

Evaluating Genetic Data: Chi-Square Analysis

- Mendel's 3:1 monohybrid and 9:3:3:1 dihybrid ratios are hypothetical predictions based on the following assumptions

1. Dominance/Recessiveness
2. Segregation
3. Independent assortment and
4. Random fertilization

* The last three are influenced by chance events and therefore subject to normal deviation. E.g. is the tossing of a coin. This concept is called **chance deviation**.

- The implication of chance deviation is that, as the total number of tosses is reduced, the impact of chance deviation increases.
- With the toss of 4 coins, it is possible to get all heads or tails
- But 1000 tosses **cannot** give all heads or tails. Usually, about 500 heads and 500 tails would be observed.

- Any fluctuation from this hypothetical ratio (e.g. 486 heads and 514 tails) would be attributed to chance deviation.
- Two major points of significance are:
 - (i) the outcomes of segregation, independent assortment, and fertilization, like coin tossing, are subject to random fluctuations from their predicted occurrences due to chance deviation.

- (ii) the smaller the sample size, the greater is the impact of chance deviation on the final outcome.

How do we evaluate these deviations?

- By statistical analysis: to see how well an **observed** data fit the **expected** ratios.
- Comparing an experiment's observed results with its expected result is said to test the **goodness of fit**.
- “The goodness of fit” helps the geneticist to know how much the experimental result can differ from the hypothetical or calculated figure and still be regarded as statistically close to expectations

- If the data do not “fit” precisely, how much deviation do we allow before we question the basic assumption underlying the hypothetical ratio.
- To answer this question, Chi-square analysis (χ^2) is used to assess observed deviation.
- The test takes into account the observed deviation in each component of an expected ratio as well as the sample size and reduces them to a single numerical value.

- The value (χ^2) is then used to estimate how frequently the observed deviation can be expected to occur strictly due to chance.
- The formula used in Chi-square analysis is given as:

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

- Where “o” is the observed value for a given category and “e” is the expected value for that category.
- Since $(o - e)$ is the deviation in each case, then the equation can be reduced to
- $\chi^2 = \sum d^2/e$

Chi-Square Analysis

a) A hypothetical Monohybrid Cross

Expected ratio (o)	Observed	Expected (e)	Deviation (d=o-e)	Squared Deviation (d^2)	Squared Deviation/Expected (d^2/e)
$\frac{3}{4}$	740	$\frac{3}{4} \times 1000 = 750$	$740 - 750 = -10$	$(-10)^2 = 100$	$100/750 = 0.13$
$\frac{1}{4}$	260	$\frac{1}{4} \times 1000 = 250$	$260 - 250 = 10$	$10^2 = 100$	$100/250 = 0.40$
Total	1000			$\sum X^2$	0.53
				P	0.48

Chi-Square Analysis

b) A hypothetical Dihybrid Cross

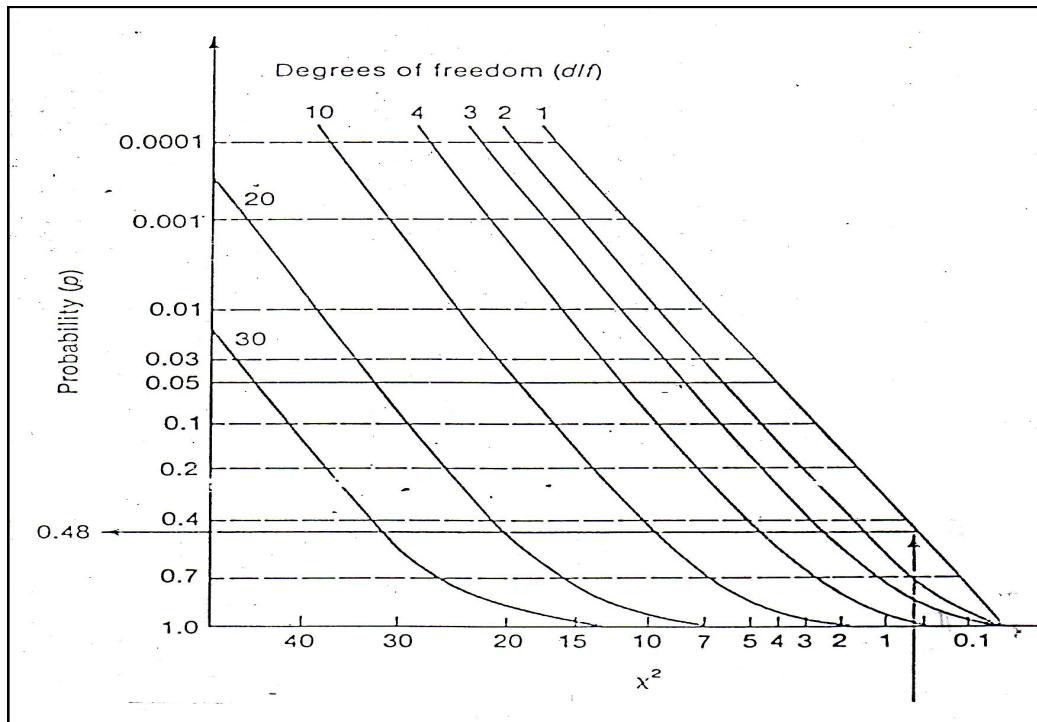
Expected ratio	o	e	d=o-e	d ²	(d ² /e)
9/16	583	563	20	400	0.71
3/16	195	187	8	64	0.34
3/16	166	187	-21	441	2.36
1/16	56	63	- 7	49	0.77
Total	1000			$\Sigma \chi^2$	4.18
				<i>P</i>	0.26

- The final step is the interpretation of the χ^2 value
- First, we determine the value of the **degrees of freedom (d/f)**, which is equal to $n-1$, where n is the number of different categories into which each datum point may fall.
- For the 3:1 ratio, $n = 2$, so $d/f = 2 - 1 = 1$
- The d/f for the 9:3:3:1 ratio is 3

- D/f must always be taken into account because the greater the number of categories, the more deviation is expected due to chance.
- The next step is to convert the χ^2 value to the corresponding probability value (P), using a prepared chart or graph.

Table of Chi-Square (χ^2)^a

DEGREES OF FREEDOM	$P = 0.99$	$P = 0.95$	$P = 0.80$	$P = 0.50$	$P = 0.20$	$P = 0.05$	$P = 0.01$
1	0.000157	0.00393	0.0642	0.455	1.642	3.841	6.635
2	0.020	0.103	0.446	1.386	3.219	5.991	9.210
3	0.115	0.352	1.005	2.366	4.642	7.815	11.345
4	0.297	0.711	1.649	3.357	5.989	9.488	13.277
5	0.554	1.145	2.343	4.351	7.289	11.070	15.086
6	0.872	1.635	3.070	5.348	8.558	12.592	16.812
7	1.239	2.167	3.822	6.346	9.803	14.067	18.475
8	1.646	2.733	4.594	7.344	11.030	15.507	20.090
9	2.088	3.325	5.380	8.343	12.242	16.919	21.666
10	2.558	3.940	6.179	9.342	13.442	18.307	23.209
15	5.229	7.261	10.307	14.339	19.311	24.996	30.578
20	8.260	10.851	14.578	19.337	25.038	31.410	37.566
25	11.524	14.611	18.940	24.337	30.675	37.652	44.314
30	14.953	18.493	23.364	29.336	36.250	43.773	50.892



- Using the dihybrid cross above, where $P = 0.26$ as an example:
- *The first interpretation is that, the probability is 26% or about 1 in 4, that the deviation was due to chance.*
- *The second interpretation is that, were the same experiment repeated many times, 26% of the trials would be expected to exhibit chance deviation.*

Interpretation

- Is 0.26 an **acceptable** or **unacceptable** P -value?
- The decision is relative and depends on the certainty of the investigator
- By convention, 0.05 has been chosen as an arbitrary standard.

- All P values between 0.05 and 1.0 are considered **acceptable** in chi-square analysis.
- All values below 0.05 are **unacceptable** with respect to goodness of fit.
- 0.26 is much above 0.05 and therefore acceptable.

- In other words, our data are consistent with the hypothesis of a 9:3:3:1 ratio of phenotypes, which is indicative of a two-locus genetic model with dominance at each locus.

- Were the P value below this standard, we would have rejected the hypothesis for the experiment. The data would then be interpreted as unacceptable in fitting a 9:3:3:1 ratio.

Homework

- A heterozygous genetic condition called “creeper” in chickens produces shortened and deformed legs and wings, giving the bird a squat appearance. Matings between creepers produced
775 creeper : 388 normal progeny.
 - (a) Is the hypothesis of a 3:1 ratio acceptable?
 - (b) Does a 2:1 ratio fit the data better?