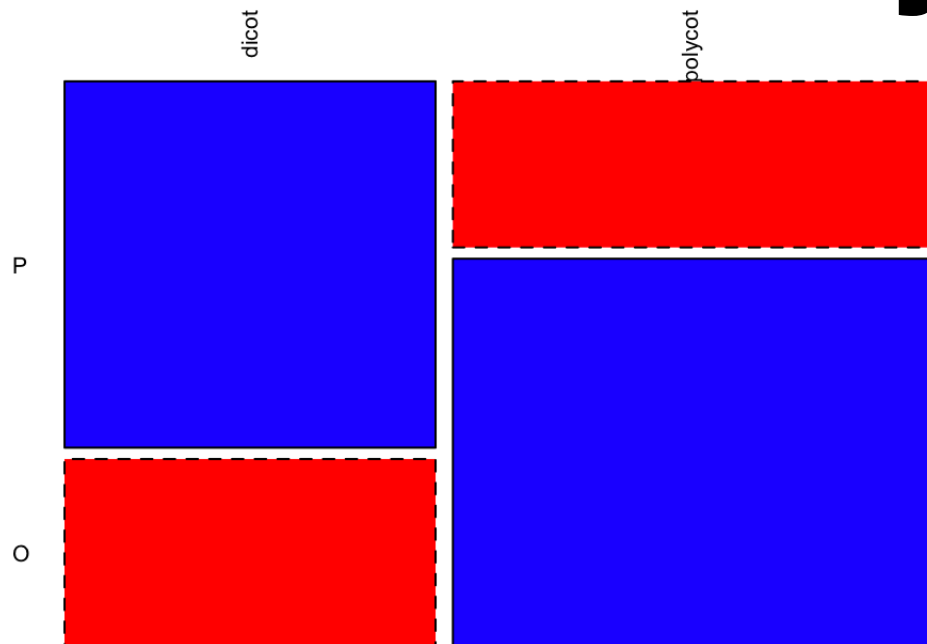
$\frac{2437}{5585} = 0.436$

$\frac{3148}{5585} = 0.564$

Statistical null hypothesis: No difference in the proportion of dycots/polycots in parents and offspring populations



Data Analysis: chi-square test



$$\chi^2 = \sum \frac{(o - e)^2}{e}$$

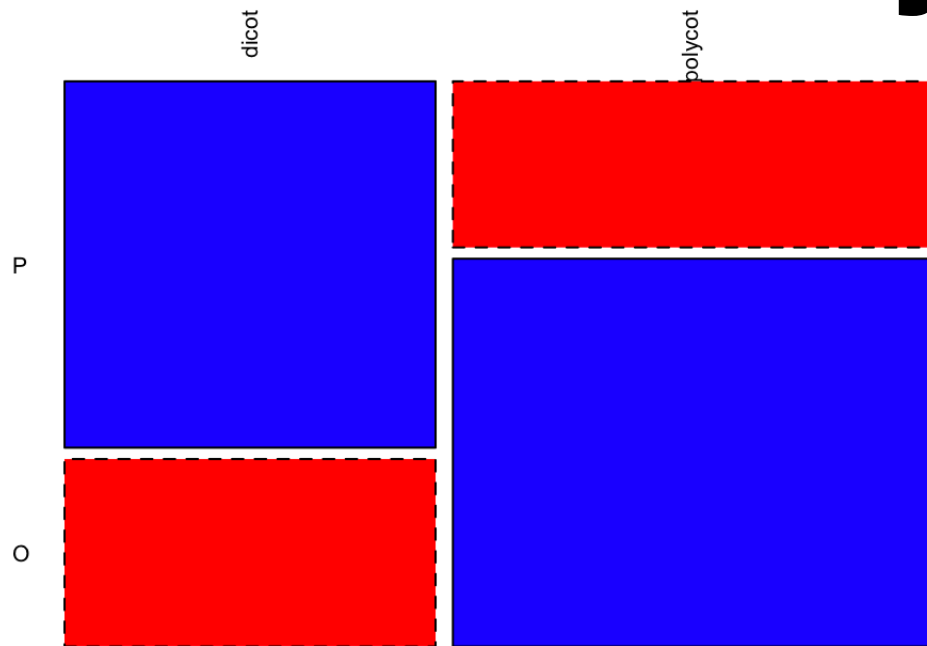
		P	O	
dicot	1612	825	2437	$\frac{2437}{5585} = 0.436$
		1115.3	1321.7	
polycot	944	2204	3148	$\frac{3148}{5585} = 0.564$
	2556	3029	5585	

$2556 * 0.436 = 2556 * \frac{2437}{5585} \rightarrow 1115.3$
 $2556 * 0.564 = 2556 * \frac{3148}{5585} \rightarrow 1440.7$

Statistical null hypothesis: No difference in the proportion of dycots/polycots in parents and offspring populations



Data Analysis: chi-square test



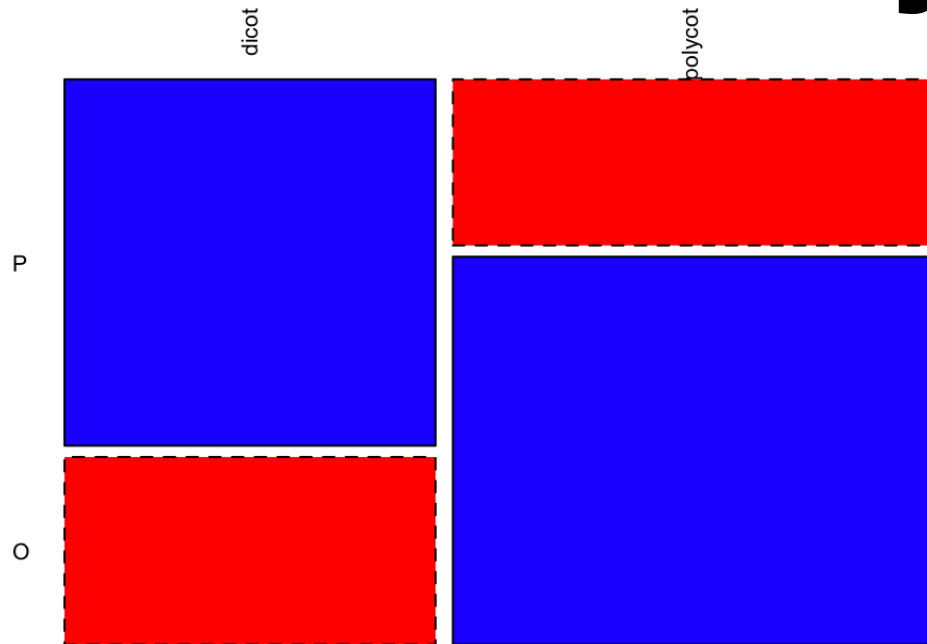
$$\chi^2 = \sum \frac{(o - e)^2}{e}$$

		P	O	
dicot	1612	825	2437	$\frac{2437}{5585} = 0.436$
				$2556 * 0.436 = 2556 * \frac{2437}{5585} \rightarrow 1115.3$
polycot	944	2204	3148	$\frac{3148}{5585} = 0.564$
				$2556 * 0.564 = 2556 * \frac{3148}{5585} \rightarrow 1440.7$
	2556	3029	5585	

Statistical null hypothesis: No difference in the proportion of dycots/polycots in parents and offspring populations



Data Analysis: chi-square test



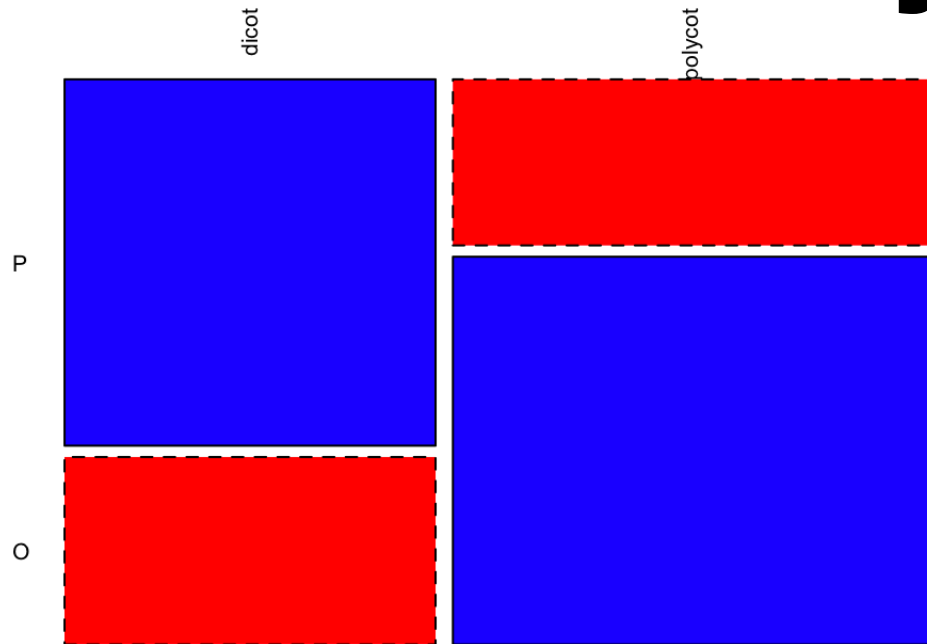
$$\chi^2 = \sum \frac{(o - e)^2}{e} = 722.1475$$

		P	O	
dicot	1612	825	2437	$\frac{2437}{5585} = 0.436$
				$2556 * 0.436 = 2556 * \frac{2437}{5585} \rightarrow 1115.3$
polycot	944	2204	3148	$\frac{3148}{5585} = 0.564$
				$2556 * 0.564 = 2556 * \frac{3148}{5585} \rightarrow 1440.7$
	2556	3029	5585	

Statistical null hypothesis: No difference in the proportion of dycots/polycots in parents and offspring populations



Data Analysis: chi-square test



$$\chi^2 = \sum \frac{(o - e)^2}{e} = 722.1475$$

pvalue = 4.6e-159
Reject null hypothesis
Evidence of selection

	P	O	
dicot	1612	825	$\frac{2437}{5585} = 0.436$
polycot	944	2204	$\frac{3148}{5585} = 0.564$
	2556	3029	5585

$2556 * 0.436 = 2556 * \frac{2437}{5585} \rightarrow 1115.3$
 $2556 * 0.564 = 2556 * \frac{3148}{5585} \rightarrow 1440.7$

1321.7
 1707.3

Statistical null hypothesis: No difference in the proportion of dycots/polycots in parents and offspring populations



Data analysis on WI Fast Stats app

<https://wi-fast-stats.shinyapps.io/webinar-aug20/>

Acknowledgement
Yizhou Liu
(CS undergraduate
student at UW-Madison)



<https://solislemuslab.github.io/>



@solislemuslab



crsl4