

Probability and Chi-Square Analysis

Contents

- [1 Mendel's Observations](#)
- [2 Probability: Past Punnett Squares](#)
- [3 The Chi-Square Test](#)
 - [3.1 Chi-Square Test: Is this coin fair or weighted? \(activity\)](#)
 - [3.2 Chi-Square Test of Kernel Coloration and Texture in an F₂ Population \(activity\)](#)

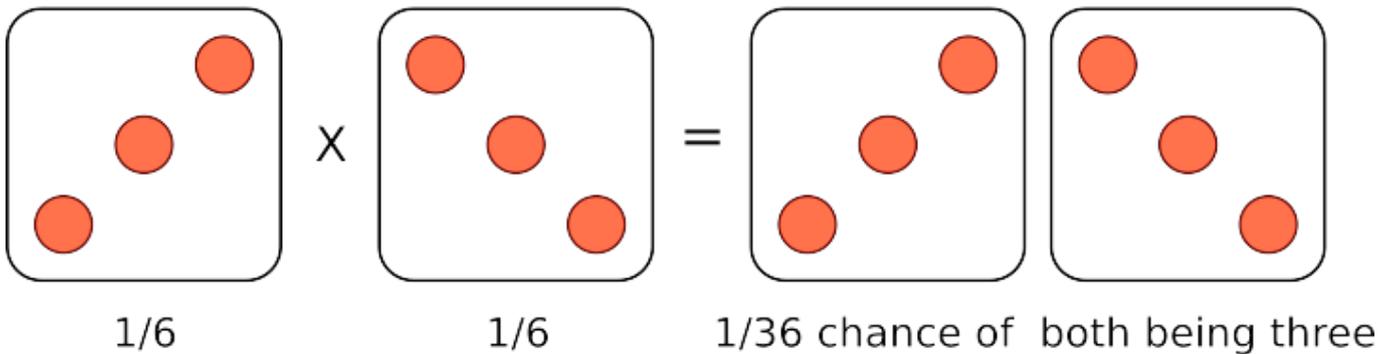
Mendel's Observations

Parental Cross	F ₁ Phenotypes	F ₂ Phenotypes	F ₂ Ratios
Round x Wrinkled Seeds	All Round	5474 Round: 1850 Wrinkled	2.96:1
Yellow x Green Seeds	All Yellow	6022 Yellow: 2001 Green	3.01:1
Purple x White Flowers	All Purple	705 Purple: 224 White	3.15:1
Tall x Dwarf Plants	All Tall	787 Tall: 227 Dwarf	2.84:1

Probability: Past Punnett Squares

Punnett Squares are convenient for predicting the outcome of monohybrid or dihybrid crosses. The expectation of two heterozygous parents is 3:1 in a single trait cross or 9:3:3:1 in a two-trait cross. Performing a three or four trait cross becomes very messy. In these instances, it is better to follow the rules of probability. **Probability** is the chance that an event will occur expressed as a fraction or percentage. In the case of a monohybrid cross, 3:1 ratio means that there is a 3/4 (0.75) chance of the dominant phenotype with a 1/4 (0.25) chance of a recessive phenotype.

$$1/6 \times 1/6 = 1/36$$



A single die has a 1 in 6 chance of being a specific value. In this case, there is a $1/6$ probability of rolling a 3. It is understood that rolling a second die simultaneously is not influenced by the first and is therefore independent. This second die also has a $1/6$ chance of being a 3.

We can understand these rules of probability by applying them to the dihybrid cross and realizing we come to the same outcome as the 2 monohybrid Punnett Squares as with the single dihybrid Punnett Square.

Round = R
Wrinkled = r

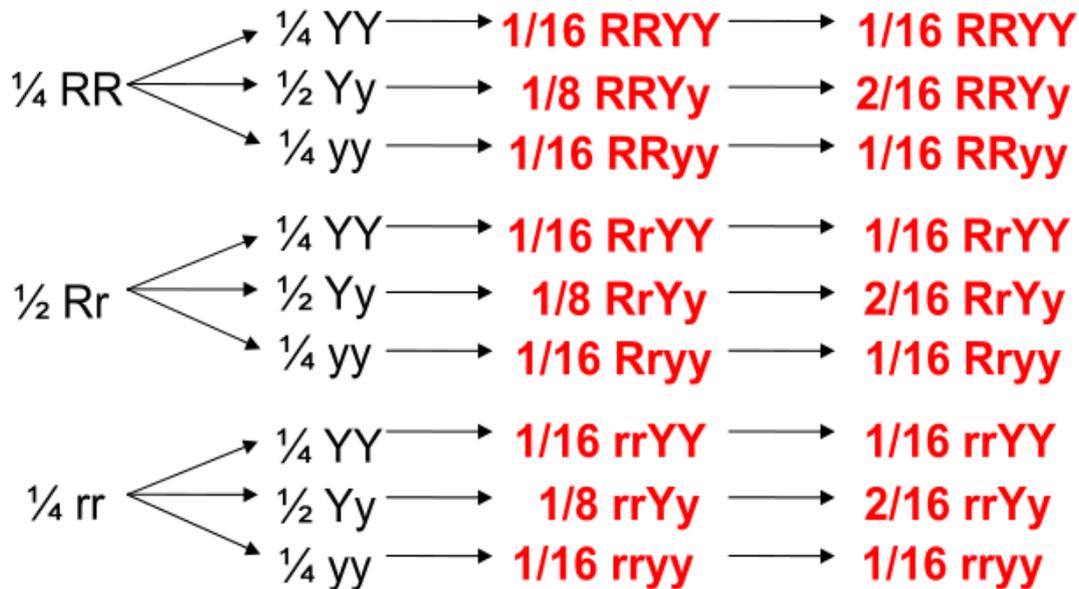
	R	r
R	RR	Rr
r	Rr	rr

$\frac{1}{4}$ RR $\frac{1}{2}$ Rr $\frac{1}{4}$ rr

Yellow = Y
Green = y

	Y	y
Y	YY	Yy
y	Yy	yy

$\frac{1}{4}$ YY $\frac{1}{2}$ Yy $\frac{1}{4}$ yy



This forked line method of calculating probability of offspring with various genotypes and phenotypes can be scaled and applied to more characteristics.

The Chi-Square Test

The χ^2 statistic is used in genetics to illustrate if there are deviations from the expected outcomes of the alleles in a population. The general assumption of any statistical test is that there are no significant deviations between the measured results and the predicted ones. This lack of deviation is called the **null hypothesis** (H_0). χ^2 statistic uses a distribution table to compare results against at varying levels of probabilities or **critical values**. If the χ^2 value is

greater than the value at a specific probability, then the null hypothesis has been rejected and a significant deviation from predicted values was observed. Using Mendel's laws, we can count phenotypes after a cross to compare against those predicted by probabilities (or a Punnett Square).

p-Value for χ^2 test				
DF	0.1	0.05	0.01	0.005
1	2.706	3.84	6.64	7.88
2	4.605	5.99	9.21	10.60
3	6.251	7.82	11.35	12.94
4	7.779	9.49	13.28	14.86
5	9.236	11.07	15.09	16.75
6	10.65	12.59	16.81	18.55
7	12.02	14.07	18.48	20.28
8	13.36	15.51	20.09	21.96
9	14.68	16.92	21.67	23.59
10	15.99	18.31	23.21	25.19

In order to use the table, one must determine the stringency of the test. The lower the p-value, the more stringent the statistics. **Degrees of Freedom (DF)** are also calculated to determine which value on the table to use. Degrees of Freedom are the number of classes or categories there are in the observations minus 1. $DF=n-1$

In the example of corn kernel color and texture, there are 4 classes: Purple & Smooth, Purple & Wrinkled, Yellow & Smooth, Yellow & Wrinkled. Therefore, $DF = 4 - 1 = 3$ and choosing p^2 must be greater than 7.82 in order to be significantly deviating from what is expected. With this dihybrid cross example, we expect a ratio of 9:3:3:1 in phenotypes where 1/16th of the population are recessive for both texture and color while 9/16th of the population display both color and texture as the dominant. 3/16th will be dominant for one phenotype while recessive for the other and the remaining 3/16th will be the opposite combination.

With this in mind, we can predict or have expected outcomes using these ratios. Taking a total

count of 200 events in a population, $9/16(200)=112.5$ and so forth. Formally, the χ^2 value is generated by summing all combinations of:

$$(\text{observed}-\text{expected})^2/\text{expected}$$

Chi-Square Test: Is this coin fair or weighted? (activity)

1. Everyone in the class should flip a coin 2x and record the result (assumes class is 24)
2. Fair coins are expected to land 50% heads and 50% tails
 - 50% of 48 results should be 24
 - 24 heads and 24 tails is already written in the “Expected” column
3. As a class, compile the results in the “Observed” column (total of 48 coin flips)
4. In the last column, subtract the expected heads from the observed heads and square it, then divide by the number of expected heads
5. In the last column, subtract the expected tails from the observed tails and square it, then divide by the number of expected tails
6. Add the values together from the last column to generate the χ^2 value
7. Compare the value with the value at 0.05 with DF=1
 - there are 2 classes or categories (head or tail), so $DF = 2 - 1 = 1$
 - Were the coin flips fair (not significantly deviating from 50:50)?

Heads or Tails	Expected Number	Observed Number	$\frac{(\text{Observed}-\text{Expected})^2}{\text{Expected}}$
Heads	24		
Tails	24		
$\chi^2 = \sum \frac{(\text{Observed}-\text{Expected})^2}{\text{Expected}} \rightarrow$			

Let’s say that the coin tosses yielded 26 Heads and 22 Tails. Can we assume that the coin was unfair? If we toss a coin an odd number of times (eg. 51), then we would expect that the results would yield 25.5 (50%) Heads and 25.5 (50%) Tails. But this isn’t a possibility. This is when the χ^2 test is important as it delineates whether 26:25 or 30:21 etc. are within the probability for a fair coin.

Chi-Square Test of Kernel Coloration and Texture in an F_2 Population (activity)

1. From the counts, one can assume which phenotypes are dominant and recessive
2. Fill in the “Observed” category with the appropriate counts
3. Fill in the “Expected Ratio” with either 9/16, 3/16 or 1/16
4. The total number of counted event was 200, so multiply the “Expected Ratio” x 200 to

generate the “Expected Number” fields

5. Calculate the $(\text{Observed}-\text{Expected})^2/\text{Expected}$ for each phenotype combination
6. Add all $(\text{Observed}-\text{Expected})^2/\text{Expected}$ values together to generate the χ^2 value and compare with the value on the table where DF=3
7. Do we reject the Null Hypothesis or were the observed numbers as we expected as roughly 9:3:3:1?
 - What would it mean if the Null Hypothesis was rejected? Can you explain a case in which we have observed values that are significantly altered from what is expected?

Phenotype	Expected Ratio (9/16, 3/16,1/16)	Expected Number (Total # x Ratio)	Observed Number	$\frac{(\text{Observed}-\text{Expected})^2}{\text{Expected}}$
Purple & Smooth				
Purple & Wrinkly				
Yellow & Smooth				
Yellow & Wrinkly				
$\chi^2 = \sum \frac{(\text{Observed}-\text{Expected})^2}{\text{Expected}} \rightarrow$				

Tags: [quantitative reasoning](#), [integration of knowledge](#), [guided inquiry](#)