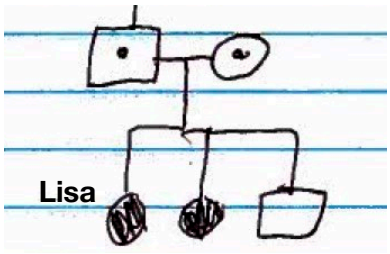
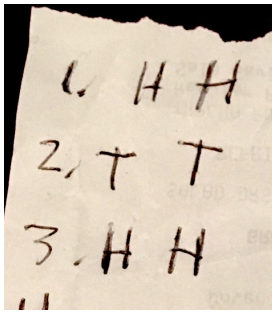
Chi-Square Tests & the Problem of Small Sample Sizes Guide by: Chris Gonzalez



Two heterozygous parents (A/a) produce three children, as shown in the pedigree. This monohybrid cross gives us an **expected** 3:1 phenotype ratio. However, consider the following question:

"If Lisa's mom and dad are carriers, how do both her and her sister have the recessive trait if only 1/4 of the parent's offspring should have it?"

The short answer is that our sample size is too small to accurately represent the expected ratio. What does this actually mean?



Consider a (real!) experiment I did, in which I flipped a coin 2 times and recorded the results. Just like homozygous recessives, the chance of getting 2 heads (H-H) in a row is (1/2)(1/2) = 1/4 (or 25%). This **expectation** is our null hypothesis.

Notice my results though — in my first 3 experiments, I got about 66% H-H! Does this mean our 1/4 expectation is wrong? Or is random chance impacting my results?

This concern is EXACTLY why we have Chi-Square tests.

Total: 3 flips (about 66% H-H)

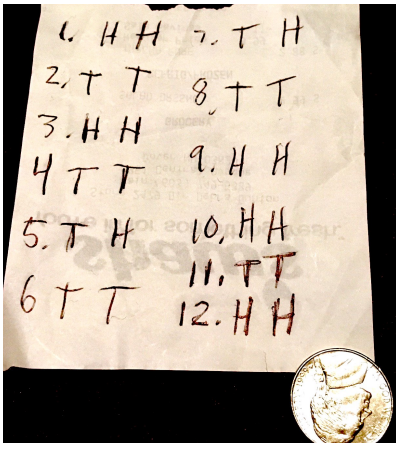
	Expected (E)	Observed (O)	(O-E)^2	((O-E)^2) / E
H-H	1	2	1	1
other combos	2	1	1	0.5
SUM				1.5
X^2 VALUE				1.5

df = 2 - 1 = 1
P value = 0.30—0.20 (using an online calculator, P value = 0.220671)
— There is about a **22% probability** that the differences between my E and O values are due to random chance, which is much higher than 5%. Thus there is **NO significant difference**, and we **do not reject our null hypothesis.** The expected 1/4 calculation is still supported.

What if we conduct 12 coin-flip experiments?
Total: 12 experiments (about 42% H-H)

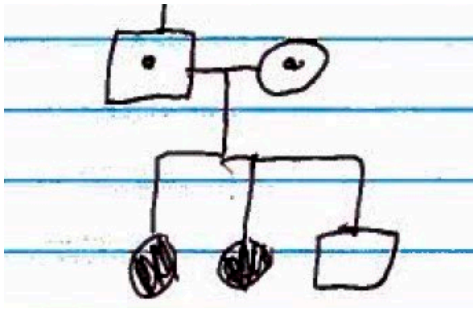
	Expected (E)	Observed (O)	(O-E)^2	((O-E)^2) / E
H-H	3	5	4	1.333
other combos	9	7	4	0.444
SUM				1.777
X^2 VALUE				1.777

df = 2 - 1 = 1
P value = 0.20—0.10 (using an online calculator, P value = 0.182518)
— there is still **NO significant difference** between the expected and observed values; random chance is again biasing our results!



However, notice that with 12 experiments, the probability that random chance is impacting our results decreased from 22% to 18%

Consider again our monohybrid cross. Each parent has a 1/2 chance of passing on their recessive allele, and thus a 1/4 chance of having a homozygous recessive child.



However, each child is equivalent to one coin-flip experiment. As we saw above, 3 experiments (or even 12!) are not enough to prevent random chance from skewing our results, and giving us numbers we wouldn't expect based on our null hypothesis.

We must remember that with small sample sizes, outcomes that don't meet our expected probability assessment are common. Our probabilities tell us what is MOST LIKELY to happen, not what ALWAYS WILL happen.

This is why Mendel's Pea Plant experiments are so powerful; he conducted **over 20,000 crosses**, greatly decreasing the impact of random chance and revealing the actual probabilities that underly unlinked gene inheritance.

Table 2.4 The Chi-Square Table

	Probability (*P*) Value									
df	0.95	0.90	0.70	0.50	0.30	0.20	0.10	0.05	0.01	0.001
1	0.004	0.016	0.15	0.46	1.07	1.64	2.17	3.84	6.64	10.83