

Statystyka w Zastosowaniach

Laboratorium 4

Lokalizacja genów z wykorzystaniem testu Studenta

1. Wygeneruj macierz genotypów dla $n=500$ osobników z krzyżówki wstecznej na 3 chromosomach o długości 150 cM i odstępem między sąsiednimi markerami równym $\Delta = 1cM$. Zweryfikuj poprawność poprzez porównanie korelacji próbkowej między genotypami pierwszego i piątego markera na pierwszym chromosomie z wartością teoretyczną tej korelacji.
2. Wygeneruj wektor wartości cechy dla $n=500$ z rozkładu $N(0,1)$ [cecha nie jest zależna od czynników genetycznych] i narysuj wykresy przedstawiające zależność wartości statystyki Studenta do testowania hipotezy o braku zależności między cechą a genotypem od odległości od lewego końca chromosomu. Wykonaj osobne wykresy dla każdego chromosomu.
3. Powtórz powyższy eksperyment 500 razy i wyestymuj p-stwo co najmniej jednej błędnej detekcji w sytuacji gdy
 - a) Każdy z testów Studenta jest wykonany na poziomie istotności 0.05.
 - b) Zastosowano korektę Bonferroniego w celu kontroli p-stwa błędnych detekcji (Family Wise Error Rate, FWER) na poziomie 0.05.
 - c) Zastosowano wartość krytyczną t_c dla $FWER = \alpha = 0.05$ zgodnie z korektą wynikającą ze wzoru Feingolda, Browna i Siegmunda (1993):

$$\alpha \approx 1 - \exp \left[-2C \{1 - \Phi(t_c)\} - 0.04Lt_c\phi(t_c)\nu(t_c\sqrt{0.04\Delta}) \right] , \quad (1)$$

gdzie C jest liczbą chromosomów, L jest długością chromosomu (w cM), ϕ jest gęstością a Φ dystrybuantą standardowego rozkładu normalnego, a $\nu(t)$ jest zadana wzorem

$$\nu(t) = 2t^{-2} \exp \left\{ -2 \sum_{n=1}^{\infty} n^{-1} \Phi(-|t|n^{1/2}/2) \right\} \approx \frac{(2/t)(\Phi(t/2) - 0.5)}{(t/2)\Phi(t/2) + \phi(t/2)} . \quad (2)$$

- d) Zastosowano test permutacyjny (w oparciu o 1000 permutacji wektora Y).