STA6625 (AU2016)
Homework #3

Ezgi Karaesmen

09/21/16

Chapter 10

Problem 10.8

Total genotype counts for both SNP 1 and SNP 2 loci are given below:

		SNP 2			
		CC	CT	TT	Total
	GG	1	4	0	5
SNP 1	GA	3	3	1	7
	AA	0	2	1	3
	Total	4	9	2	15

10.8.1

Assuming that HWE holds for this population, the genome frequencies were calculated as follows:

$$\hat{P}_{GG} = \frac{5}{15} = 0.333 \qquad\qquad \hat{P}_{CC} = \frac{4}{15} = 0.266$$

$$\hat{P}_{GA} = \frac{7}{15} = 0.466 \qquad\qquad \hat{P}_{CT} = \frac{9}{15} = 0.60$$

$$\hat{P}_{AA} = \frac{3}{15} = 0.20 \qquad\qquad \hat{P}_{TT} = \frac{2}{15} = 0.133$$

And the allele frequencies were computed as follows:

$$\hat{P}_{G} = \sqrt{\hat{P}_{GG}} = \sqrt{0.333} = 0.577 \qquad\qquad \hat{P}_{C} = \sqrt{\hat{P}_{CC}} = \sqrt{0.266} = 0.516$$

$$\hat{P}_{A} = \sqrt{\hat{P}_{AA}} = \sqrt{0.20} = 0.447 \qquad\qquad \hat{P}_{T} = \sqrt{\hat{P}_{TT}} = \sqrt{0.133} = 0.365$$

10.8.2

The haplotypes can be determined only for individuals that carry at least one homozygous genotype for either one of the loci. A table for the allele counts for both loci (excluding individuals that are heterogenous for both loci) is shown below:

		SNP 2		
		C	T	Total
	G	9	5	14
SNP 1	A	5	5	10
	Total	14	10	24

10.8.3

Similarly the haplotype frequencies can be calculated as follows and are shown below:

		SNP 2		
		C	T	Total
	G	$9/24 = 0.375$	$5/24 = 0.208$	$14/24 = 0.583$
SNP 1	A	$5/24 = 0.208$	$5/24 = 0.208$	$10/24 = 0.416$
	Total	$14/24 = 0.583$	$10/24 = 0.416$	$24/24 = 1$

Therefore, LD measures D, D' and Δ^2 between SNPs 1 and 2 can be determined as follows:

$$D = D_{GC} = D_{AC} = D_{GT} = D_{AT}$$
$$\hat{D} = \hat{D}_{GC} = \hat{P}_{GC} - \hat{P}_G \cdot \hat{P}_C$$
$$= 0.375 - 0.583 \cdot 0.583$$
$$= 0.035$$

$$D'_{GC} = \begin{cases} \dfrac{D_{GC}}{max\{-P_A P_T, -P_G P_C\}} & \text{if } D < 0, \\ \dfrac{D_{GC}}{min\{P_G P_T, P_A P_C\}} & \text{if } D > 0. \end{cases}$$

$$\hat{D}'_{GC} = \frac{\hat{D}_{GC}}{min\{\hat{P}_G \hat{P}_T, \hat{P}_A \hat{P}_C\}}$$
$$= \frac{0.035}{min\{0.583 \cdot 0.416, 0.416 \cdot 0.583\}}$$
$$= \frac{0.035}{0.243} = 0.145$$

$$r^2_{GC} = \Delta^2 = \frac{D^2_{GC}}{P_G P_A P_C P_T}$$
$$= \frac{0.035^2}{0.583 \cdot 0.416 \cdot 0.583 \cdot 0.416}$$
$$= \frac{0.0012}{0.0588} = 0.0204$$

Extra Problems

Problem 1

The LD coefficient of the current generation can be determined by the function $D_t = (1 - \theta)^t D_0$, where D_0 is the LD coefficient at the time mutation was introduced, t is the number of generations since the mutation and θ is the recombination fraction. An R function was generated to compute the D_t for different number of generations and recombination fractions:

```
Dt_foo = function(D0, theta, t){round((1-theta)^t * D0,4)}

Dt_foo(D0 = 0.999, theta = 0.0001, t = 200)
Dt_foo(D0 = 0.999, theta = 0.01, t = 200)
Dt_foo(D0 = 0.999, theta = 0.05, t = 200)

Dt_foo(D0 = 0.999, theta = 0.0001, t = 100)
Dt_foo(D0 = 0.999, theta = 0.01, t = 100)
Dt_foo(D0 = 0.999, theta = 0.05, t = 100)
```

D_t results for different number of generations and recombination fractions are shown in a tabular form:

		Generations	
		200	100
Recombination	0.0001	0.9792	0.9891
fractions	0.01	0.1338	0.3657
	0.05	0	0.0059

Therefore,

1. For an original LD coefficient of 0.999, recombination fraction of 0.0001 and 200 generations $D_{200} = 0.9792$.

2. For an original LD coefficient of 0.999, recombination fraction of 0.01 and 200 generations $D_{200} = 0.1338$.

3. For an original LD coefficient of 0.999, recombination fraction of 0.05 and 100 generations $D_{100} = 0.0059$.

Problem 2

For the given haplotype and allele probabilities:

$$\hat{P}_{ab} = 0.60 \qquad\qquad \hat{P}_b = 0.60 + 0.10 = 0.70$$

$$\hat{P}_{aB} = 0.25 \qquad\qquad \hat{P}_B = 0.25 + 0.05 = 0.30$$

$$\hat{P}_{Ab} = 0.10 \qquad\qquad \hat{P}_a = 0.60 + 0.25 = 0.85$$

$$\hat{P}_{AB} = 0.05 \qquad\qquad \hat{P}_A = 0.10 + 0.05 = 0.15$$

Part a: Calculate the linkage disequilibrium coefficient D

$$D = D_{AB} = D_{Ab} = D_{aB} = D_{ab}$$

$$\hat{D} = \hat{D}_{AB} = \hat{P}_{AB} - \hat{P}_A \cdot \hat{P}_B$$

$$= 0.05 - 0.15 \cdot 0.30$$

$$= 0.005$$

Part b: Determine the range of $\hat{D}$

$$max\{-P_a\ P_b\ ,\ -P_A\ P_B\} \le D_{AB} \le min\{P_A\ P_b\ ,\ P_a\ P_B\}$$

$$max\{-0.85 \cdot 0.70\ ,\ -0.15 \cdot 0.30\} \le D_{AB} \le min\{0.15 \cdot 0.70\ ,\ 0.85 \cdot 0.30\}$$

$$max\{-0.595\ ,\ -0.045\} \le D_{AB} \le min\{0.105\ ,\ 0.255\}$$

$$-0.045 \le D_{AB} \le 0.105$$

Part c: Calculate the standardized coefficient D'

$$D' = \begin{cases} \dfrac{D}{max\{-P_aP_b\ ,\ -P_AP_B\}} & \text{if } D < 0, \\[2mm] \dfrac{D}{min\{P_AP_b\ ,\ P_aP_B\}} & \text{if } D > 0. \end{cases}$$

$$D > 0 ;$$

$$\hat{D}' = \frac{\hat{D}}{min\{\hat{P}_A\hat{P}_b\ ,\ \hat{P}_a\hat{P}_B\}}$$

$$= \frac{0.005}{min\{0.15 \cdot 0.70, 0.85 \cdot 0.30\}}$$

$$= \frac{0.005}{0.105} = 0.476$$

Part d: Calculate the linkage disequilibrium measure r^2 (Δ^2)

$$r^2 = \Delta^2 = \frac{D^2}{P_A P_a P_B P_b}$$

$$= \frac{0.005^2}{0.15 \cdot 0.0.85 \cdot 0.30 \cdot 0.70}$$

$$= 0.00093$$

Part e: Determine whether there is significant evidence of LD by using the asymptotic score test.

For $H_0 : D = 0$ $H_a : D \neq 0$, where rejecting H_0 indicates that the loci are in linkage disequilibrium. The function for the asymptotic score test statistis is given below:

$$T_s = \frac{D}{\sqrt{\frac{1}{2n}(P_A P_a P_B P_b)}}$$

T_s has asymptotic standard normal under H_0. Therefore, the p-value is obtained via `pnorm()` function in **R**.

$$T_s = \frac{0.005}{\sqrt{\frac{1}{2 \cdot 100}(0.15 \cdot 0.85 \cdot 0.30 \cdot 0.70)}}$$

$$= \frac{0.0050}{0.0116}$$

$$= 0.4321$$

The p-value of $T_s = 0.4321$ was determined with the following **R** function:

```
round(pnorm(0.4321, lower.tail = F), 4)
[1] 0.3328
```

P-value was determined to be 0.3328 and the H_0 could not be rejected, indicating that the loci Locus 1 and Locus 2 are not in LD.