TOPIC 10.2: CHI-SQUARED TEST

Worked Example

A chi-squared test can be applied to phenotypic ratios to determine if there is a statistically significant likelihood that two particular genes are linked or unlinked

Case Study: Pea Plant Inheritance Patterns

Peas can be considered smooth (R) or wrinkled (r) **and** yellow (Y) or green (y)

Two heterozygous plants (RrYy) were crossed to give the following results:

- Smooth / yellow = 701 plants
- Smooth / green = 204 plants
- Wrinkled / yellow = 243 plants
- Wrinkled / green = 68 plants

Pea Phenotype	Observed
= smooth/yellow	701
= smooth/green	204
= wrinkled/yellow	243
= wrinkled/green	68
Total	**1216**

Step 1: Identify Expected Frequencies

There are two possibilities regarding the phenotypic ratios of the two genes:

- *Null Hypothesis (H_0)* – There is **no** association (i.e. genes are unlinked)
- *Alternative Hypothesis (H_1)* – There **is** an association (i.e. genes are linked)

A table is constructed to identify *expected* frequencies of distribution *(unlinked)*

- This data will be compared against the *observed* values previously identified

The expected ratios are calculated using a dihybrid cross (ratios = 9 : 3 : 3 : 1)

- The ratios are applied to total population to determine expected frequencies

Pea	Ratio	Expected
	9/16 × 1216	684
	3/16 × 1216	228
	3/16 × 1216	228
	1/16 × 1216	76
Total	9 : 3 : 3 : 1	1216

Step 2: Apply the Chi-Squared Formula

The chi-squared (χ^2) formula calculates a value based on a comparison of the observed frequencies (O) and the expected frequencies (E)

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

Based on the worked example, the value calculated by the chi-squared test is:

- $\chi^2 = 0.42 + 2.53 + 0.99 + 0.84 =$ **4.76**

A degree of freedom *(df)* will also be required to determine statistical significance

- **df** = (number of rows – 1) × (number of columns – 1)

Raw data table had 4 rows and 2 columns, so degree of freedom equals **three**

Pea	O	E	$\frac{(O-E)^2}{E}$
	701	684	0.42
	204	228	2.53
	243	228	0.99
	68	76	0.84

Step 3: Determine Significance

The chi-squared value is used to determine statistical significance (p value)

- $p<0.05$ is considered significant (less than 5% likelihood results due to chance)

Based on the worked example, a value of 4.76 lies below a p value of 0.05

- This means results are **not** significant (>5% probability it is due to chance)

The alternative hypothesis can be rejected and the null hypothesis accepted

- It is statistically unlikely that the genes are linked (they are likely unlinked)

		df
		3
p value	0.01	11.345
	0.05	7.815
	0.25	4.110

Values that are greater than this are statistically significant