In the mid 1800’s, Gregor Mendel demonstrated the existence of genes based on the regular occurrence of certain characteristic ratios of dichotomous characters (or traits) among the offspring of crosses between parents of various characteristics and lineages.

These ratios are known as segregation ratios.

The analysis of segregation ratios remains an important research tool in human genetics.

The demonstration of such ratios for a discrete trait among the offspring of certain types of families constitutes strong evidence that the trait has a simple genetic basis.
Simple Mendelian disorders or traits can be adequately modeled using Mendel's laws. Generally, these traits are close to completely penetrant.

Mendel’s Laws

- **Law of Segregation (The "First Law")**: The alleles at a gene segregate (separate from each other) into different gametes during meiosis. An individual receives with equal probability one of the two alleles at a gene from the mother and one of two alleles at a gene from the father.

- **Law of Independent Assortment (The "Second Law")**: The segregation of the genes for one trait is independent of the segregation of genes for another trait, i.e., when genes segregate, they do so independently.

Generally, Mendelian traits are close to completely penetrant and are a function of a small number of factors.
Mode of Inheritance is the manner in which a particular genetic trait or disorder is passed from one generation to the next.

Classical Mendelian experiments use inbred strains of animals or self-fertilized plants so that individuals in each of the starting generation parental groups are homozygous at every locus and are genetically identical.
Example 1: Rabbits

grey × albino

↓

grey F1

↓

grey : black : albino F2

9 : 3 : 4

Proposed genetic model to explain the observed segregation ratios: color is controlled by two genes. Gene 1 controls the presence of color: alleles C and c. Gene 2 controls whether color is grey or black: alleles G and g

IBD Sharing Probabilites for Outbreds

<table>
<thead>
<tr>
<th>Gene 1</th>
<th>Gene 2</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC or Cc</td>
<td>GG or Gg</td>
<td>grey rabbit</td>
</tr>
<tr>
<td></td>
<td>gg</td>
<td>black rabbit</td>
</tr>
<tr>
<td>cc</td>
<td>any genotype</td>
<td>albino rabbit</td>
</tr>
</tbody>
</table>

Parental generation groups → homozygous for both genes

Does the proposed genetic model explain the observed segregation ratios of the phenotypes?
Example 1: Rabbits

CCGG × ccbb

↓

CcGg

↓

F1

<table>
<thead>
<tr>
<th>F2</th>
<th>CG</th>
<th>Cg</th>
<th>cG</th>
<th>cg</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG</td>
<td>grey</td>
<td>grey</td>
<td>grey</td>
<td>grey</td>
</tr>
<tr>
<td>Cg</td>
<td>grey</td>
<td>black</td>
<td>grey</td>
<td>black</td>
</tr>
<tr>
<td>cG</td>
<td>grey</td>
<td>grey</td>
<td>albino</td>
<td>albino</td>
</tr>
<tr>
<td>cg</td>
<td>grey</td>
<td>black</td>
<td>albino</td>
<td>albino</td>
</tr>
</tbody>
</table>

So expected phenotype relative frequencies for grey: black: albino are 9 : 3 : 4
Proposed genetic model: color is controlled by three genes. Gene 1 controls the presence of color: alleles C and c. Gene 2, with alleles G and g, and Gene 3, with alleles B and b, interact to produce the color of mice that have alleles to produce color.

Parental generation $\rightarrow$ homozygous CC at Gene 1

<table>
<thead>
<tr>
<th>Gene 2</th>
<th>Gene 3</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>GG or Gg</td>
<td>Any genotype</td>
<td>grey mouse</td>
</tr>
<tr>
<td>gg</td>
<td>BB or Bb</td>
<td>black mouse</td>
</tr>
<tr>
<td></td>
<td>bb</td>
<td>chocolate mouse</td>
</tr>
</tbody>
</table>

If this model where the correct model for color, what segregation ratios of the phenotypes in the F2 generation would we expect for grey : black : chocolate?
Example 2: Mice

\[
\begin{align*}
\text{GGBB} & \times \text{ggbb} \\
\downarrow & \\
\text{GgBb} & \quad \text{F1} \\
\downarrow & \\
\text{GB} & \quad \text{grey} \\
\text{Gb} & \quad \text{grey} \\
\text{gB} & \quad \text{black} \\
\text{gb} & \quad \text{chocolate}
\end{align*}
\]

The expected phenotype relative frequencies for grey: black: albino are 12 : 3 : 1
Example 3: Bean Flower Color

Flowers come in shades from white to purple. Quantify color: white (0) to purple (10)

\[ 10 \times 0 \]

(purple \times white)

↓

5

↓

<table>
<thead>
<tr>
<th>color</th>
<th>10</th>
<th>9</th>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>relative counts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Proposed genetic model: color is controlled by two genes with additive effects.
  - Gene 1: A=3, a=0
  - Gene 2: B=2, b=0
- If this model where the correct model for color, what relative counts would we expect?
### Example 3: Bean Flower Color

<table>
<thead>
<tr>
<th></th>
<th>AB</th>
<th>Ab</th>
<th>aB</th>
<th>ab</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td>10</td>
<td>8</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Ab</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>aB</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>ab</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>color</th>
<th>10</th>
<th>9</th>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>relative counts</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Aggregation and segregation studies are generally the first step when studying the genetics of a human trait.

Aggregation studies evaluate the evidence for whether there is a genetic component to a study.

They do this by examining whether there is familial aggregation of the trait.

Questions of interest include

- Are relatives of diseased individuals more likely to be diseased than the general population?
- Is the clustering of disease in families different from what you’d expect based on the prevalence in the general population?
Aggregation Study Example: Alzheimer’s Disease - Studies based on twins have found differences in concordance rates between monozygotic and dizygotic twins. In particular, 80% of monozygotic twin pairs were concordant whereas only 35% of dizygotic twins were concordant. In a separate study, first-degree relatives of individuals (parents, offspring, siblings) with Alzheimer’s disease were studied. First degree relatives of patients had a 3.5 fold increase in risk for developing Alzheimer’s disease as compared to the general population. This was age-dependent with the risk decreasing with age-of-onset.

Segregation analysis moves beyond aggregation of disease and seeks to more precisely identify the factors responsible for familial aggregation. For instance,

- Is the aggregation due to environmental, cultural or genetic factors?
- What proportion of the trait is due to genetic factors?
- What mode of inheritance best represents the genetic factors?
- Does there appear to be genetic heterogeneity?
Exercise: Characterize the pattern of inheritance one would expect to see in pedigrees for autosomal dominant and recessive genes. Do the same for X-linked inheritance. Assume full penetrance.

- Dominant autosomal
- Recessive autosomal
- Dominant X-linked
- Recessive X-linked
Consider a disease that is believed to be caused by a fully penetrant rare mutant allele at an autosomal locus.

Let D be the allele causing the disorder and let d represent the normal allele.

There are 9 possible mating types (can collapse to six mating types due to symmetry)

Each of these mating types will produce offspring with a characteristic distribution of genotypes and therefore a distribution of phenotypes.

The proportions of the different genotypes and phenotypes in the offspring of the six mating types are known as the segregation ratios of the mating types.
These specific values of the segregation ratios can be used to test whether a disease is caused by a single autosomal dominant gene.

Suppose that a random sample of matings between two parents where one is affected and one is unaffected is obtained. Out of a total of \( n \) offspring, \( r \) are affected.

Since autosomal dominant genes are usually rare, it is reasonable to assume that the frequency of allele \( D \) is quite low and that most affected individuals are expected to have genotype of \( Dd \) instead of \( DD \).

What are the matings in the sample under this assumption?

How can we test if the observed segregation ratios in the offspring are what is expected if the disease were indeed caused by an autosomal dominant allele?

The Binomial distribution can be used to model this data.
The binomial distribution is a very common discrete probability distribution that arises in the following situation:

- A fixed number, $n$, of trials
- The $n$ trials are independent of each other
- Each trial has exactly two outcomes: “success” and “failure”
- The probability of a success, $p$, is the same for each trial

If $X$ is the total number of successes in a binomial setting, then we say that the probability distribution of $X$ is a **binomial distribution** with parameters $n$ and $p$: $X \sim B(n, p)$

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

Segregation and Aggregation Analysis
Segregation analysis for autosomal dominant disease

1. Let \( X \) be the number of offspring that are affected.
2. Under the null hypothesis, \( X \) will have a binomial distribution

\[
P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}
\]

where \( p \) is the probability that an offspring is affected.
3. We are interested in testing

\[H_0: p = \frac{1}{2} \quad \text{vs.} \quad H_a: p \neq \frac{1}{2}\]

4. Out of a total of \( n \) offspring, \( r \) are affected. The p-value is the probability of observing a value at least as extreme as \( r \). If \( r < \frac{n}{2} \), the p-value is

\[
\sum_{x=0}^{r} \binom{n}{x} \left( \frac{1}{2} \right)^x \left( \frac{1}{2} \right)^{n-x} + \sum_{x=n-r}^{n} \binom{n}{x} \left( \frac{1}{2} \right)^x \left( \frac{1}{2} \right)^{n-x}
\]

\[
= \left( \frac{1}{2} \right)^{n-1} \sum_{x=0}^{r} \binom{n}{x}
\]


Autosomal dominant disease example

- Marfan syndrome, a connective tissue disorder, is a rare disease that is believed to be autosomal dominant (and actually is!).
- 112 offspring of an affected parent and an unaffected parent are sample.
- 52 of the offspring are affected and 60 are unaffected.
- Are these observations consistent with an autosomal dominant disease.
- The p-value is

\[
\left( \frac{1}{2} \right)^{112-1} \sum_{x=0}^{52} \binom{112}{x} = 0.5085
\]

- What if only 42 of the offspring are affected?

\[
\left( \frac{1}{2} \right)^{112-1} \sum_{x=0}^{42} \binom{112}{x} = 0.0104
\]
Normal Approximation to Binomial

- If $X \sim B(n, p)$, and $n$ is large enough such that
  
  
  \[ np \geq 10 \quad \text{and} \quad n(1 - p) \geq 10 \]

- Then $X$ is approximately $N \left( \mu_X = np, \sigma_X = \sqrt{np(1-p)} \right)$

- For the Marfan syndrome data with 52 offspring affected,
  
  \[ z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{52.5 - (112)(.5)}{\sqrt{112(.5)(.5)}} = -0.661 \]

  P-value is $2P(Z \geq |z|) = 2(0.2539) = .5079$, where $Z$ follows a standard normal distribution

- For the Marfan syndrome data with 42 offspring affected, the p-value is .0107.
How can you do segregation analysis to test if a disease that is fully penetrant autosomal recessive?

For this model we know that affected individuals are DD, but unaffected individuals could be Dd or dd.

One proposal is to look at the segregation ratios in families with unaffected parents and at least one affected offspring. What are some problems with this proposal?