




“Barley” significant:

Chi-squared goodness of fit test

Fonti Kar + Emi Tanaka

Outline

- Motivating example 
- Significance testing
 - Test statistic (Chi-squared goodness of fit)
 - Degrees of freedom
 - Significance level and critical value
 - P -value
- A sweet little exercise
- GTA x ZoomAgri





Shelley Down ▸ Cadbury Dairy Milk

Yesterday · 🌐

Hiiii Cadbury

Was super excited to watch my son open his Favourites Easter Egg but had a bit of a WTF moment when we realised there were 7 Cherry Ripes and 1 Dairy Milk. I know Cherry Ripes get a bit of a tough time and you still need to force them upon us but the ratio seems a bit drastic! Did you have a newbie working on the distribution line that day because surely this ain't right?!

Emma Hayes-Cooke, Lee Bright, Naima Maanaima and 3,268 others like this.

Top Comments ▾



38 shares

398 comments



Cadbury Dairy Milk ✓ Thanks for your feedback. Whilst the process of the pieces going into the packet is randomised, it is not intentional that you would get that much of one particular product. Could you please PM us with the barcode, Best Before Date and batch code from the pack, as well as your address details, and we'll be in touch?

👍 77 · 21 hrs

↩ 16 Replies · 2 hrs



How do we know when we have a leg to stand on?

38 shares

398 comments



Cadbury Dairy Milk Thanks for your feedback. Whilst the process of the pieces going into the

What you
uct.
Best
Before Date and batch code from the pack, as well
as your address details, and we'll be in touch?

77 · 21 hrs

16 Replies · 2 hrs

Cadbury's prevailing view vs. consumers

H_0 = Equal chance of drawing any brand of chocolate



Cadbury's prevailing view vs. consumers

H_0 = Equal chance of drawing any brand of chocolate

H_1 = **Unequal** chance of drawing any brand of chocolate



How well does the observed data fit the theory?



Test statistic

- A standardised, numeric **summary** our data
- Decide whether we have enough evidence to reject the default view (H_0)

Chi-squared
test statistic

observed value

expected value

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

The equation is annotated with orange lines and numbered boxes: [1] points to the minus sign, [2] points to the square on the difference, [3] points to the denominator E_i , and [4] points to the summation symbol \sum .

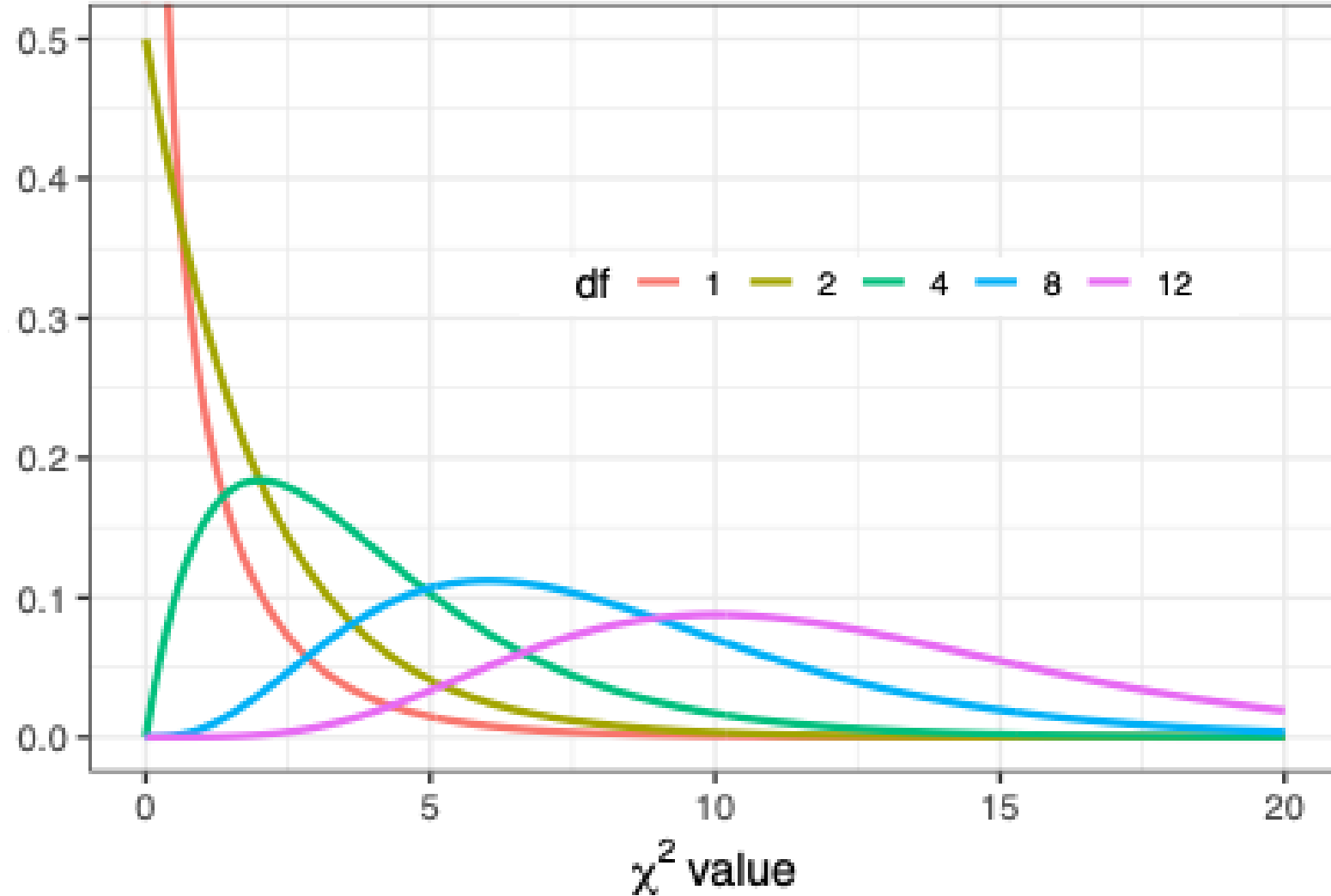
- [1] Finding a difference or goodness of fit 😊
- [2] Making it absolute
- [3] Standardise differences by expected value
- [4] Add them all up

Degrees of Freedom (k)

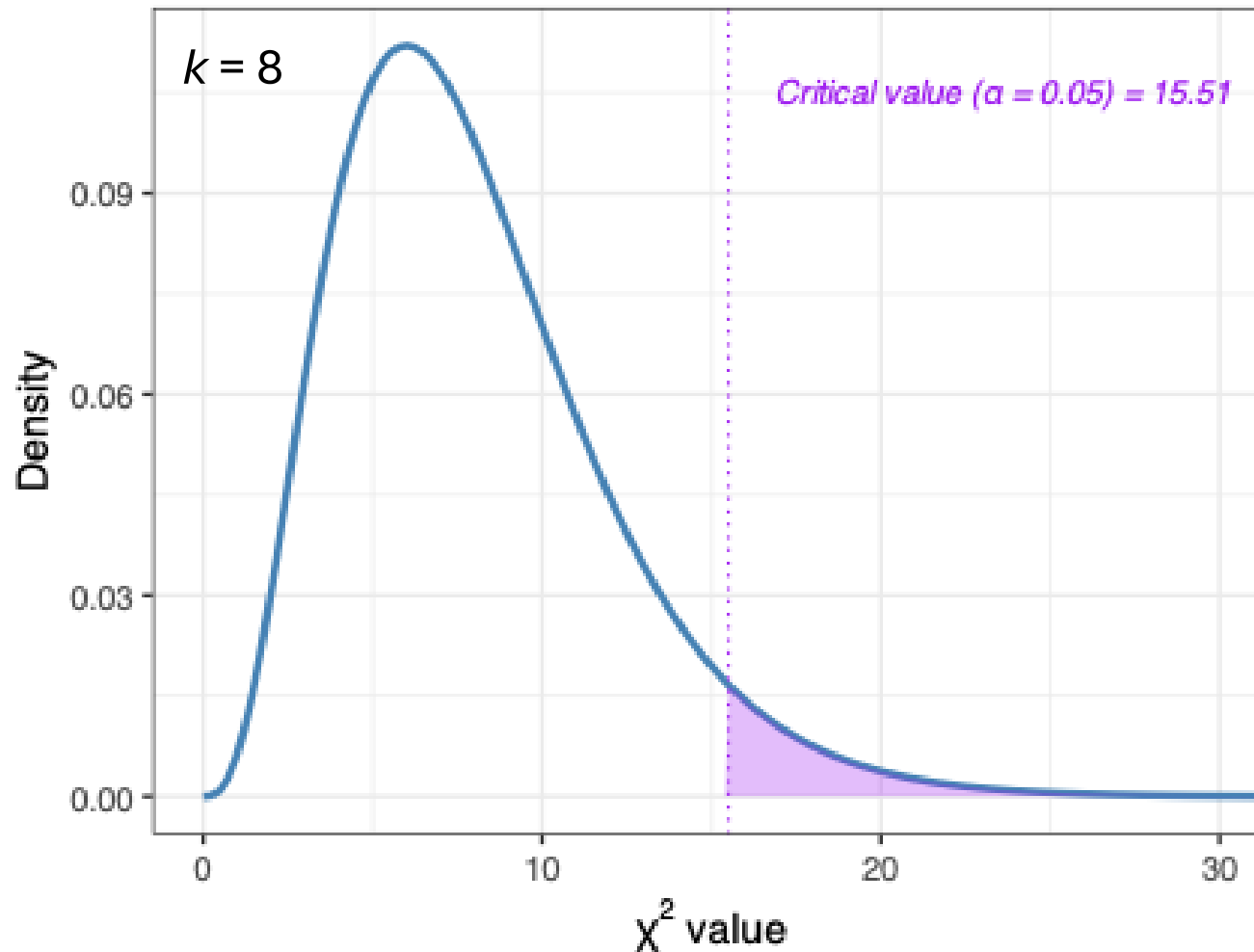
k = Number of categories - 1
= 9 brands - 1
= 8

Brand	Observed
Crunchie	1
Caramello	1
Dairy Milk	1
Moro	1
Flake	1
Boost	3
Cherry Ripe	2
Picnic	3
Turkish	2
TOTAL	15

- Independent, free information in data
- Determines shape of χ^2 distribution



Is what we observed extreme enough to cause a fuss?



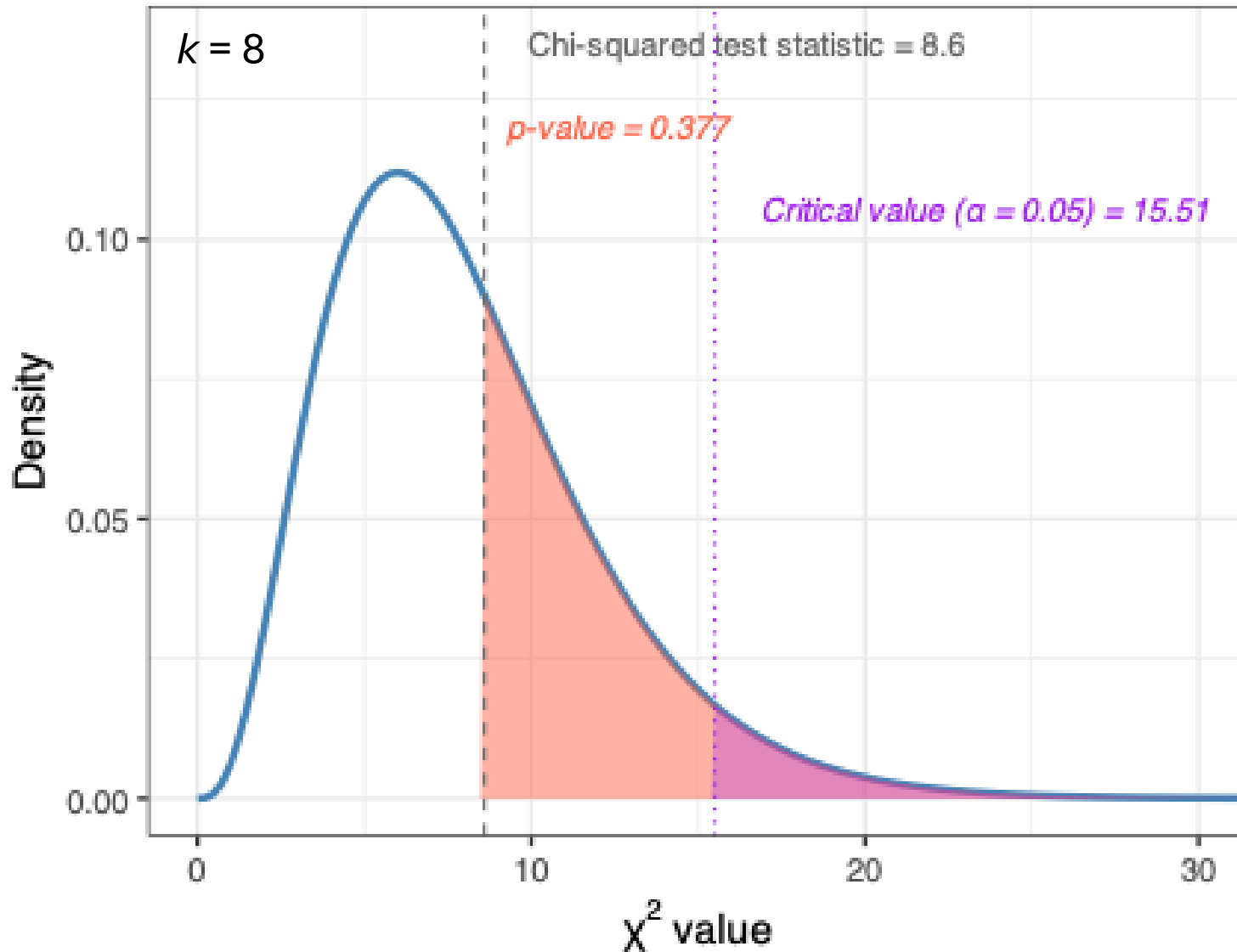
Significance level α

- how much probability we allow in the “rejection region”
- $\alpha = 0.05$
 - Balance of being too lenient vs. too cautious

Critical value

- A benchmark for us to determine whether our test statistic is extreme enough to reject the null.

Is what we observed extreme enough to cause a fuss?



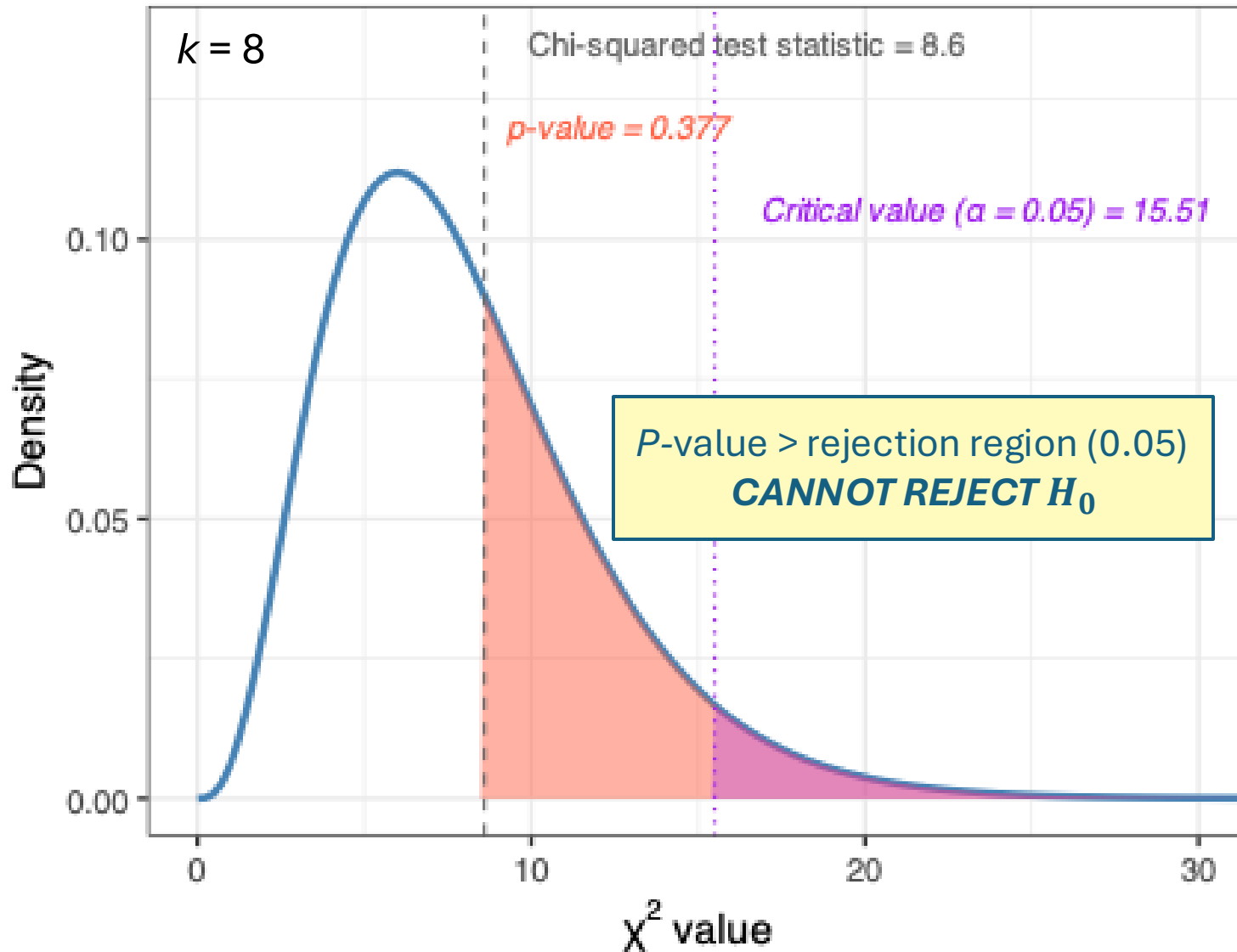
Compare our *test statistic* to the *critical value*

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

p-value 📄✎

- Probability of observing test statistic at least as extreme, if the H_0 were true

Is what we observed extreme enough to cause a fuss?



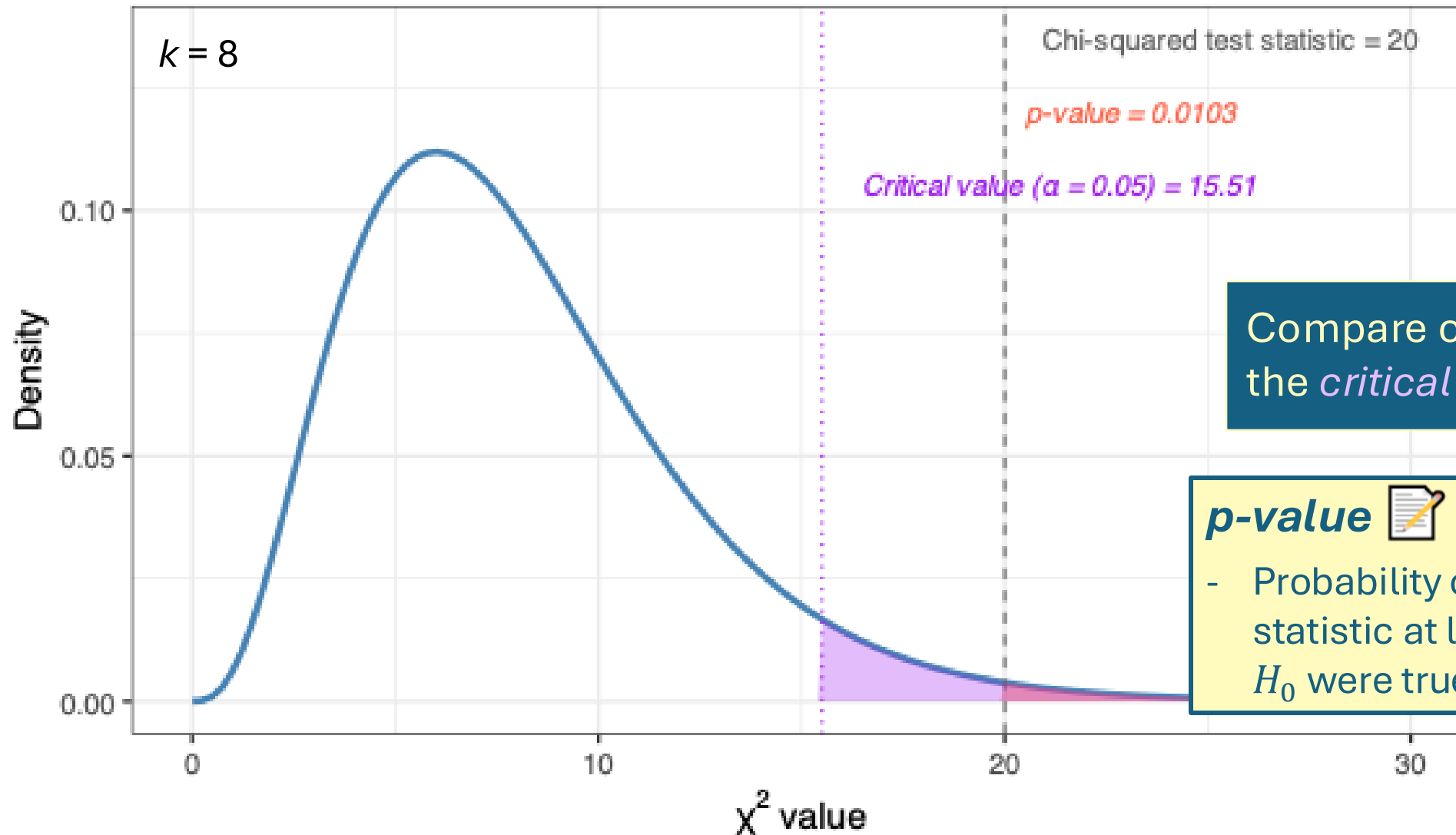
Compare our *test statistic* to the *critical value*

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

p-value 

- Probability of observing test statistic at least as extreme, if the H_0 were true

Is what we observed extreme enough to cause a fuss?



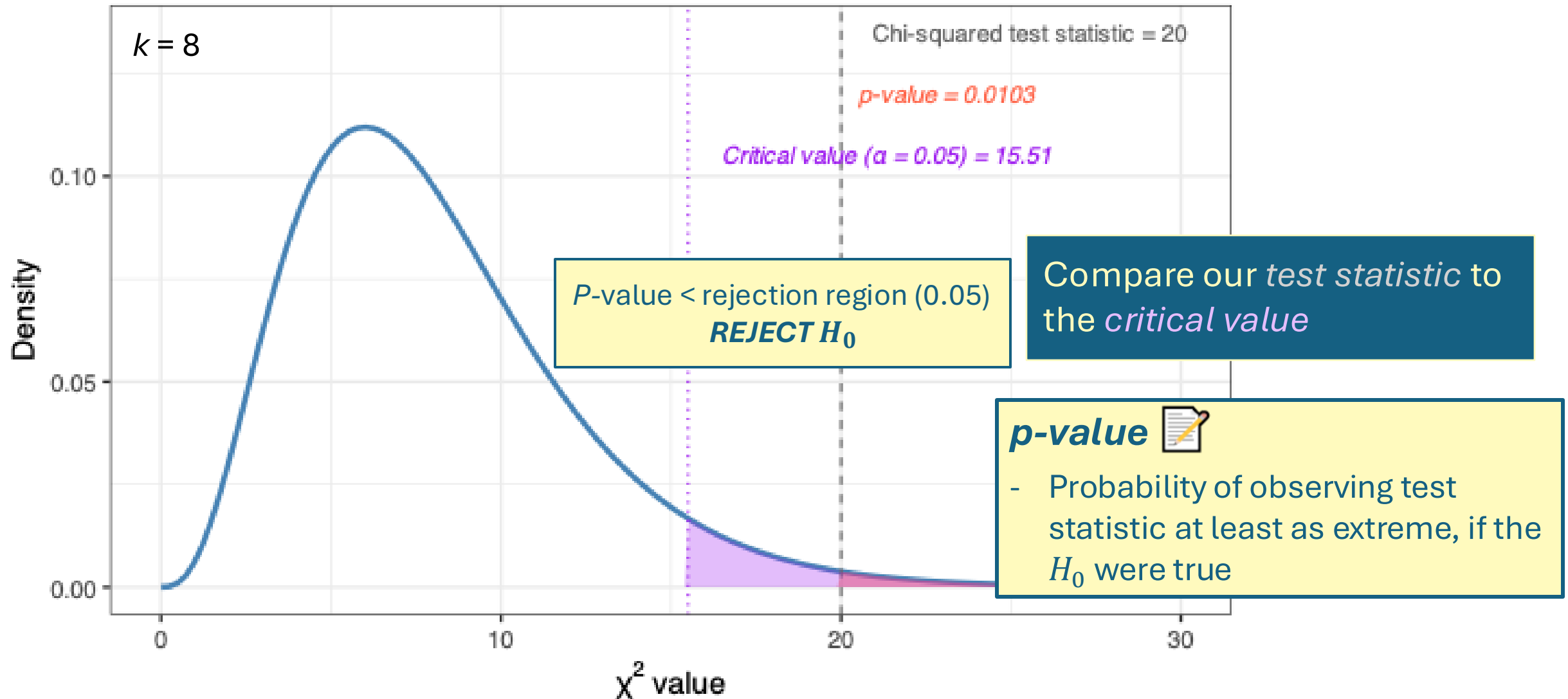
Compare our *test statistic* to the *critical value*

p-value



- Probability of observing test statistic at least as extreme, if the H_0 were true

Is what we observed extreme enough to cause a fuss?



Put it to practice

H_0 = Equal chance of drawing any brand of chocolate

H_1 = **Unequal** chance of drawing any brand of chocolate

Chi-squared
test statistic

observed value

expected value

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

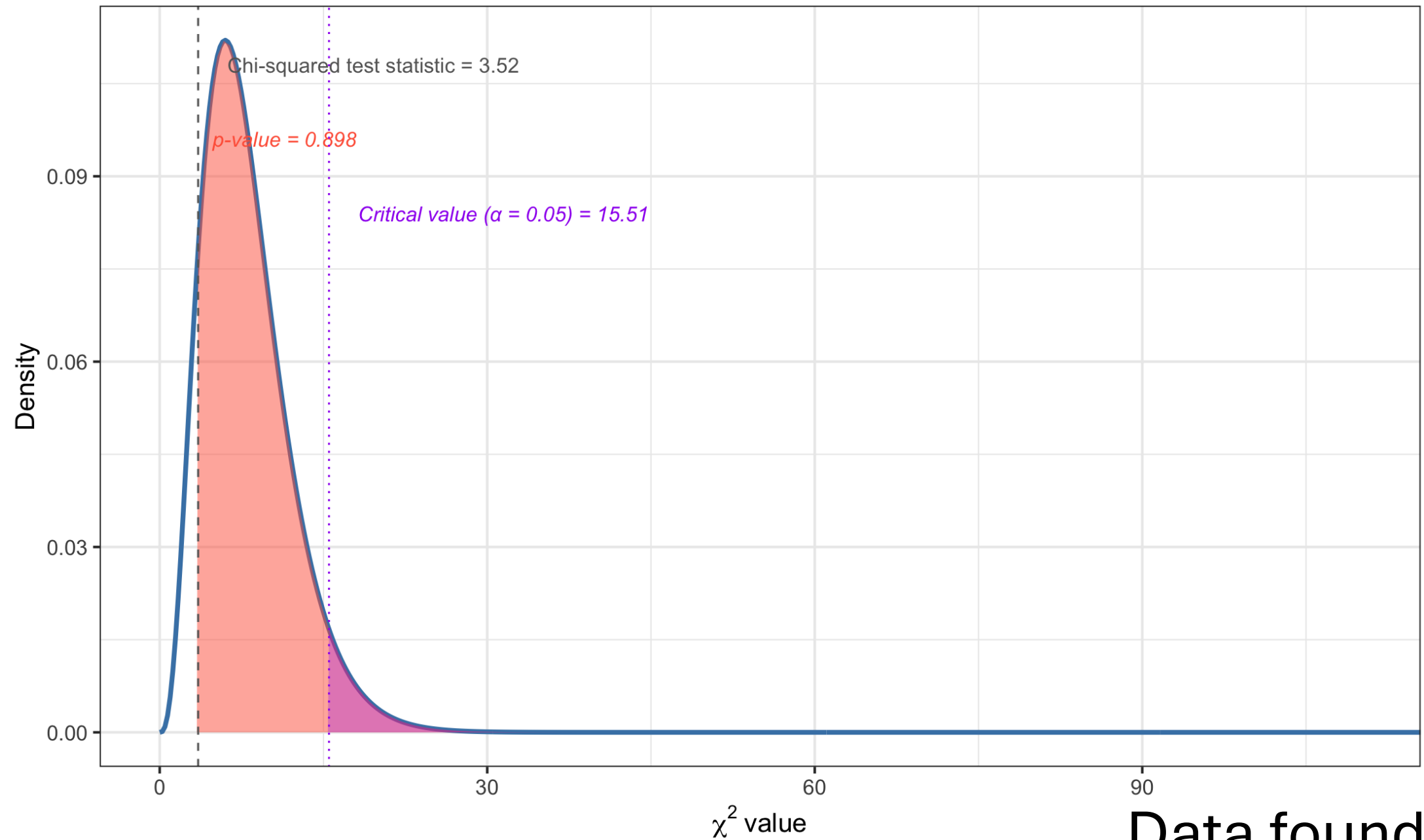


Let's have a go! 🤪 (10 mins)

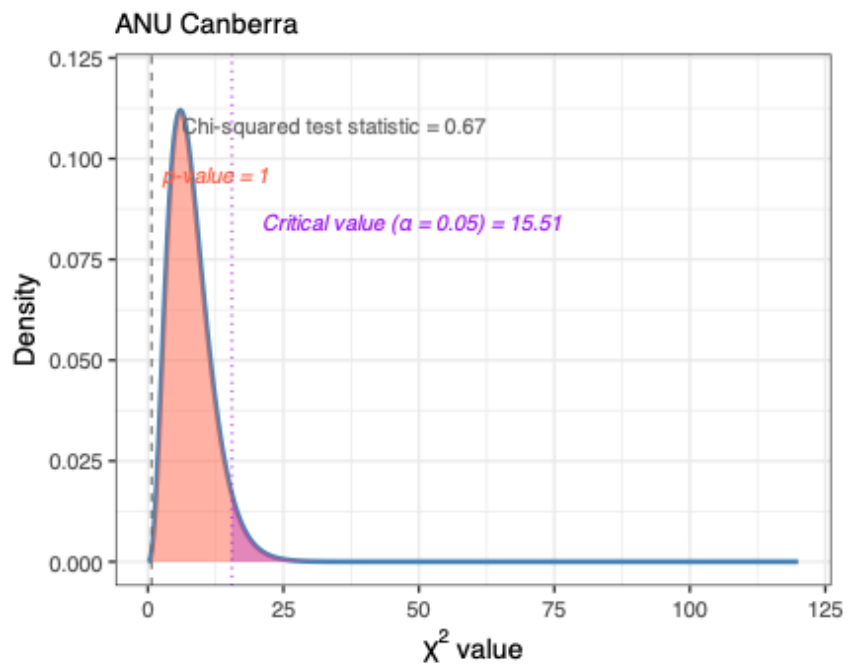
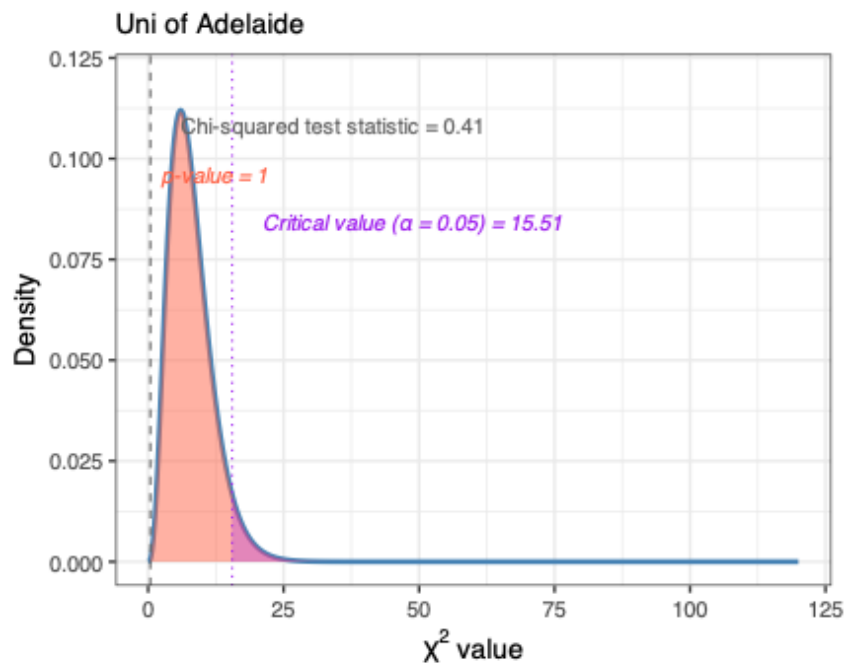
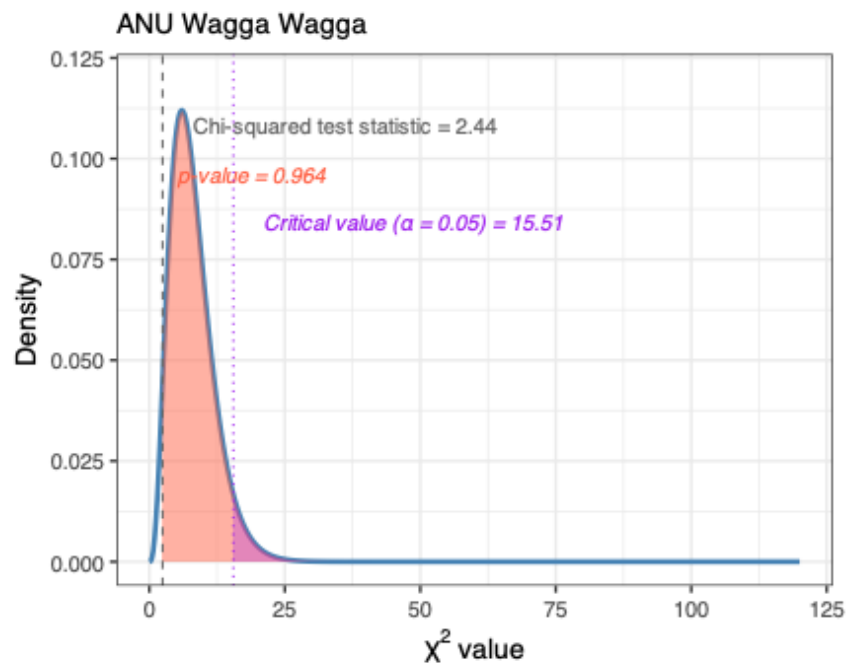
Open your boxes and enter the data!

Data found [here](#)

Overall data for entire training session



Data found [here](#)



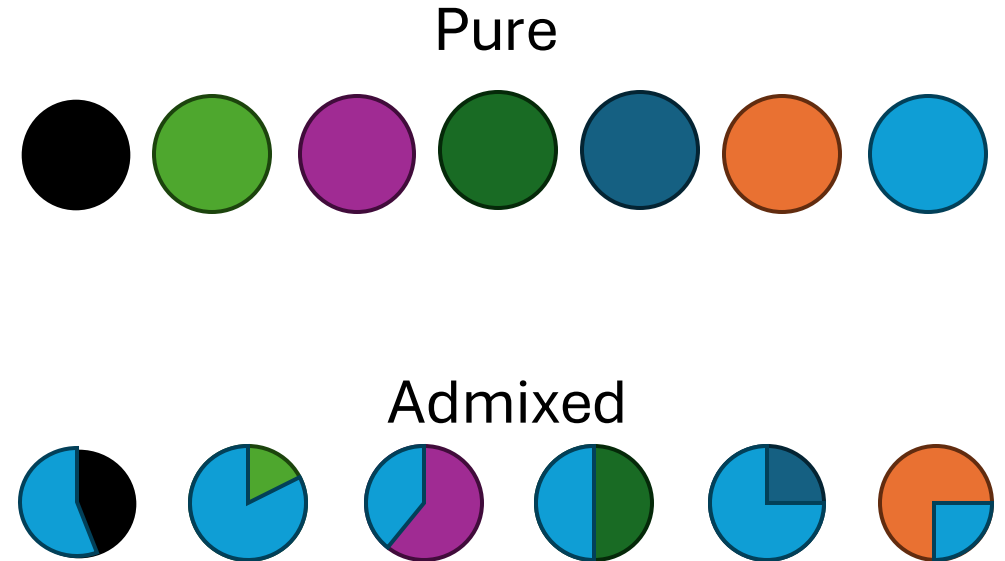
Data found [here](#)



ZOOMAGRI

methods

- 10 devices
- 7 barley varieties
- 24 samples
 - 12 x Pure
 - 12 x Admixed



What is the **efficacy** and consistency of the ZoomAgri devices in detecting the correct barley varietal(s)?

Testing at different levels

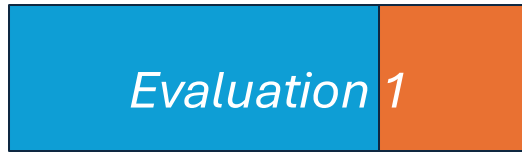
By evaluation

Expected



Observed

Device 1



Device 2



Device 3



Device...d



By sample

Expected



Testing at different levels

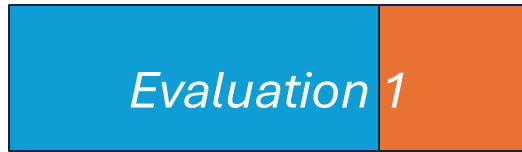
By evaluation

Expected



Observed

Device 1



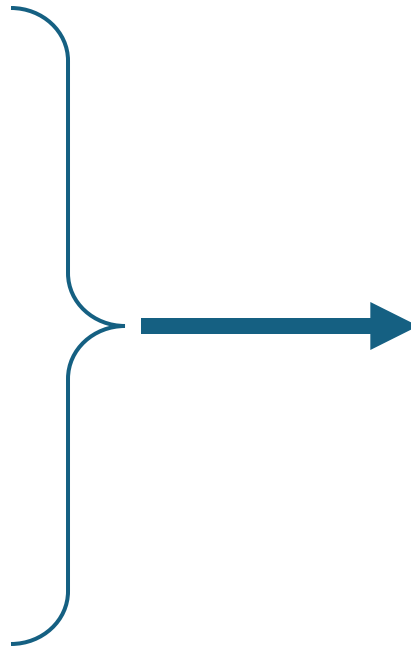
Device 2



Device 3



Device... d

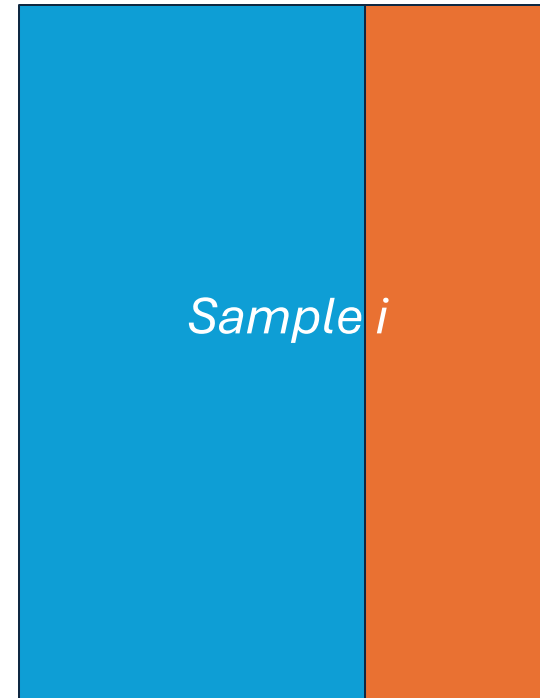


By sample

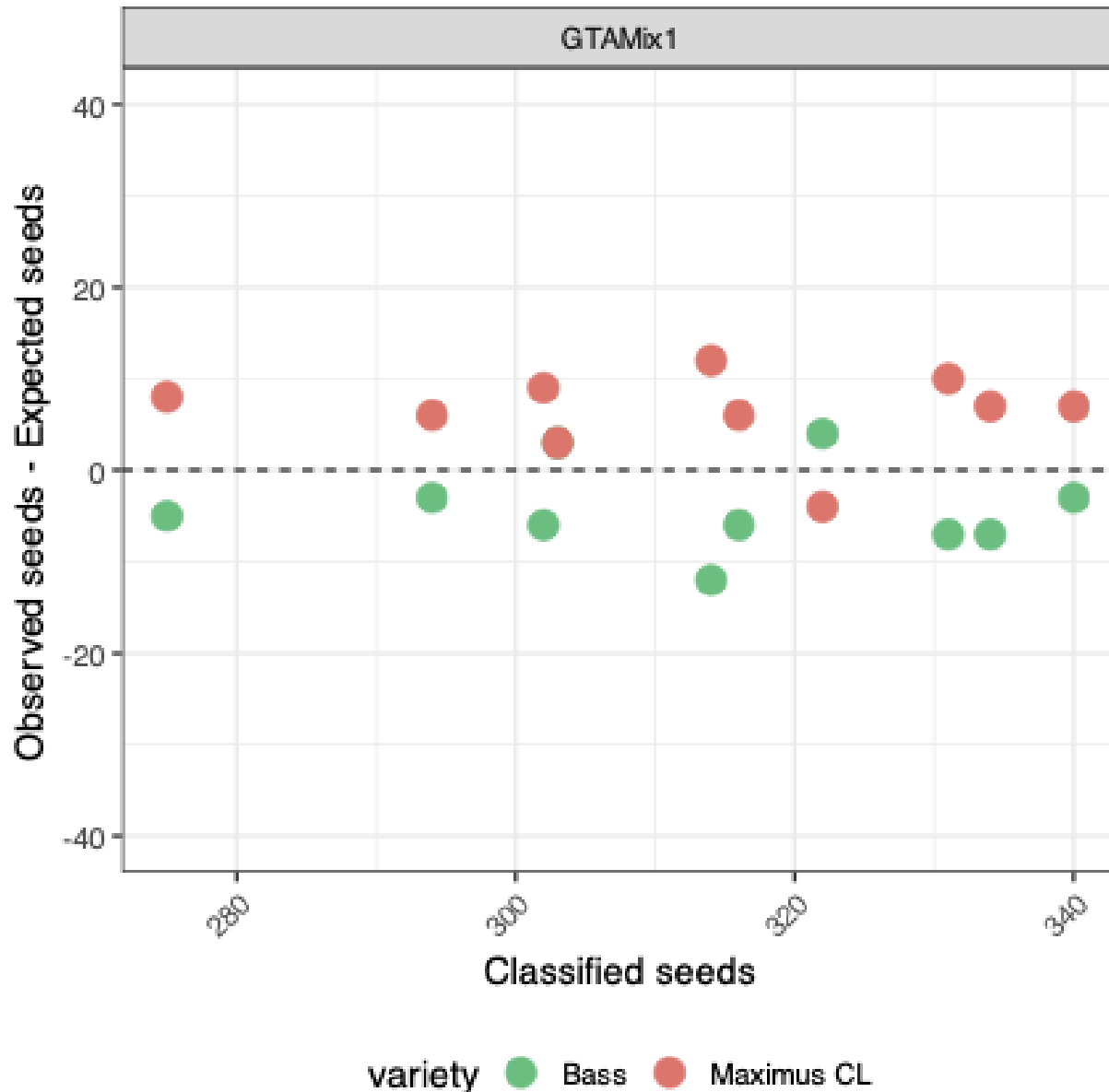
Expected



Observed



Let's visualise the differences



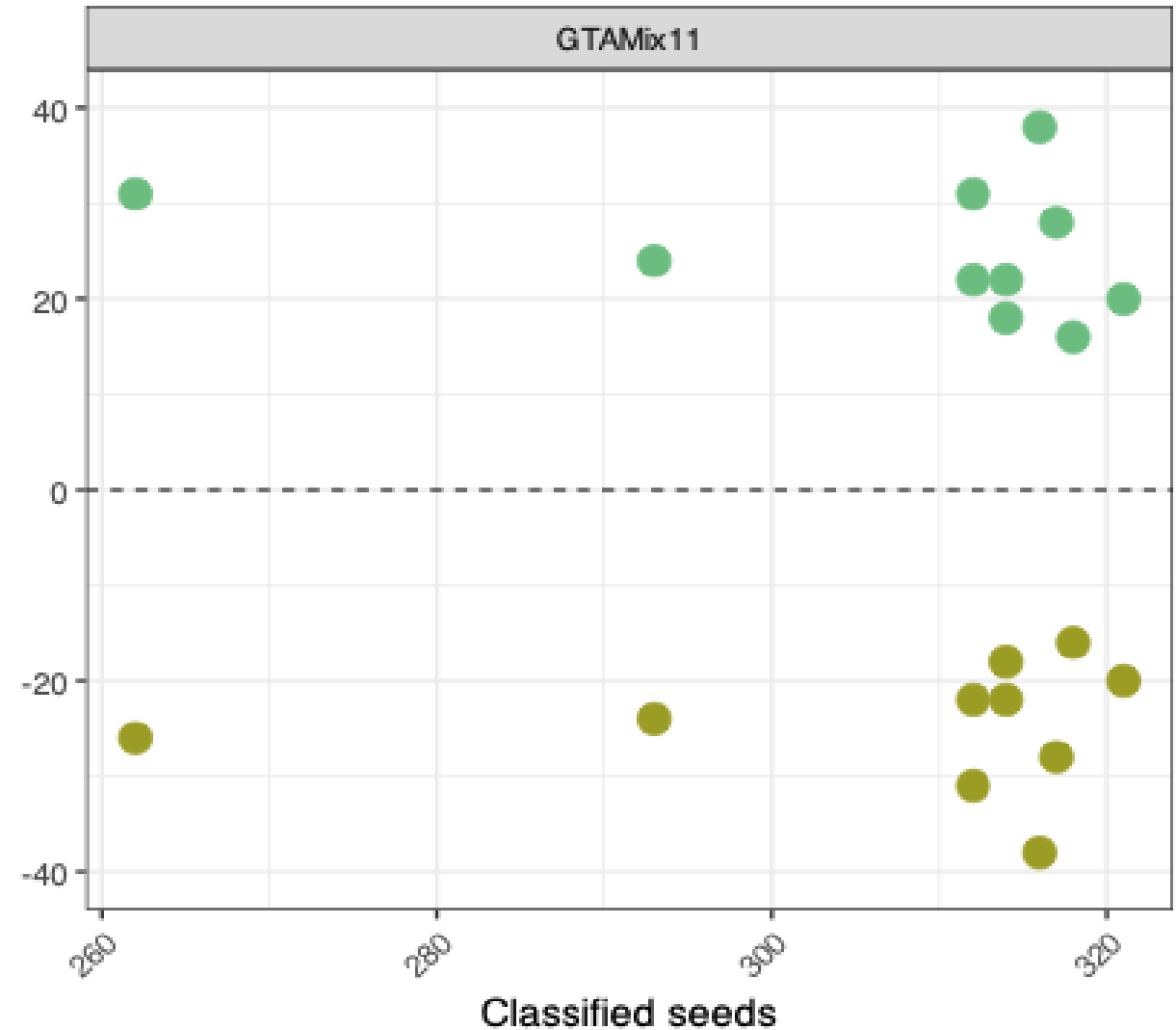
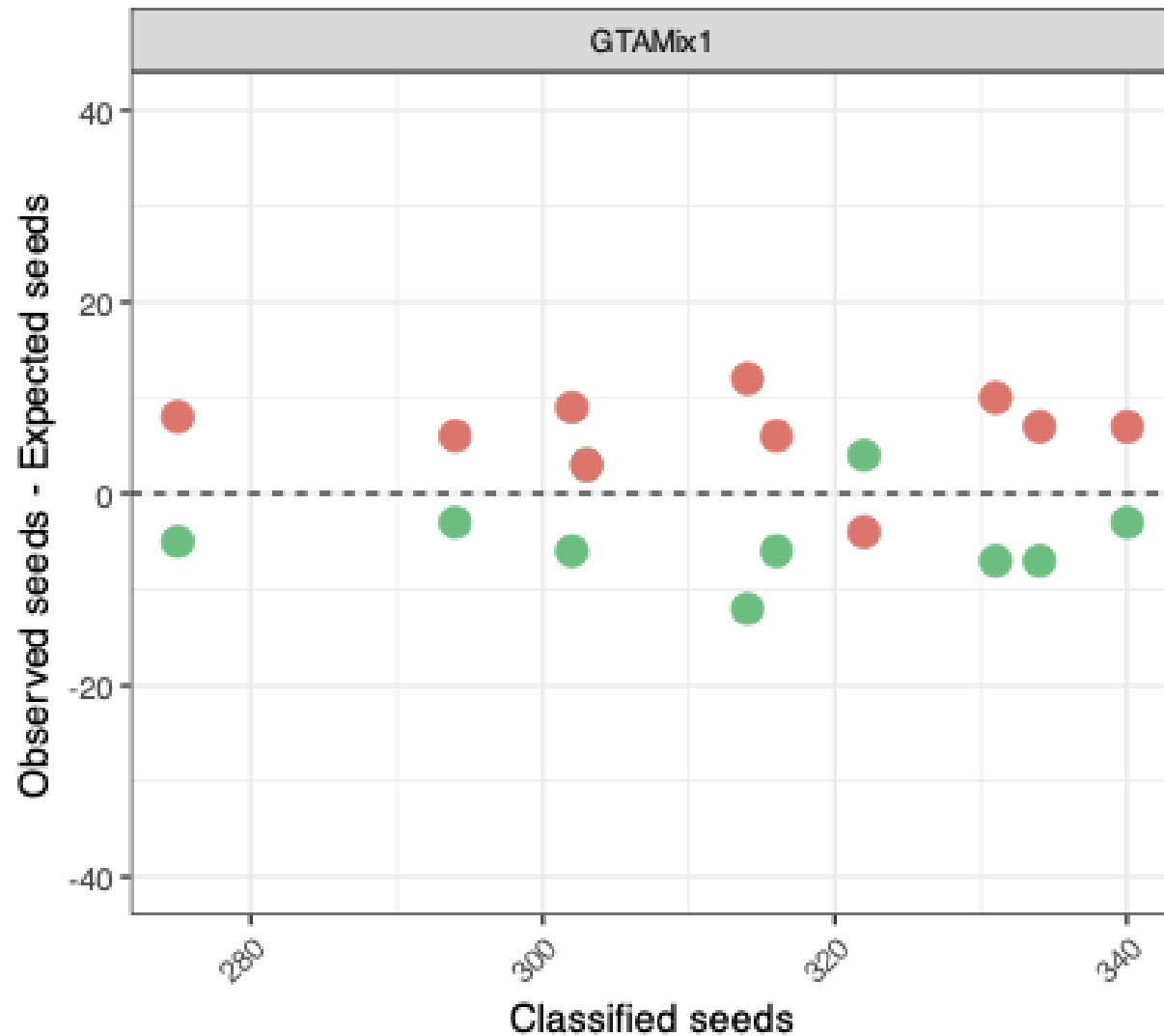
Data viz 🎨

- Middle, **horizontal line** is our **benchmark**
- **Distance** from middle is relative to **discrepancy**
- **Scatterplot** to show how **discrepancy** **changes** with classified seeds

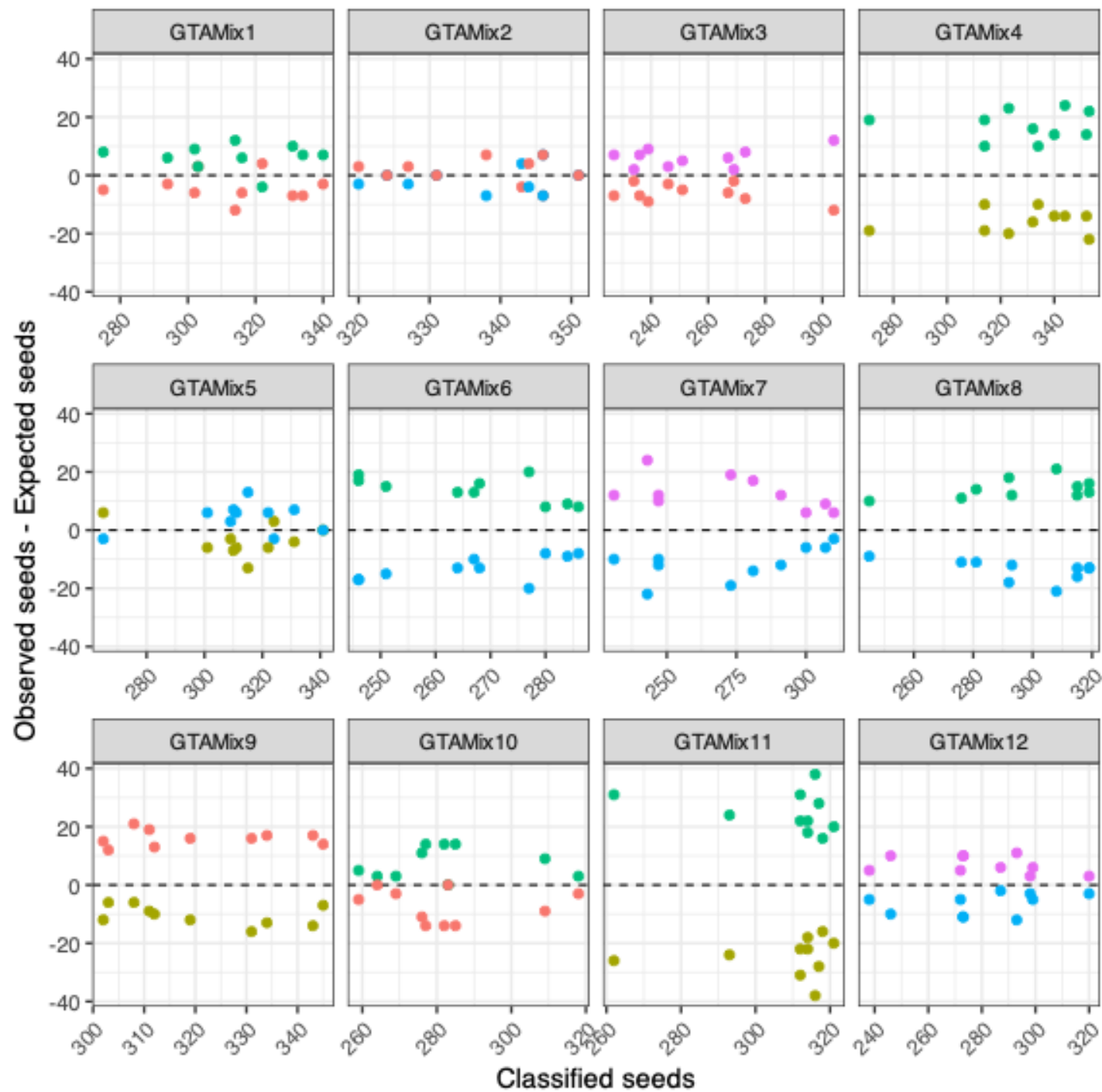
$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

What would their χ^2 look like?

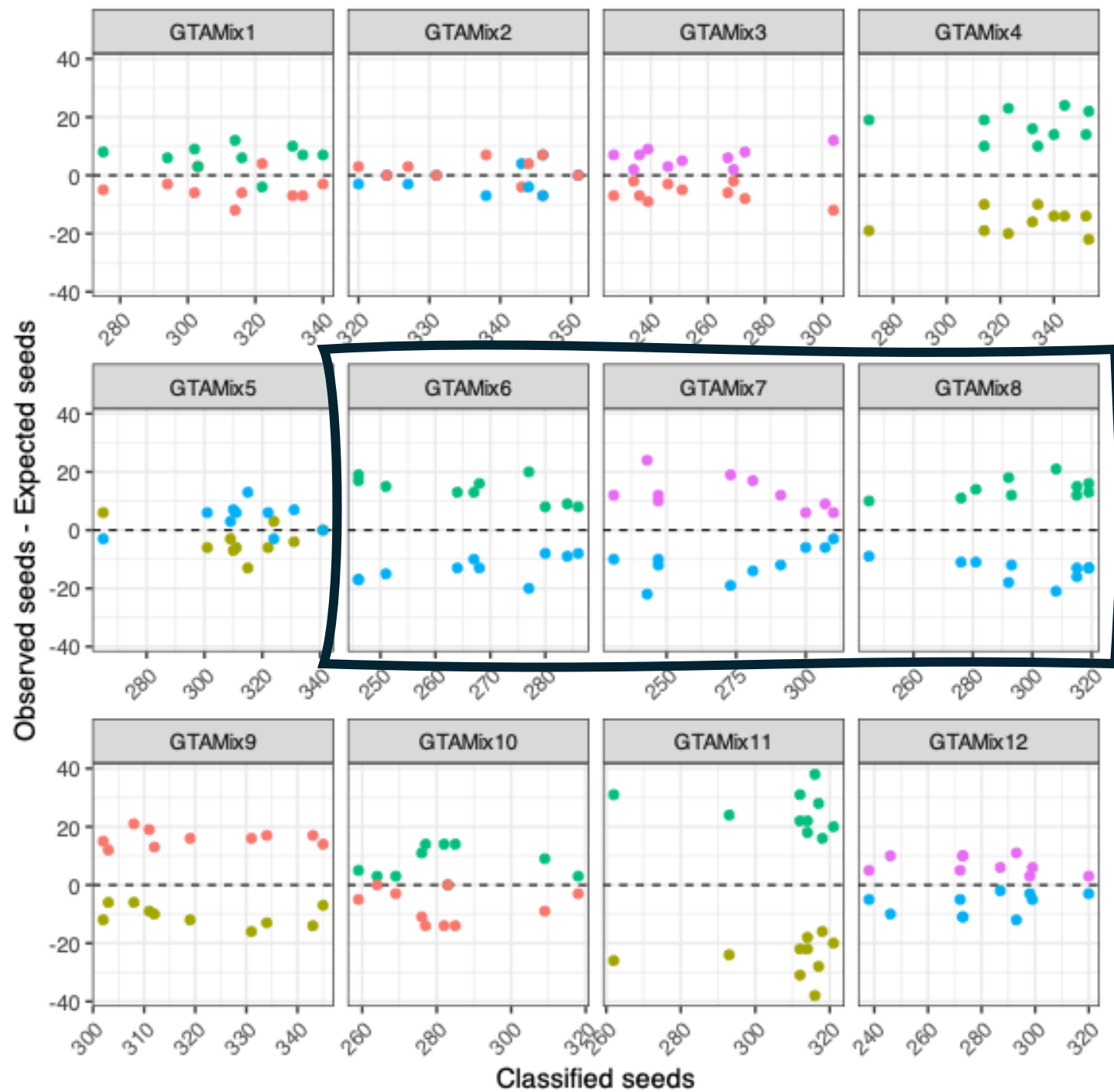
$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$



Across all samples

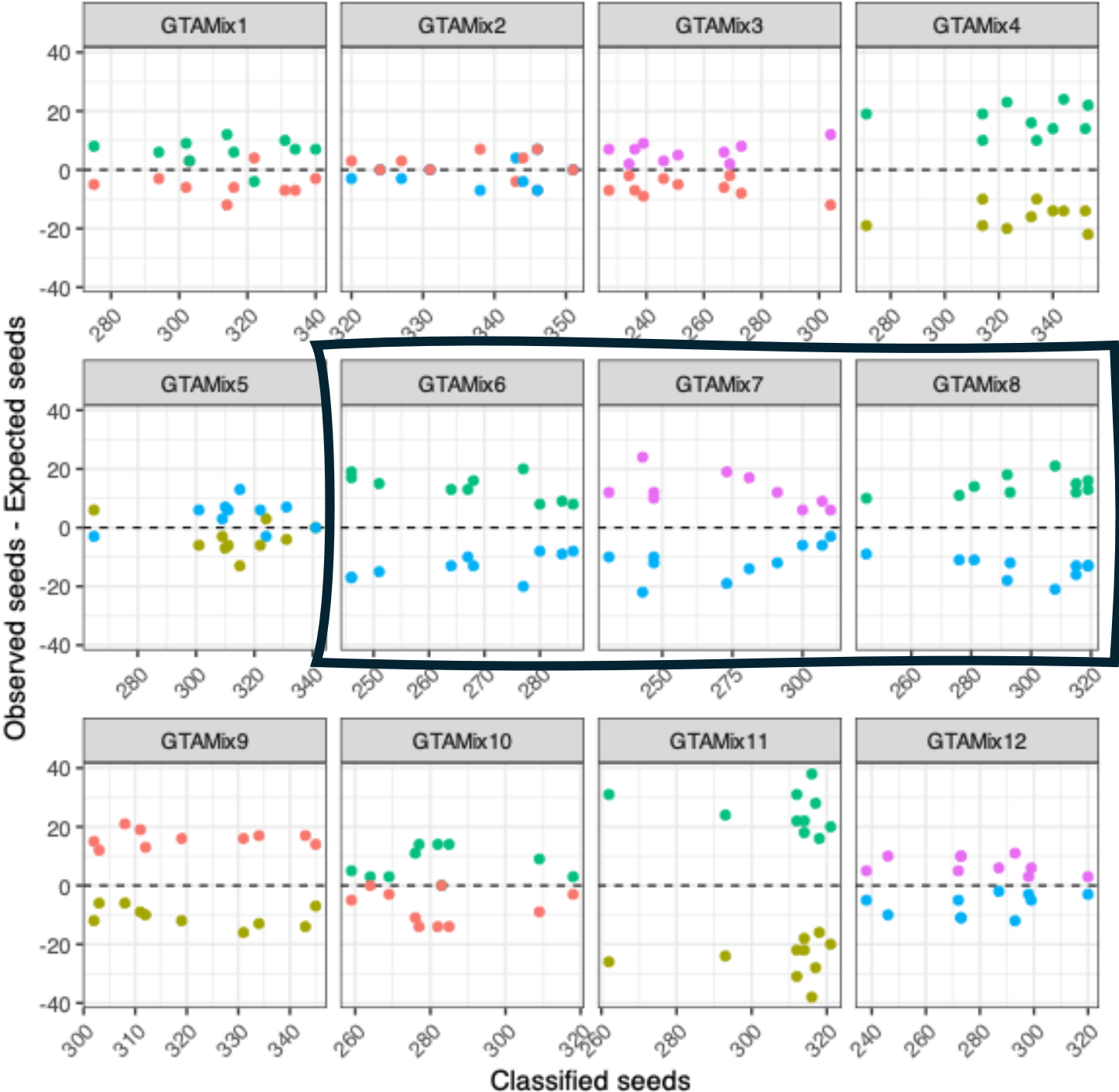


Across all samples



Across all samples

Sample code	Variety mixture	X ²	df	p-value
GTAMix5	La Trobe (75%), Planet (25%)	6.82	10	0.742
GTAMix6	Maximus CL (75%), Planet (25%)	38.82	10	<0.001 ***
GTAMix7	Spartacus CL (70%), Planet (30%)	31.12	10	<0.001 ***
GTAMix8	Planet (70%), Maximus CL (30%)	33.10	10	<0.001 ***



summary

- Test statistics are used to bridge data to statistical theory
- Significance, critical value and P -value are closely interrelated
- Cadbury was right. (unfortunately)