COLUMBIA UNIVERSITY

DEPARTMENT OF BIOSTATISTICS

P8149 – HUMAN POPULATION GENETICS

Exercise Sheet 4 (Model Answers)

Question 1 [1 mark]

The frequency of allele A is

$$p_A = \frac{2(42)+76}{2(42+76+448)} = .141$$

$$p_B = 1-.141 = .859$$

The inbreeding coefficient is

$$f = 1-\frac{H_0}{H_e} = 1-\frac{76/(42+76+448)}{1-.141^2-.859^2} = .446$$

Question 2 [2 marks]

Common ancestor	path	n	F_{ai}	F_i
A	GECADH	6	0	$(1/2)^6$=1/64
A	GFCADH	6	0	1/64
B	GECBDH	6	0	1/64
B	GFCBDH	6	0	1/64
Total				4(1/64)=.0625

The inbreeding coefficient of individual I is .0625

Question 3 [1 mark]

$\underline{A \Rightarrow B}$

$$
\begin{aligned}
p &= P_{11} + \frac{1}{2} P_{12} \\
\therefore P_{12} &= 2p - 2P_{11} \\
&= 2p - 2p^2 - 2pqf \\
&= 2p(1-p) - 2pqf \\
&= 2pq - 2pqf \\
&= 2pq(1-f)
\end{aligned}
$$

$\underline{A \Rightarrow C}$

Since we have proved that $A \Rightarrow B$, to prove $A \Rightarrow C$ we only need to prove $B \Rightarrow C$

$$
\begin{aligned}
q &= P_{22} + \frac{1}{2} P_{12} \\
\therefore P_{22} &= q - P_{12}/2 \\
&= q - pq(1-f) \\
&= q - pq + pqf \\
&= q(1-p) + pqf \\
&= q^2 + pqf
\end{aligned}
$$

Question 4 [2 marks]

The initial heterozygosity is

$$H_0 = 2pq = 2(.7)(.3) = .42$$

Therefore, expected heterozygosity after 10 generations is

$$
\begin{aligned}
H_{10} &= \left(1 - \frac{1}{2N}\right)^{10} H_0 \\
&= \left(1 - \frac{1}{100}\right)^{10} (.42) \\
&= .38
\end{aligned}
$$

Question 5 [1 mark]

(a) Probability of loss = $1-p_0 = 1-.01 = .99$ (independent of population size)

(b) Probability of loss = $1-p_0 = 1-.01 = .99$ (independent of population size)

Question 6 [2 marks]

(a) Probability of loss of a new mutation = $\left(\frac{1}{2}\right)^2 = \frac{1}{4}$.

(b) Probability of loss of a new mutation = $e^{-1} = .368$

(c) In (b) above, it is possible for the family size to be less than 2, in which case there is a higher probability of the loss of a new mutation.

Question 7 [3 marks]

(a) The genotypic and allele frequencies in the two subpopulations are

	A_1A_1	A_1A_2	A_2A_2	A_1
sub. 1	.01	.18	.81	.1
sub. 2	.25	.5	.25	.5

The value of f_{is} in each subpopulation is

$$f_{is}^{(1)} = 1-\frac{H_o}{H_e} = 1-\frac{.18}{2(.1)(.9)} = 0,$$

$$f_{is}^{(2)} = 1-\frac{.5}{2(.5)(.5)} = 0$$

Therefore, the overall value of f_{is} is

$$f_{is} = \frac{\sum_i w_i p_i q_i f_{is}^{(i)}}{\sum_i w_i p_i q_i} = 0$$

The value of f_{st} is

$$
\begin{aligned}
f_{st} &= \frac{\sigma_p^2}{\bar{p}\,\bar{q}} \\
&= \frac{\sum_i w_i \left\{ p^{(i)} \right\}^2 - \bar{p}^2}{\bar{p}\bar{q}} \\
&= \frac{\frac{1}{2}\left(.1^2 + .5^2\right) - .3^2}{(.3)(.7)} \\
&= .19
\end{aligned}
$$

Hence, we have

$$
\begin{aligned}
f_{it} &= f_{st} + \left(1 - f_{st}\right) f_{is} \\
&= .19 + .81(0) \\
&= .19
\end{aligned}
$$

(b) When the subpopulations are indistinguishable, the genotype counts are:

	A_1A_1	A_1A_2	A_2A_2	total
	260	680	1060	2000

The frequency of allele A_1 is

$$
p = \frac{260(2) + 680}{2(2000)} = .3
$$

and the value of the system-of-mating inbreeding is

$$
\begin{aligned}
f &= 1 - \frac{H_{obs}}{H_{\exp}} \\
&= 1 - \frac{680/2000}{2(.3)(.7)} \\
&= .19
\end{aligned}
$$

Question 8 [2 marks]

The alleles in the two populations, together with their respective frequencies, can be represented as follows:

population 1					population 2				
A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
.2	.2	.2	.2	.2	.2	.2	.2	.2	.2

The expected heterozygosities in each population are

$$H_s^{(1)} = 1 - \sum_{i=1}^{5} p_i^2 = 1 - 5(.2)^2 = .8$$
$$H_s^{(2)} = 1 - \sum_{i=6}^{10} p_i^2 = 1 - 5(.2)^2 = .8$$

Therefore,

$$H_s = \frac{H_s^{(1)} + H_s^{(2)}}{2} = .8$$

For the total expected heterozygosity, we assume the two populations are merged, giving an allele frequency of .1 for each allele and

$$H_t = 1 - 10(.1)^2 = .9$$

Therefore,

$$f_{st} = 1 - \frac{H_s}{H_t} = 1 - \frac{.8}{.9} = .11$$

We see that the value of f_{st} is relatively low, although the two populations have non-overlapping alleles. We thus see that f_{st} is a poor measure of population differentiation when allele frequencies are low.

Question 9 [2 marks]

(a) The probability of no coalescence one generation back is

$$1-P=1-\frac{1}{100}=.99$$

Therefore, the probability that coalescence occurred 10 generations back is

$$(1-P)^9 P=(.01)^9(.99)=9.14\times10^{-3}$$

(b) The probability of no coalescence one generation back is

$$1-P=\left(1-\frac{1}{10}\right)\left(1-\frac{2}{10}\right)-\ldots-\left(1-\frac{9}{10}\right)=3.63\times10^{-4}$$

The probability of a coalescence is then

$$P=1-3.63\times10^{-4}=1.00$$

Hence the expected time during which there are exactly 10 lineages is $\frac{1}{P}=1$.

(c) The probability of no coalescence one generation back is

$$1-P=\left(1-\frac{1}{10}\right)=.9$$

Therefore, the expected time during which there are two lineages is

$$\frac{1}{P}=\frac{1}{.1}=10\,.$$

Question 10 [2 marks]

The probability of a coalescence is

$$P_{coal}=\frac{1}{2N}\,.$$

Since there are two lineages available for mutation, the probability of a mutation is

$$P_{mut}=2\mu\,.$$

Hence, the probability of ibd between the two genes copies is the probability that a coalescence occurs before a mutation occurs, which is

$$\frac{P_{coal}}{P_{coal}+P_{mut}} = \frac{1/(2N)}{2\mu + 1/(2N)} = \frac{1}{1+4\mu N},$$

as required.

Question 11 [2 marks]

(a) When $h = .03$, $s = 1$, and the equilibrium frequency of mutant allele is

$$\tilde{q} \approx \frac{u}{hs} = \frac{10^{-5}}{(.03)(1)} = 3.3\times 10^{-4}$$

When $h = .4$, $s = .05$, and the equilibrium frequency of mutant allele is

$$\tilde{q} \approx \frac{u}{hs} = \frac{10^{-5}}{(.4)(.05)} = 5\times 10^{-4}$$

(b) We have $f = .1$, $h = 0$ and $s = 1$, and the equilibrium frequency of mutant allele is

$$\tilde{q} \approx \frac{u}{fs} = \frac{10^{-5}}{(.1)(1)} = 1.0\times 10^{-4}$$

Also for the second case, $f = .1$, $h = 0$ and $s = .05$, and the equilibrium frequency of mutant allele is

$$\tilde{q} \approx \frac{u}{fs} = \frac{10^{-5}}{(.1)(.05)} = 2.0\times 10^{-3}$$

Dr. P Gorroochurn: 12/13/2024