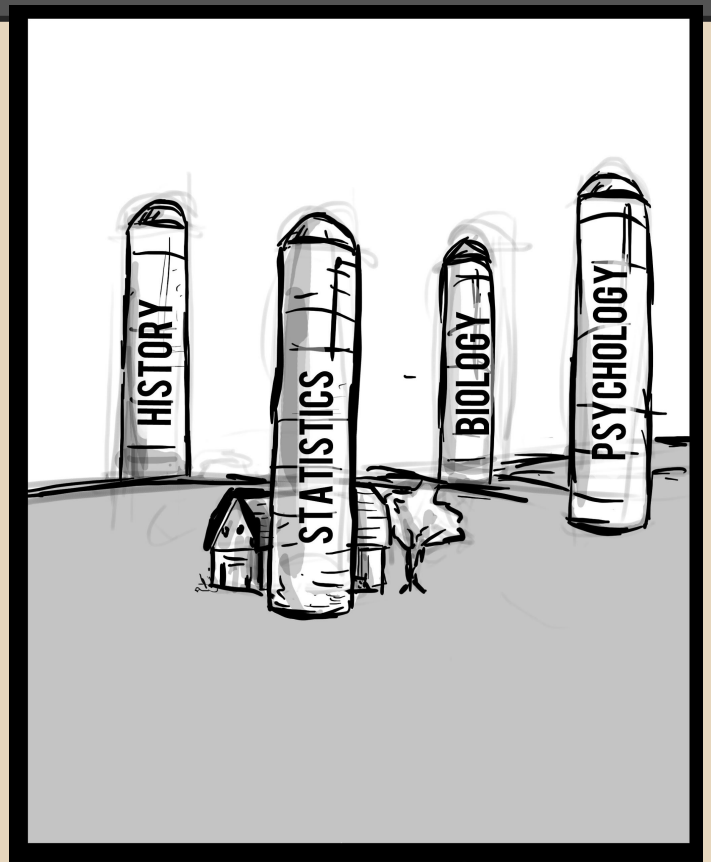
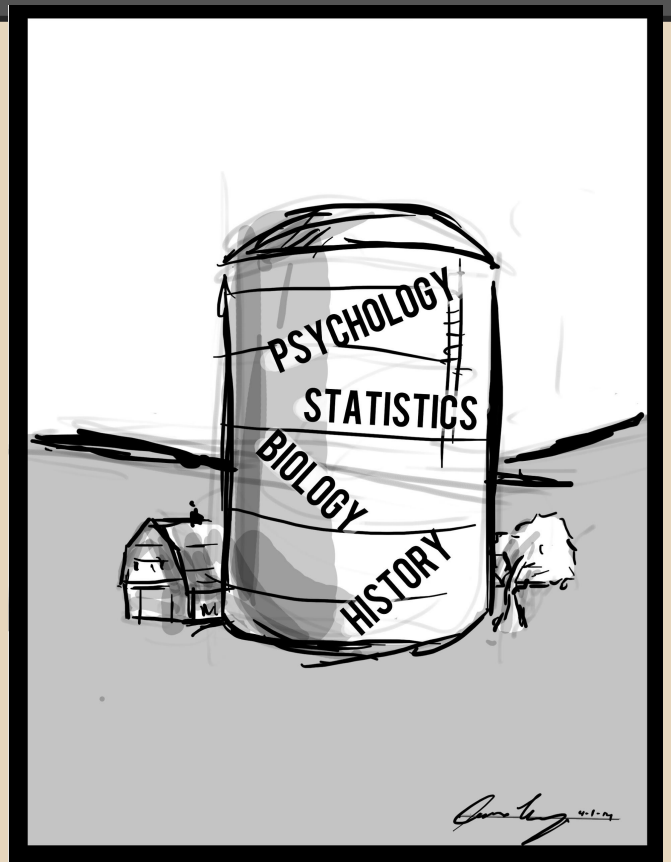
Doing Statistics with Real Biology Data

Luke Wilcox Kentwood, Michigan

How is knowledge stored?





Biologists crossed a large random sample of heterozygous red eye, normal winged fruit flies (RrNn with RrNn). They believe that the phenotypic ratio should be 9:3:3:1.

Phenotypic Results:

Red Eyes and Normal Wings = 731
Red Eyes and Vestigial Wings = 241
Sepia Eyes and Normal Wings = 229
Sepia Eyes and Vestigial Wings = 63

Does the data provide significant evidence that the biologists' predictions are wrong?

Does the data provide significant evidence that the biologists' predic

Null Hypothesis: There is no significant difference observed and expected values

Total # of Flies = 731 + 241 + 229 + 63 = 1264

phenotype	observed	expected	(o-e)	$(o-e)^2$	$(o-e)^2/e$
red/Normal	731	812.57	-81.57	6653.66	8.19
red/Vestigial	241	270.86	-29.86	891.62	3.29
sepia/Normal	229	270.86	-41.86	1752.26	6.47
sepia/vestigial	63	90.29	-27.29	744.74	8.25

+

$\chi^2 = 26.2$

$\text{map units} = \frac{241 + 229}{1264} = .371$

$df = 3 \rightarrow 7.82 \leq 26.2$

The calculated χ^2 value of 26.2 is greater than 7.82, so we reject H_0. The variation in our data couldn't have happened purely by chance. The genes are linked and are .371 map units away from each other.

CHI-SQUARE TABLE

	Degrees of Freedom							
p	1	2	3	4	5	6	7	8
0.05	3.84	5.99	7.82	9.49	11.07	12.59	14.07	15.51
0.01	6.64	9.32	11.34	13.28	15.09	16.81	18.48	20.09

1269

Does the data provide significant evidence that the biologists' predictions are wrong?

State: H_0: the claimed distribution for fly phenotypes (9:3:3:1) is correct ($\chi^2 = 0$)

H_a: the claimed distribution for fly phenotypes is incorrect ($\chi^2 > 0$)

$\alpha = .05$

Plan: χ^2 GOF test

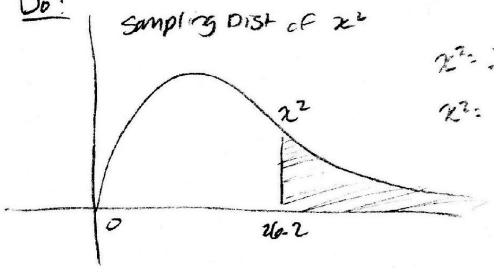
✓ ① Random → "large random sample"

✓ ② Independent → $n \leq \frac{1}{10}N$ (large #) $\leq \frac{1}{10}$(larger #)

individual observations independent

✓ ③ Large Sample Size → smallest expected value = 90.29 ≥ 5

Do:



$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$\chi^2 = \frac{(781-812.57)^2}{812.57} + \frac{(241-270.86)^2}{270.86} + \frac{(229-270.86)^2}{270.86} + \frac{(63-90.29)^2}{90.29}$$

$$\chi^2 = 8.19 + 3.29 + 6.47 + 8.25$$

$\chi^2 = 26.2$ P-value: 8.066×10^{-6}

Conclude:

Assuming H_0 true ($\chi^2 = 0$), there is a 0 probability of getting a χ^2 value of 26.2 or more purely by chance. This provides good evidence against null and is statistically significant @ $\alpha = .05$. Therefore we reject H_0 + can conclude the claimed distribution is incorrect. The largest component of χ^2 is 8.25 because the observed # of sepia, vestigial flies is much lower than expected.

Content Goals for the lesson

- To able to formulate appropriate null and alternative hypothesis for a chi-square goodness of fit (Punnett Squares)
- To be able to check the appropriate conditions for performing inference.
- To be able to perform a chi-square goodness of fit test and make a conclusion.

Mathematical Practices Goals

Standard 1: Make sense of problems and persevere in solving them

Standard 2: Reason abstractly and quantitatively

Standard 3: Construct viable arguments and critique the reasoning of others

Standard 4: Model with mathematics

Standard 5: Use appropriate tools strategically

Standard 6: Attend to precision

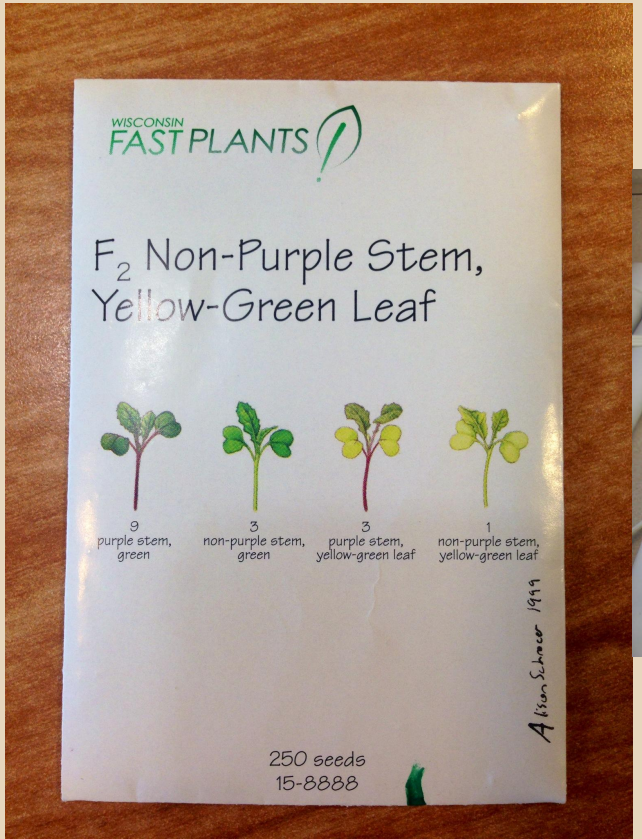
Standard 7: Look for and make use of structure

Standard 8: Look for and express regularity in repeated reasoning

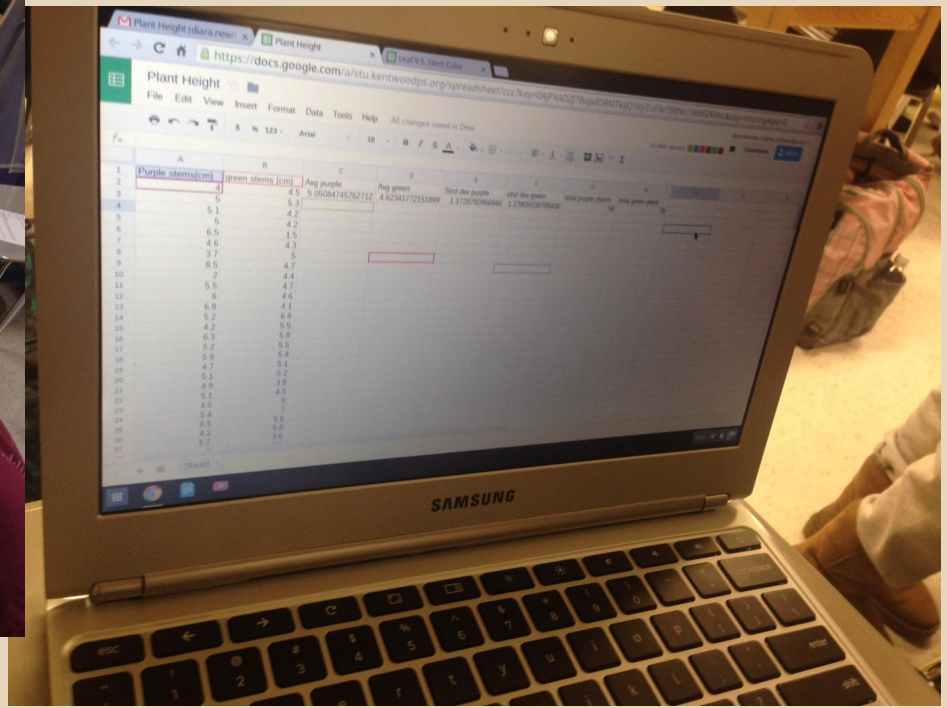
Lesson Plan

- Get groups with plants in front of them.
- Collect data.
- Groups work to write a significance test.
- Students present whiteboard solutions.
- Teacher wrap up.

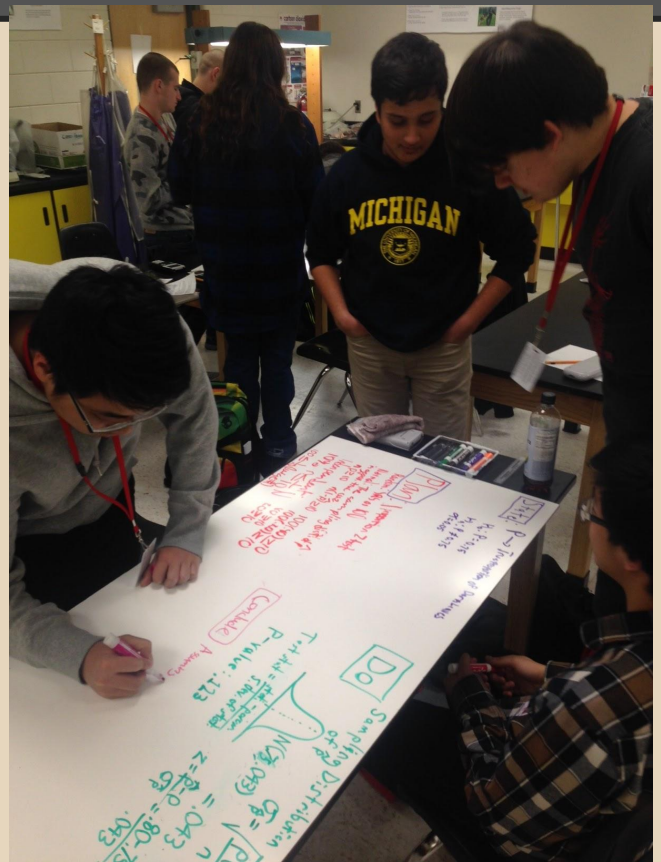
Get groups with plants in front of them

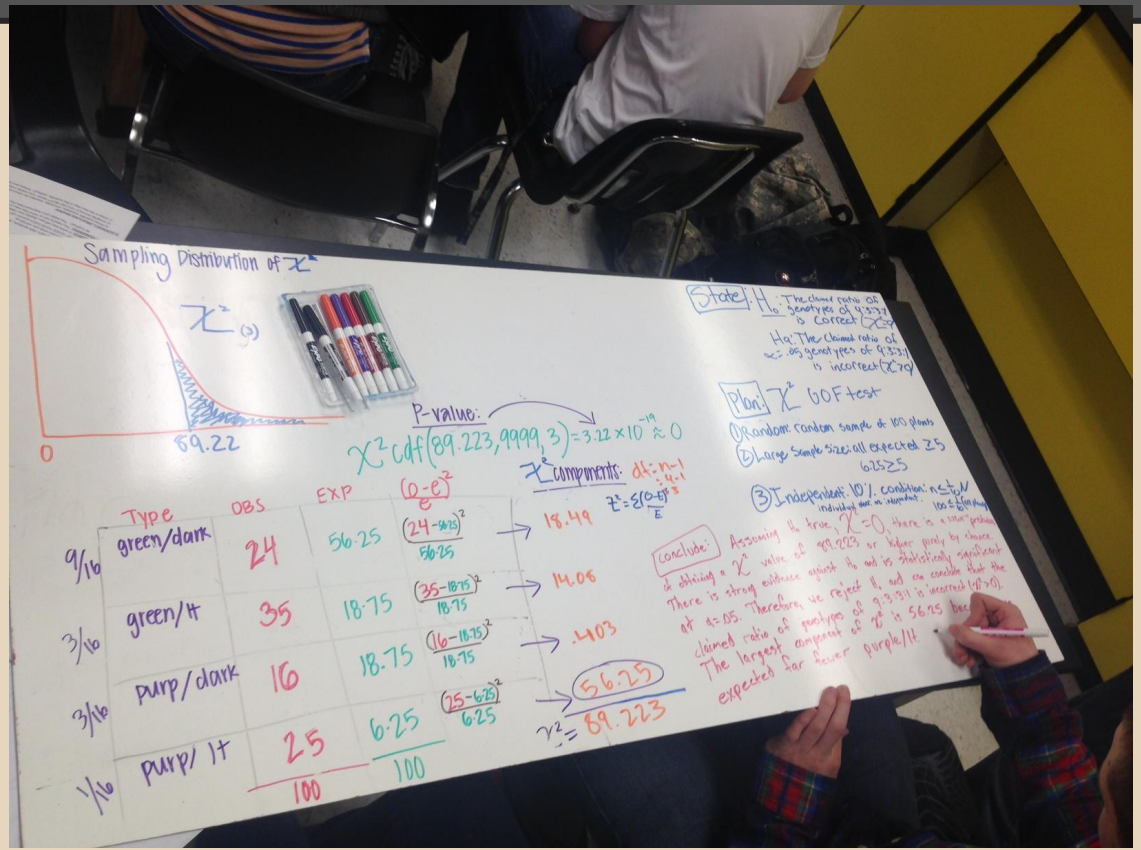


Collect Data



Groups work to write significance test





Present white board solutions

Teacher wrap up

What did they learn?

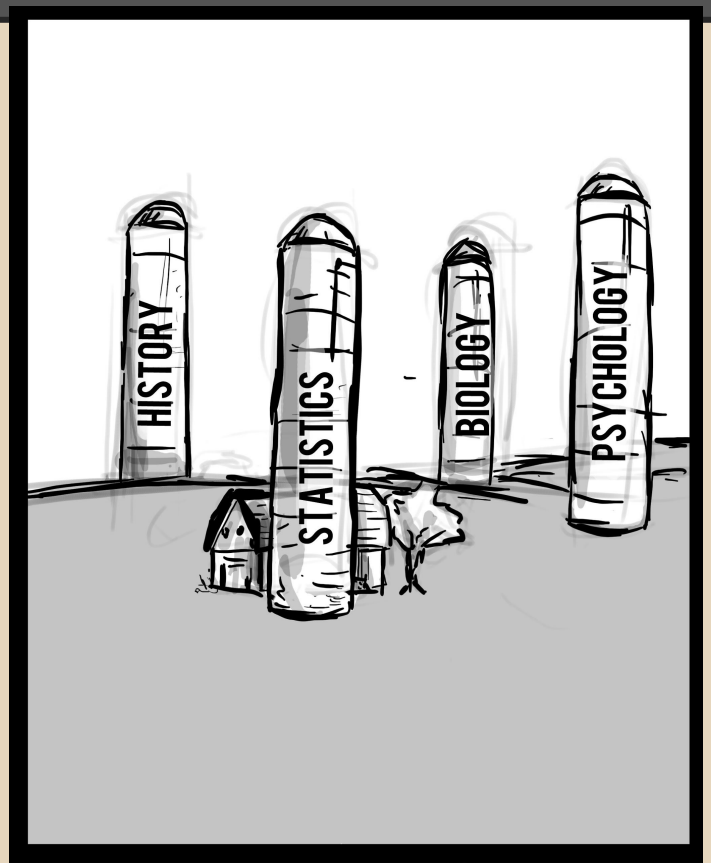
AP Statistics students:

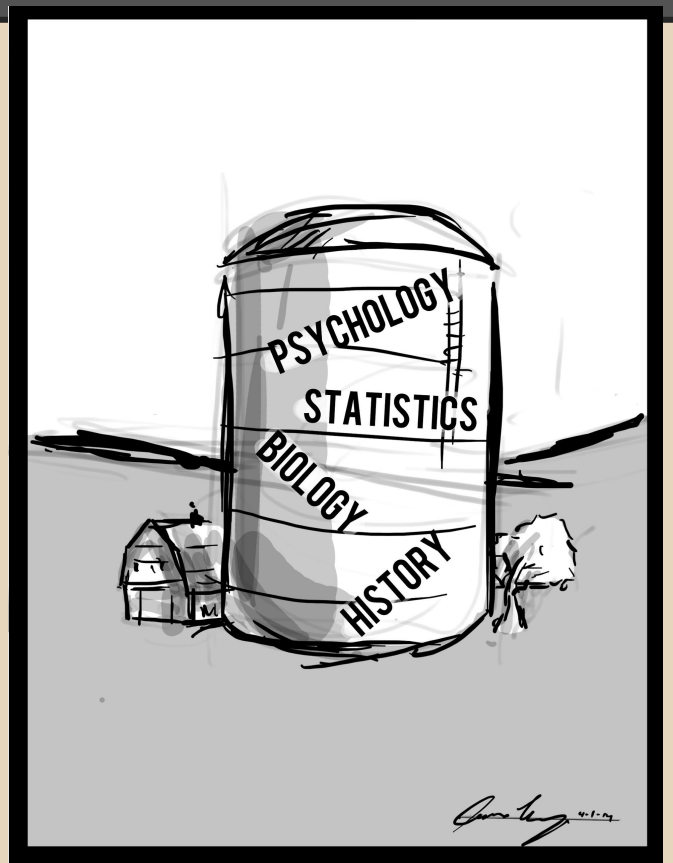
- Importance of controlling lurking variables when collecting data.
- Punnett squares and rules for probability lead to hypotheses.
- Relationship between chi-square test statistic and P-value.

AP Biology students:

- Structure of a significance test.
- Meaning of a P-value.
- Importance of checking conditions for inference.
- Follow up analysis.

What did they learn?





Extension. Inquiry Based Instruction.

- (1) What question do you want to answer?
- (2) How will you collect the data?
- (3) What kind of significance test can you use to support your conclusion?
    - t-test for claimed height
    - 2 sample t-test for heights of purple/green stem
    - 1 proportion z-test for distribution of leaf color

Questions/Comments?

- Powerpoint is available through NCTM website.
- Feel free to email me at luke.wilcox@kentwoodps.org