Chi-Squared test formal procedure

► Create null and alternative hypotheses:
  H_0: Observed counts match predicted counts.
  H_A: Observed counts don't match predicted counts.

► Calculate X^2 test statistic: $X^2 = \sum_{i=1}^{k} \left[\frac{(O_i - E_i)^2}{E_i} \right]$

► Compare X^2 to various χ^2 values (α=0.05).

► Determine probability, **p value**, of seeing X^2 as large as we do.

► Decide to "reject H_0" or "fail to reject H_0" based on the p value.
  H_0: p > 0.05, observed counts roughly match predicted counts.
  H_A: p < 0.05, observed counts don't match predicted counts.
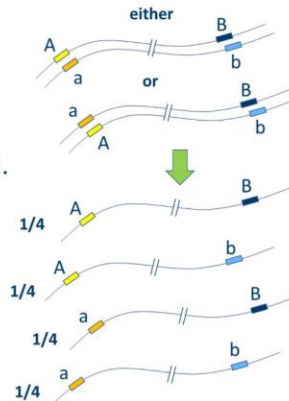
Chi-Squared test conclusions

What does it mean to reject H_0?

Goodness of fit test:
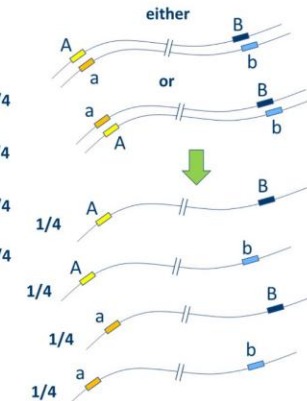 - 1+ assumptions of the model are violated.

For Mendelian segregation, the main assumption is that alleles from different loci segregate (i.e., go into gametes) independently because they are on different chromosomes or far apart on the same chromosome (i.e., they're not "linked").

Dihybrid cross model: AaBb x AaBb

	AB	Ab	aB	ab	
AB	AABB	AABb	AaBB	AaBb	1/4
Ab	AABb	AAbb	AaBb	Aabb	1/4
aB	AaBB	AaBb	aaBB	aaBb	1/4
ab	AaBb	Aabb	aaBb	aabb	1/4
	1/4	1/4	1/4	1/4	

If A dominant to a, and B dominant to b, then offspring will have phenotype proportions of:
  9/16 , 3/16 , 3/16 , 1/16

Example #1 - flower colors $X^2 = 2.2146$

What are the degrees of freedom?
df = 4 - 1 = 3 for this goodness of fit model (the 1/4 proportions come from outside the data itself).

From the table: $\chi^2_{3,\alpha=0.05} = 7.815$

Since 2.2146 < 7.815, p > 0.05.

"The observed frequencies of offspring phenotypes from the dihybrid cross *do not significantly deviate* from the predictions of a model assuming free recombination between these loci (p>0.05)."

Example #1 - flower colors

round blue flowers: 871
round yellow flowers: 295
oval blue flowers: 301
oval yellow flowers: 112
                     1,579

Expected for each phenotype.

Obs	round	oval
blue	871	322
yellow	295	112

Exp	round	oval
blue	888.188	296.063
yellow	296.063	98.688

Exp	round	oval
blue	0.33260	0.08234
yellow	0.00381	1.79580

X^2 contributions

$X^2 = \sum_{i=1}^{4} \left[\frac{(O_i - E_i)^2}{E_i} \right]$

X^2 = 0.33260+0.08234+0.00381+1.79580 = 2.2146

Example #1 - flower colors

Consider two diallelic flower loci. A locus for shape round (RR, Ro) or oval (oo). A locus for color blue (BB, By) or yellow (yy).

Two heterozygous round blue individuals are crossed.

$X^2 = \sum_{i=1}^{4} \left[\frac{(O_i - E_i)^2}{E_i} \right]$

Number of offspring with each phenotype:

round blue flowers: 871   ⎤ These are the observed
round yellow flowers: 295   ⎥ counts. We need to compare
oval blue flowers: 301   ⎥ them to the predicted counts
oval yellow flowers: 112   ⎦ and calculate the X^2

Example #2 - *Drosophila* traits

Consider two diallelic loci. A locus for eye color, red (RR, Rs) or sepia (ss). A locus for wings, wild-type (WW,Wc) or curled (cc).

Two heterozygous red eyed wild-type individuals are crossed.

$X^2 = \sum_{i=1}^{4} \left[\frac{(O_i - E_i)^2}{E_i} \right]$

Number of offspring with each phenotype:

red eyes wild-type: 42   ⎤ These are the observed
red eyes curled wings: 9   ⎥ counts. We need to compare
sepia eyes wild-type: 11   ⎥ them to the predicted counts
sepia eyes curled wings: 10   ⎦ and calculate the X^2

Example #2 - *Drosophila* traits

red eyes wild-type :   42
red eyes curled wing: 9
sepia eyes wild-type :   11
sepia eyes curled wing : 10
                         72

Expected for each phenotype.

Obs	red	sepia
wild-type	42	11
curled	9	10

Exp	red	sepia
wild-type	40.5	13.5
curled	13.5	4.5

Exp	red	sepia
wild-type	0.05556	0.46296
curled	1.5000	6.72222

X^2 contributions

$X^2 = \sum_{i=1}^{4} \left[\frac{(O_i - E_i)^2}{E_i} \right]$

X^2 = 0.05556+0.46296+1.5000+6.72222 = 8.7407

Example #2 - *Drosophila* traits $X^2 = 8.7407$

What are the degrees of freedom?
df = 4 - 1 = 3 for this goodness of fit model (the 1/4 proportions come from outside the data itself).

From table: $\chi^2_{3,\alpha=0.05} = 7.815$, $\chi^2_{3,\alpha=0.025} = 9.348$

Since 7.815 < 8.7407 < 9.348, 0.025 < p < 0.05.

"The observed frequencies of offspring phenotypes from the dihybrid cross *significantly deviate* from the predictions of a model assuming free recombination between these loci (0.025<p<0.05)."