P8149 Human Population Genetics - Homework 4

Yongyan Liu (yl6107)

2025-12-16

Question 1

Suppose that a random-breeding population is sampled and the following genotype frequencies of a protein variant are found:

AA	AB	BB
42	76	448

What is the inbreeding coefficient indicated by these numbers?

Solution

Total sample size: $n = 42 + 76 + 448 = 566$

Observed genotype frequencies:

$$P_{AA} = \frac{42}{566}, \quad P_{AB} = \frac{76}{566}, \quad P_{BB} = \frac{448}{566}$$

Allele frequencies:

$$p_A = P_{AA} + \frac{1}{2}P_{AB} = \frac{42}{566} + \frac{1}{2} \cdot \frac{76}{566} = \frac{42 + 38}{566} = \frac{80}{566} \approx 0.1413$$

$$p_B = 1 - p_A = \frac{486}{566} \approx 0.8587$$

Using the formula for system-of-mating inbreeding coefficient:

$$f = 1 - \frac{H_0}{H_{exp}} = 1 - \frac{P_{AB}}{2p_A p_B}$$

Expected heterozygosity under HWE:

$$H_{exp} = 2p_A p_B = 2 \cdot \frac{80}{566} \cdot \frac{486}{566} = \frac{77760}{320356} \approx 0.2428$$

Observed heterozygosity:

$$H_0 = P_{AB} = \frac{76}{566} \approx 0.1343$$

Therefore:

$$f = 1 - \frac{76/566}{77760/320356} = 1 - \frac{0.1343}{0.2428} = 0.447$$

Question 2

Calculate the inbreeding coefficient for individual I in the pedigree diagram

Solution

From the pedigree, identify common ancestors and paths from parent G to parent H through each ancestor.

Common ancestors: **A** and **B**

Paths through A:

- G $\leftarrow$ E $\leftarrow$ C $\leftarrow$ A $\rightarrow$ D $\rightarrow$ H: $n = 6$ individuals
- G $\leftarrow$ F $\leftarrow$ C $\leftarrow$ A $\rightarrow$ D $\rightarrow$ H: $n = 6$ individuals

Paths through B:

- G $\leftarrow$ E $\leftarrow$ C $\leftarrow$ B $\rightarrow$ D $\rightarrow$ H: $n = 6$ individuals
- G $\leftarrow$ F $\leftarrow$ C $\leftarrow$ B $\rightarrow$ D $\rightarrow$ H: $n = 6$ individuals

Using the formula $F_i = \left(\frac{1}{2}\right)^{n_i}(1 + F_{a_i})$ where ancestors are not inbred ($F_a = 0$):

$$F_I = 4 \times \left(\frac{1}{2}\right)^6 = 4 \times \frac{1}{64} = \frac{4}{64} = \frac{1}{16} = 0.0625$$

Question 3

Consider a population with inbreeding coefficient f. Let a bi-allelic locus have allele frequencies p and $q = 1-p$, and genotype frequencies (P_{11}, P_{12}, P_{22}). Prove that equation (A) below implies both (B) and (C):

$$(A) : P_{11} = p^2 + pqf \quad (B) : P_{12} = 2pq(1-f) \quad (C) : P_{22} = q^2 + pqf$$

Solution

We use two constraints:

1. **Genotype frequencies sum to 1:** $P_{11} + P_{12} + P_{22} = 1$
2. **Allele frequency of A_1 is preserved:** $p = P_{11} + \frac{1}{2}P_{12}$

Derive (B):

From constraint 2:

$$p = P_{11} + \frac{1}{2}P_{12} = (p^2 + pqf) + \frac{1}{2}P_{12}$$

Solving for P_{12}:

$$\frac{1}{2}P_{12} = p - p^2 - pqf = p(1 - p - qf) = p(q - qf) = pq(1-f)$$

$$P_{12} = 2pq(1-f)$$

Derive (C):

From constraint 1:

$$P_{22} = 1 - P_{11} - P_{12} = 1 - (p^2 + pqf) - 2pq(1-f)$$

$$= 1 - p^2 - pqf - 2pq + 2pqf$$

$$= 1 - p^2 - 2pq + pqf$$

Since $1 - p^2 - 2pq = (1-p)^2 - 2pq + 2pq - p^2 = q^2$ (using $q = 1 - p$):

$$= q^2 + pqf$$

$$P_{22} = q^2 + pqf$$

Thus, equation (A) together with the constraints implies both (B) and (C).

Question 4

In a diploid population of size 50, the frequencies of A_1 and A_2 are 0.7 and 0.3. What is the expected heterozygosity after ten generations?

Solution

From Eq. (8-9), heterozygosity decreases under genetic drift:

$$\bar{H}_t = \bar{H}_0 \left(1 - \frac{1}{2N}\right)^t$$

Given: $N = 50$, $t = 10$, $p_0 = 0.7$, $q_0 = 0.3$

Initial expected heterozygosity:

$$H_0 = 2p_0q_0 = 2(0.7)(0.3) = 0.42$$

After 10 generations:

$$H_{10} = 0.42 \times \left(1 - \frac{1}{2 \times 50}\right)^{10} = 0.42 \times \left(1 - \frac{1}{100}\right)^{10} = 0.42 \times (0.99)^{10}$$

$$(0.99)^{10} \approx 0.9044$$

$$H_{10} = 0.42 \times 0.9044 \approx 0.38$$

Question 5

An allele has a frequency of 0.01 in a population.

(a) If genetic drift is the only force operating, what is the probability under random mating that the allele will ultimately be lost if the (diploid) population size is 50?

(b) What is the probability of loss if the size of the (diploid) population is 5000?

Solution

From Eq. (8-3), under pure genetic drift:

$$\Pr\{\text{allele is fixed}\} = p_0$$

$$\Pr\{\text{allele is lost}\} = 1 - p_0$$

Given $p_0 = 0.01$, for $N = 50$ and $N = 5000$:

$$\Pr\{\text{loss}\} = 1 - 0.01 = 0.99$$

It's not related to population size.

Question 6

(a) What is the probability of loss of a new mutation in the first generation if all families are of size 2?

(b) What is the probability of loss of a new mutation in the first generation if the family size follows a Poisson distribution with mean 2?

(c) Explain the difference, if any, between your results in a. and b. above.

Solution

A new mutation appears in a heterozygote (Aa). Loss occurs if the mutant allele is not transmitted to any offspring.

(a) Fixed family size $k = 2$:

Probability of not transmitting a to one offspring $= \frac{1}{2}$

Probability of not transmitting a to either offspring:

$$\Pr\{\text{loss}|k = 2\} = \left(\frac{1}{2}\right)^2 = 0.25$$

(b) Poisson-distributed family size with mean 2:

From Eq. (8-7), using law of total probability:

$$\Pr\{\text{loss}\} = \sum_{k=0}^{\infty} \Pr\{\text{loss}|k\} \cdot \Pr\{K = k\} = \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^k \cdot \frac{e^{-2}2^k}{k!}$$

$$= e^{-2} \sum_{k=0}^{\infty} \frac{1^k}{k!} = e^{-2} \cdot e^{1} = e^{-1} \approx 0.368$$

(c) Explanation:

The probability of loss is **higher** with Poisson family size (0.368) than with fixed family size (0.25). This is because:

- With Poisson distribution, there is a probability $e^{-2} \approx 0.135$ of having **zero offspring** (guaranteed loss)
- The variance in family size introduces additional randomness that increases the chance of loss
- Fixed family size of 2 guarantees at least some chance of transmission

---

Question 7

For the subdivided population below:

	A_1A_1	A_1A_2	A_2A_2
subpopulation 1	10	180	810
subpopulation 2	250	500	250

(a) Calculate f_{is}, f_{it}, and f_{st}.

(b) Assume the two subpopulations are indistinguishable and calculate f for the combined population.

Solution

Part (a)

Subpopulation 1: $n_1 = 1000$

$$p_1 = \frac{2(10) + 180}{2(1000)} = \frac{200}{2000} = 0.1, \quad q_1 = 0.9$$

Expected heterozygosity: $H_{exp}^{(1)} = 2(0.1)(0.9) = 0.18$

Observed heterozygosity: $H_{obs}^{(1)} = \frac{180}{1000} = 0.18$

$$f_{is}^{(1)} = 1 - \frac{H_{obs}}{H_{exp}} = 1 - \frac{0.18}{0.18} = 0$$

Subpopulation 2: $n_2 = 1000$

$$p_2 = \frac{2(250) + 500}{2000} = \frac{1000}{2000} = 0.5, \quad q_2 = 0.5$$

Expected heterozygosity: $H_{exp}^{(2)} = 2(0.5)(0.5) = 0.5$

Observed heterozygosity: $H_{obs}^{(2)} = \frac{500}{1000} = 0.5$

$$f_{is}^{(2)} = 1 - \frac{0.5}{0.5} = 0$$

Average f_{is}: Since both subpopulations are in HWE:

$$f_{is} = 0$$

Calculate f_{st}:

Overall allele frequency (equal subpopulation sizes):

$$\bar{p} = \frac{0.1 + 0.5}{2} = 0.3, \quad \bar{q} = 0.7$$

Total expected heterozygosity: $H_T = 2\bar{p}\bar{q} = 2(0.3)(0.7) = 0.42$

Average subpopulation heterozygosity:

$$H_S = \frac{0.18 + 0.5}{2} = 0.34$$

$$f_{st} = \frac{H_T - H_S}{H_T} = \frac{0.42 - 0.34}{0.42} = \frac{0.08}{0.42} \approx 0.19$$

Calculate f_{it}:

Using $(1 - f_{it}) = (1 - f_{is})(1 - f_{st})$:

$$f_{it} = f_{st} + f_{is}(1 - f_{st}) = 0.19 + 0(1 - 0.19) = 0.19$$

Part (b)

Combined population (ignoring subdivision):

Total: $n = 2000$ - A_1A_1: $10 + 250 = 260$ - A_1A_2: $180 + 500 = 680$ - A_2A_2: $810 + 250 = 1060$

$$p = \frac{2(260) + 680}{4000} = \frac{1200}{4000} = 0.3$$

Expected heterozygosity: $H_{exp} = 2(0.3)(0.7) = 0.42$

Observed heterozygosity: $H_{obs} = \frac{680}{2000} = 0.34$

$$f = 1 - \frac{H_{obs}}{H_{exp}} = 1 - \frac{0.34}{0.42} = 0.19$$

Question 8

Assume that one population has 5 alleles at a particular locus, each with a frequency of 0.2. A second population has five different alleles (i.e., different from the set of alleles in the first population) at the same locus, each with a frequency of 0.2. Calculate f_{st} for this locus and comment on your result.

Solution

Population 1: Alleles A_1, A_2, A_3, A_4, A_5, each at frequency 0.2

Population 2: Alleles B_1, B_2, B_3, B_4, B_5, each at frequency 0.2

Total population (assuming equal sizes): 10 alleles, each at frequency 0.1

Subpopulation expected heterozygosity:

$$H_S = 1 - \sum p_i^2 = 1 - 5(0.2)^2 = 1 - 0.2 = 0.8$$

(Same for both populations)

Total expected heterozygosity:

$$H_T = 1 - \sum \bar{p}_i^2 = 1 - 10(0.1)^2 = 1 - 0.1 = 0.9$$

$$f_{st} = \frac{H_T - H_S}{H_T} = \frac{0.9 - 0.8}{0.9} = \frac{0.1}{0.9} \approx 0.11$$

Comment: Despite the two populations having **completely non-overlapping allele sets** (maximum possible differentiation), f_{st} is only 0.11. This illustrates a limitation of F_{ST}: when loci are highly polymorphic, F_{ST} underestimates differentiation because both H_S and H_T approach 1.

Question 9

(a) What is the probability that two alleles coalesced 10 generations before the present in a diploid population of size $N = 50$?

(b) What is the expected time in a diploid population of size $N = 5$ that there are exactly 10 lineages?

(c) What is the expected time in a diploid population of size $N = 5$ that there are exactly 2 lineages?

Solution

(a) Probability of coalescence exactly at generation 10:

For $n = 2$ alleles, from Eq. (8-17), coalescence time follows a geometric distribution:

$$\Pr\{T_2 = t\} = (1 - P_2)^{t-1} P_2$$

where $P_2 = \frac{1}{2N} = 0.01$

$$\Pr\{T_2 = 10\} = (1 - 0.01)^9 \times 0.01 \approx 0.9135 \times 0.01 = 9.14 \times 10^{-3}$$

(b) Expected time with exactly 10 lineages ($N = 5$):

From Eq. (8-18), expected time for coalescence from j to $j - 1$ lineages:

$$E[T_j] = \frac{4N}{j(j-1)} = \frac{20}{10 \times 9} = \frac{20}{90} = 0.22 \text{ generations}$$

Since the calculated expectation (0.22 generations) is less than 1, this implies that the probability of a coalescence event occurring in the very first generation is extremely high (approaching 100%). Therefore, the state of having 'exactly 10 lineages' will most likely last for only 1 generation.

(c) Expected time with exactly 2 lineages ($N = 5$):

$$E[T_2] = \frac{4N}{2(2-1)} = \frac{4 \times 5}{2} = 10.0 \text{ generations}$$

Question 10

Consider a haploid population with size $2N$ in mutation-drift balance. Assuming an infinite-alleles model, use the coalescent approach to prove that the probability of IBD between two gene copies is:

$$\bar{F}_{eq} = \frac{1}{1 + 4N\mu}$$

Solution

Consider two randomly chosen alleles. Going back in time, at each generation:

- **Probability of coalescence** (both alleles derived from same ancestor): $\frac{1}{2N}$
- **Probability of mutation** in either lineage: 2μ (since either can mutate)
- **Probability of neither event**: $1 - \frac{1}{2N} - 2\mu$

For two alleles to be IBD, they must coalesce **before** either experiences a mutation.

The probability that the **first event** back in time is a coalescence (rather than mutation):

$$\bar{F}_{eq} = \frac{\Pr(\text{coalescence})}{\Pr(\text{coalescence}) + \Pr(\text{mutation})} = \frac{\frac{1}{2N}}{\frac{1}{2N} + 2\mu}$$

$$= \frac{1}{1 + 4N\mu} = \frac{1}{1 + \theta}$$

where $\theta = 4N\mu$ is the population-scaled mutation rate.

Question 11

Assume $h = 0.03$ for a mutation that is lethal when homozygous and $h = 0.4$ for a mildly deleterious mutation when homozygous ($s = 0.05$). In both cases, the mutation rate to the deleterious form is $\mu = 10^{-5}$.

(a) Assuming random mating, what are the equilibrium frequencies for the mutant allele in these two cases?

(b) Assuming $f = 0.1$ and that both mutants are now completely recessive, what are the equilibrium frequencies for the mutant allele in these two cases?

Solution

Part (a): Random mating ($f = 0$)

Case 1: Lethal homozygote ($s = 1$, $h = 0.03$)

Using Eq. (10-7) for partial dominance:

$$\tilde{q} = \frac{\mu}{hs} = \frac{10^{-5}}{(0.03)(1)} = \frac{10^{-5}}{0.03} = 3.3 \times 10^{-4}$$

Case 2: Mildly deleterious ($s = 0.05$, $h = 0.4$)

$$\tilde{q} = \frac{\mu}{hs} = \frac{10^{-5}}{(0.4)(0.05)} = \frac{10^{-5}}{0.02} = 5.0 \times 10^{-4}$$

Part (b): Inbreeding with $f = 0.1$, completely recessive ($h = 0$)

Using Eq. (10-8) for completely recessive alleles with inbreeding:

$$\tilde{q} = \frac{\mu}{fs}$$

Case 1: Lethal ($s = 1$)

$$\tilde{q} = \frac{10^{-5}}{(0.1)(1)} = \frac{10^{-5}}{0.1} = 1.0 \times 10^{-4}$$

Case 2: Mildly deleterious ($s = 0.05$)

$$\tilde{q} = \frac{10^{-5}}{(0.1)(0.05)} = \frac{10^{-5}}{0.005} = 2.0 \times 10^{-3}$$

Comment: Inbreeding with completely recessive alleles leads to lower equilibrium frequencies for lethal alleles but higher frequencies for mildly deleterious alleles compared to random mating with partial dominance. This is because inbreeding exposes recessive alleles to selection more efficiently.