Population Genetics

Hardy-Weinberg Model

Jessie Lenarz

Concordia College

November 24, 2009

Outline

Outline

Terminology

- gene – the physical entity transmitted from parent to offspring during the reproductive process that influences hereditary traits
- allele – alternative forms of genes
- genotype – set of genes present in an individual
- phenotype – the physical or biochemical expression of the genotype
- homozygous – two alleles at a locus indistinguishable in their effects on the organism
- heterozygous – two alleles at a locus are distinguishable because of their differing effects on the organism

Terminology

- gene – the physical entity transmitted from parent to offspring during the reproductive process that influences hereditary traits
- allele – alternative forms of genes
- genotype – set of genes present in an individual
- phenotype – the physical or biochemical expression of the genotype
- homozygous – two alleles at a locus indistinguishable in their effects on the organism
- heterozygous – two alleles at a locus are distinguishable because of their differing effects on the organism

Terminology

- gene – the physical entity transmitted from parent to offspring during the reproductive process that influences hereditary traits
- allele – alternative forms of genes
- genotype – set of genes present in an individual
- phenotype – the physical or biochemical expression of the genotype
- homozygous – two alleles at a locus indistinguishable in their effects on the organism
- heterozygous – two alleles at a locus are distinguishable because of their differing effects on the organism

Terminology

- gene – the physical entity transmitted from parent to offspring during the reproductive process that influences hereditary traits
- allele – alternative forms of genes
- genotype – set of genes present in an individual
- phenotype – the physical or biochemical expression of the genotype
- homozygous – two alleles at a locus indistinguishable in their effects on the organism
- heterozygous – two alleles at a locus are distinguishable because of their differing effects on the organism

Terminology

- gene – the physical entity transmitted from parent to offspring during the reproductive process that influences hereditary traits
- allele – alternative forms of genes
- genotype – set of genes present in an individual
- phenotype – the physical or biochemical expression of the genotype
- homozygous – two alleles at a locus indistinguishable in their effects on the organism
- heterozygous – two alleles at a locus are distinguishable because of their differing effects on the organism

Terminology

- gene – the physical entity transmitted from parent to offspring during the reproductive process that influences hereditary traits
- allele – alternative forms of genes
- genotype – set of genes present in an individual
- phenotype – the physical or biochemical expression of the genotype
- homozygous – two alleles at a locus indistinguishable in their effects on the organism
- heterozygous – two alleles at a locus are distinguishable because of their differing effects on the organism

Outline

Mendelian Genetics

- In most higher organisms, each individual cell contains two copies of each type of chromosome, one inherited from its mother and one inherited from its father.
- At each gene locus, each individual contains two alleles.
- Principle of Segregation
  - Each reproductive cell from a heterozygous individual contains only one of the two alleles, and that overall the reproductive cells contain the two alleles at equal frequency.

Mendelian Genetics

- In most higher organisms, each individual cell contains two copies of each type of chromosome, one inherited from its mother and one inherited from its father.
- At each gene locus, each individual contains two alleles.
- Principle of Segregation
  - Each reproductive cell from a heterozygous individual contains only one of the two alleles, and that overall the reproductive cells contain the two alleles at equal frequency.

Mendelian Genetics

- In most higher organisms, each individual cell contains two copies of each type of chromosome, one inherited from its mother and one inherited from its father.
- At each gene locus, each individual contains two alleles.
- Principle of Segregation
  - Each reproductive cell from a heterozygous individual contains only one of the two alleles, and that overall the reproductive cells contain the two alleles at equal frequency.

Outline

Punnett Squares

Suppose we have a gene with two possible alleles – A (dominant) and a (recessive). If we have two heterozygous parents, then

Father

Mother		A	a
	A	AA	Aa
	a	Aa	aa

genotype frequencies: 25% AA, 50% Aa, 25% aa

phenotype frequencies: 75% A, 25% a

Punnett Squares

ABO Blood Type

The ABO blood type system has three possible alleles: *A*, *B*, and *O*. The *A* allele codes for the “A” protein on red blood cells, the *B* allele codes for the “B” protein on red blood cells, and the *O* allele codes for neither protein on red blood cells. The *A* and *B* alleles are codominant and the *O* allele is recessive.

Phenotype	Associated Genotypes
A	*AA* *AO*
B	*BB* *BO*
AB	*AB*
O	*OO*

Punnett Squares
ABO Blood Type

Blood Type	Frequency
A	40%
B	11%
AB	4%
O	45%

Punnett Squares

ABO Blood Type

My mother is type O and my father is type A. My siblings are both type O, thus my father's genotype is AO and my mother's genotype is OO. The corresponding Punnett Square is

		Father: A	Father: O
Mother	O	AO	OO
	O	AO	OO

genotype frequencies: 50% AO, 50% OO

phenotype frequencies: 50% A, 50% O

Outline

Assumptions of Hardy-Weinberg Model

- diploid organism
- sexual reproduction
- non-overlapping generations
- random mating
- large population
- equal allele frequency in the sexes
- no migration
- no mutation
- no natural selection

Assumptions of Hardy-Weinberg Model

- diploid organism
- sexual reproduction
- non-overlapping generations
- random mating
- large population
- equal allele frequency in the sexes
- no migration
- no mutation
- no natural selection

Assumptions of Hardy-Weinberg Model

- diploid organism
- sexual reproduction
- non-overlapping generations
- random mating
- large population
- equal allele frequency in the sexes
- no migration
- no mutation
- no natural selection

Assumptions of Hardy-Weinberg Model

- diploid organism
- sexual reproduction
- non-overlapping generations
- random mating
- large population
- equal allele frequency in the sexes
- no migration
- no mutation
- no natural selection

Assumptions of Hardy-Weinberg Model

- diploid organism
- sexual reproduction
- non-overlapping generations
- random mating
- large population
- equal allele frequency in the sexes
- no migration
- no mutation
- no natural selection

Assumptions of Hardy-Weinberg Model

- diploid organism
- sexual reproduction
- non-overlapping generations
- random mating
- large population
- equal allele frequency in the sexes
- no migration
- no mutation
- no natural selection

Assumptions of Hardy-Weinberg Model

- diploid organism
- sexual reproduction
- non-overlapping generations
- random mating
- large population
- equal allele frequency in the sexes
- no migration
- no mutation
- no natural selection

Assumptions of Hardy-Weinberg Model

- diploid organism
- sexual reproduction
- non-overlapping generations
- random mating
- large population
- equal allele frequency in the sexes
- no migration
- no mutation
- no natural selection

Assumptions of Hardy-Weinberg Model

- diploid organism
- sexual reproduction
- non-overlapping generations
- random mating
- large population
- equal allele frequency in the sexes
- no migration
- no mutation
- no natural selection

Outline

For 2 alleles

Let p be the frequency of allele A and q be the frequency of allele a. Also let D be the frequency of the genotype AA, H the frequency of genotype Aa, and R the frequency of genotype aa. Note that

$$\begin{aligned} p + q &= 1 \\ p &= \frac{2D + H}{2} \\ q &= \frac{2R + H}{2} \end{aligned}$$

For 2 alleles

Mating	Frequency of Mating	Offspring genotype frequencies AA	Aa	aa
$AA \times AA$	D^2	1	0	0
$AA \times Aa$	$2DH$	$1/2$	$1/2$	0
$AA \times aa$	$2DR$	0	1	0
$Aa \times Aa$	H^2	$1/4$	$1/2$	$1/4$
$Aa \times aa$	$2HR$	0	$1/2$	$1/2$
$aa \times aa$	R^2	0	0	1
	Totals (next generation)	D'	H'	R'

$$D' = D^2 + 2DH/2 + H^2/4 = (D + H/2)^2 = p^2$$

For 2 alleles

Mating	Frequency of Mating	Offspring genotype frequencies AA	Aa	aa
$AA \times AA$	D^2	1	0	0
$AA \times Aa$	$2DH$	$1/2$	$1/2$	0
$AA \times aa$	$2DR$	0	1	0
$Aa \times Aa$	H^2	$1/4$	$1/2$	$1/4$
$Aa \times aa$	$2HR$	0	$1/2$	$1/2$
$aa \times aa$	R^2	0	0	1
	Totals (next generation)	D'	H'	R'

$$H' = 2DH/2 + 2DR + H^2/2 + 2HR/2 = 2(D + H/2)(R + H/2) = 2pq$$

For 2 alleles

Mating	Frequency of Mating	Offspring genotype frequencies		
		AA	Aa	aa
$AA \times AA$	D^2	1	0	0
$AA \times Aa$	$2DH$	$1/2$	$1/2$	0
$AA \times aa$	$2DR$	0	1	0
$Aa \times Aa$	H^2	$1/4$	$1/2$	$1/4$
$Aa \times aa$	$2HR$	0	$1/2$	$1/2$
$aa \times aa$	R^2	0	0	1
	Totals (next generation)	D'	H'	R'

$$R' = H^2/4 + 2HR/2 + R^2 = (R + H/2)^2 = q^2$$

For 2 alleles

		Father	
		$A\,(p)$	$a\,(q)$
Mother	$A\,(p)$	$AA\,(p^2)$	$Aa\,(pq)$
	$a\,(q)$	$Aa\,(pq)$	$aa\,(q^2)$

$$(p+q)^2 = p^2 + 2pr + r^2$$

Outline

For 3 alleles

Let p be the frequency of allele A_1, q be the frequency of allele A_2, and r be the frequency of allele A_3.

		Father		
		$A_1\ (p)$	$A_2\ (q)$	$A_3\ (r)$
	$A_1\ (p)$	$A_1A_1\ (p^2)$	$A_1A_2\ (pq)$	$A_1A_3\ (pr)$
Mother	$A_2\ (q)$	$A_1A_2\ (pq)$	$A_2A_2\ (q^2)$	$A_1A_3\ (qr)$
	$A_3\ (r)$	$A_1A_3\ (pr)$	$A_2A_3\ (qr)$	$A_3A_3\ (r^2)$

$$(p+q+r)^2 = p^2 + 2pq + 2pr + q^2 + 2qr + r^2$$

Outline

Hardy-Weinberg Equilibrium

For 2 alleles

For two alleles A and a with frequencies p and q, respectively, we can calculate the allele frequencies in the next generation. Let p' and q' be the frequencies of the alleles A and a, respectively, in the next generation.

$$p' = \frac{2p^2 + 2pq}{2} = p^2 + pq = p(p+q) = p$$

$$q' = \frac{2q^2 + 2pq}{2} = q^2 + pq = q(p+q) = q$$

Hardy-Weinberg Equilibrium

For 2 alleles

For two alleles A and a with frequencies p and q, respectively, we can calculate the allele frequencies in the next generation. Let p' and q' be the frequencies of the alleles A and a, respectively, in the next generation.

$$p' = \frac{2p^2 + 2pq}{2} = p^2 + pq = p(p+q) = p$$

$$q' = \frac{2q^2 + 2pq}{2} = q^2 + pq = q(p+q) = q$$

Hardy-Weinberg Equilibrium

For 3 alleles

For three alleles A_1, A_2, and A_3 with frequencies p, q, and r, respectively, we can calculate the allele frequencies in the next generation. Let p', q', and r' be the frequencies of the alleles A_1, A_2, and A_3, respectively, in the next generation.

$$p' = \frac{2p^2 + 2pq + 2pr}{2} = p^2 + pq + pr = p(p+q+r) = p$$

$$q' = \frac{2pq + 2q^2 + 2qr}{2} = pq + q^2 + qr = q(p+q+r) = q$$

$$r' = \frac{2pr + 2qr + 2r^2}{2} = pr + qr + r^2 = r(p+q+r) = r$$

Hardy-Weinberg Equilibrium

For 3 alleles

For three alleles $A_1,\ A_2$, and A_3 with frequencies $p,\ q$, and r, respectively, we can calculate the allele frequencies in the next generation. Let $p',\ q'$, and r' be the frequencies of the alleles $A_1,\ A_2$, and A_3, respectively, in the next generation.

$$p' = \frac{2p^2 + 2pq + 2pr}{2} = p^2 + pq + pr = p(p+q+r) = p$$

$$q' = \frac{2pq + 2q^2 + 2qr}{2} = pq + q^2 + qr = q(p+q+r) = q$$

$$r' = \frac{2pr + 2qr + 2r^2}{2} = pr + qr + r^2 = r(p+q+r) = r$$

Hardy-Weinberg Equilibrium
For 3 alleles

For three alleles $A_1,\ A_2$, and A_3 with frequencies $p,\ q$, and r, respectively, we can calculate the allele frequencies in the next generation. Let $p',\ q'$, and r' be the frequencies of the alleles $A_1,\ A_2$, and A_3, respectively, in the next generation.

$$p' = \frac{2p^2 + 2pq + 2pr}{2} = p^2 + pq + pr = p(p + q + r) = p$$

$$q' = \frac{2pq + 2q^2 + 2qr}{2} = pq + q^2 + qr = q(p + q + r) = q$$

$$r' = \frac{2pr + 2qr + 2r^2}{2} = pr + qr + r^2 = r(p + q + r) = r$$

Hardy-Weinberg Equilibrium

χ^2 Test

The Hardy-Weinberg Equilibrium can be used to determine if one of the assumptions has been violated. Use the χ^2 test

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

As an example, the *CCR*5 gene encodes a major macrophage coreceptor for HIV-1. Genotypes that are homozygous for the $CCDR5 - \Delta 32$ deletion are strongly resistant to infection by HIV-1. The $\Delta 32$ allele is found in virtually all European populations, but allele frequencies vary.

Hardy-Weinberg Equilibrium

χ^2 Test

In one sample of 294 Parisians studied for $+$ (the nondeletion) and Δ32 (the deletion) alleles, the number of individuals with each genotype were:

$+/+$: 224 people $+/\Delta 32$: 64 people $\Delta 32/\Delta 32$: 6 people

Estimating the allele frequencies from this sample we have

$$< \text{Frequency of + allele} >= \frac{2 \times 224 + 64}{2 \times 294} = 0.871$$

$$< \text{Frequency of } \Delta 32 \text{ allele} >= \frac{2 \times 6 + 64}{2 \times 294} = 0.129$$

Hardy-Weinberg Equilibrium

χ^2 Test

	genotype			Total
	+/+	$+/\Delta32$	$\Delta32/\Delta32$	
observed	224	64	6	294
expected	222.9	66.2	4.9	294

$$\chi^2 = \frac{(224-222.9)^2}{222.9} + \frac{(64-66.2)^2}{66.2} + \frac{(6-4.9)^2}{4.9} = 0.32$$

The degrees of freedom are

$$\begin{aligned} \text{df} &= \text{\# of classes of data} - 1 - \text{\# of parameters estimated from data} \\ &= 3 - 1 - 1 \\ &= 1 \end{aligned}$$

So the probability value is $p = 0.63$. Since this is considered large, we would not reject the HW-model.

Hardy-Weinberg Equilibrium

χ^2 Test

In a different sample of 276 people from Rheims, the number of individuals with each genotype were:

$+/+$: 234 people $+/\Delta 32$: 36 people $\Delta 32/\Delta 32$: 6 people

Estimating the allele frequencies from this sample we have

$$< \text{Frequency of + allele} >= \frac{2 \times 234 + 36}{2 \times 276} = 0.913$$

$$< \text{Frequency of } \Delta 32 \text{ allele} >= \frac{2 \times 6 + 36}{2 \times 276} = 0.087$$

Hardy-Weinberg Equilibrium

χ^2 Test

	genotype			Total
	+/+	$+/\Delta 32$	$\Delta 32/\Delta 32$	
observed	234	36	6	276
expected	230.1	43.8	2.1	276

$$\chi^2 = \frac{(234 - 230.1)^2}{230.1} + \frac{(36 - 43.8)^2}{43.8} + \frac{(6 - 2.1)^2}{2.1} = 8.8$$

The degrees of freedom are

$$\begin{aligned} \text{df} &= \text{\# of classes of data} - 1 - \text{\# of parameters estimated from data} \\ &= 3 - 1 - 1 \\ &= 1 \end{aligned}$$

So the probability value is $p = 0.002$. Since this is considered small, we would reject the HW-model.

References

- Hartl, Daniel L. *A Primer of Population Genetics*, 3rd edition. 2000.

Questions?