

Projekt 2 – zadania do wykonania

Uwagi ogólne:

- projekt należy oddać w formie papierowej oraz na płycie CD do 7.06.2009r.,
- na płycie CD mają się znaleźć WSZYSTKIE wykorzystane pliki,
- płytę CD należy podpisać imieniem i nazwiskiem,
- jeżeli plik ma nazwę narzuconą w zadaniu, a na płycie będzie zapisany pod inną nazwą, będzie to miało negatywny wpływ na ocenę z projektu,
- wykresy i tabele należy opisać.

Zadanie 1

Plik *daneXX.prn* należy przekształcić w taki sposób, aby był plikiem wejściowym do programu PHASE oraz zapisać pod nazwą *danePHASE.inp*. Uruchomić program PHASE tak, aby pliki wynikowe miały nazwę *wynikPHASE*.

Zadanie 2

Na podstawie plików wynikowych z poprzedniego zadania:

- (a) podać haplotyp o największej frekwencji,
- (b) podać haplotyp o najmniejszej frekwencji,
- (c) przedstawić ogólną interpretację wyników estymacji frekwencji haplotypów,
- (d) podać liczbę możliwych haplotypów,
- (e) podać pary (lub parę) haplotypów dla osobnika nr XX, 33, 23, 198, 299 (XX to numer danych) wraz prawdopodobieństwami ich wystąpienia oraz zinterpretować wyniki.

Zadanie 3

Plik *daneXX.prn* należy przekształcić w taki sposób, aby był plikiem wejściowym do programu EMLD oraz zapisać pod nazwą *daneEMLD.dat*. Uruchomić program EMLD.

Zadanie 4

Napisać program w pakiecie R, który:

- (a) wczyta frekwencje haplotypów otrzymana przy pomocy programu PHASE jako tablicę o nazwie *haplotypy*,
- (b) poda wymiar wczytanej tablicy,
- (c) wyświetli wartości błędów popełnionych przez program PHASE przy estymacji frekwencji haplotypów,
- (d) obliczy średnią frekwencję haplotypów,
- (e) obliczy medianę frekwencji haplotypów,
- (f) poda liczbę haplotypów o frekwencji większej od 0.XX (XX to numer danych).

Zinterpretować wynik otrzymany w punkcie (d) i (e). Zapisać program pod nazwą *haplo.R*.

Zadanie 5

Napisać program w pakiecie R, który:

- (a) wczyta plik *HapFreq.txt* jako tablicę o nazwie *frekwencje*,
- (b) obliczy średnie frekwencje haplotypów (11, 12, 21, 22) dla wszystkich par markerów i zapisze je jako tablicę *średnia*,
- (c) obliczy odchylenia standardowe frekwencji haplotypów (11, 12, 21, 22) dla wszystkich par markerów i zapisze je jako tablicę *odchylenie*,
- (d) połączy tablice *średnia* i *odchylenie* kolumnami i zapisze utworzoną tablicę w pliku o nazwie *statystyki.txt*,
- (e) obliczy wartość statystyki D dla wybranej pary markerów,
- (f) obliczy wartość statystyki D' dla wybranej pary markerów,
- (g) obliczy wartość statystyki r^2 dla wybranej pary markerów,
- (h) obliczy wartość statystyki D dla wszystkich par markerów i zapisze wyniki do pliku o nazwie *D.txt*,
- (i) wczyta plik *LD.txt* jako tablicę o nazwie *LD*,
- (j) wyświetli wartość statystyki r^2 pomiędzy markerami znajdującymi się najbliżej równowagi sprzężeń,
- (k) wyświetli wartość statystyki r^2 pomiędzy markerami znajdującymi się w największej nierównowadze sprzężeń,
- (l) sprawdzi czy statystyka D przyjmuje wartość świadczącą o ścisłej równowadze sprzężeń.

Zinterpretować wynik otrzymany w punkcie (e), (f), (g). Zapisać program jako *AnalizaSprzężeń.R*.