A C++ program for partitioning additive and non-additive genetic effects to alleles and genotypes

Bryan D. Neff and Trevor E. Pitcher

Department of Biology, University of Western Ontario, London N6A 5B7, Canada
(bneff@uwo.ca) Phone: (519) 850-2532; Fax: (519) 661-2014

17 Nov 2005

Description

This program partitions variance in familial data, such as offspring body sizes or survivorship, to both additive (s_i+d_j; i.e. sum of sire and dam additive effects) and non-additive (I_{ij}) genetic effects at a candidate locus. The program fits the familial data to the following function:

$$z_k = \mu + \tfrac{1}{4}\sum_i\sum_j\left(s_i + d_j\right) + n\sum_i\sum_{j\neq i} I_{ij} + e_k;$$

where z_k is the predicted mean phenotype of the kth family, μ is the mean phenotype across all families, n is the expected proportion of heterozygotes in the kth family assuming Mendelian segregation of sire (i) and dam (j) alleles, and e_k is the deviation of the observed mean phenotype of the kth family from the model's prediction.

The program calculates the values of s_i, d_j and I_{ij} that minimized the sum of the squares of the deviations e_k from all families. Based on this best fit, the coefficient of determination (r^2) is also calculated. To determine the significance of the model, the program employs a randomization routine. The routine randomizes the familial phenotypic data and re-calculates values of s_i, d_j and I_{ij}, and the coefficient of determination. The routine is repeated for a total of 1000 iterations and the resulting distributions are used to determine the significance of each value calculated from the original observed data. The program is written in C++ and is an executable file (Neff&Pitcher_GA.exe). It is run by double clicking the icon or using the 'Run' command line.

Directions

Before running the program one data file must be set-up. This is most easily done using Microsoft Excel (or equivalent). The file must be saved as a **tab delimited** file and placed in the same file folder (directory) as the C++ program. The following data must be entered into the spreadsheet:

(1) Input: gadata.txt

row 1: number of alleles, number of families, lower bound on parameters (i.e. the estimates of s_i, d_j and I_{ij}), upper bound on parameters, number of allowable routine calls
row 2: dam allele 1, dam allele 2, sire allele 1, sire allele 2, familial phenotype value
.
.

.

The following is a sample spreadsheet for data.txt. The example assumes 3 alleles, 10 families, a lower bound of –1, and upper bound of +1, 5000 allowable routine calls. The first dam and sire have genotypes 3/3 and 1/3 respectively, and their family had a survivorship of 91%. The second dam and sire have genotype 2/3 and 1/1, respectively, and their family had a survivorship of 81%.

3	10	-1	1	5000
3	3	1	3	0.91
2	3	1	1	0.81
1	3	2	2	0.39
1	2	1	2	0.78
1	2	2	3	0.77
1	1	3	3	0.76
2	3	3	3	0.86
2	3	1	2	0.37
1	3	2	3	0.28
3	3	3	3	0.91

The parameter estimates were restricted to –1 to +1 because survivorship values must reside between these two values. The allowable routine calls should be set to a minimum of 1000 and should be increased if an error message appears on the screen during execution ("Maximum function calls exceeded"). The alleles must be numbered consecutively from 1 through to the number of alleles. In this example, they must be numbered from 1 through 3.

(2) Output: garesults.txt

The output is displayed on the screen and written to the file garesults.txt. This file can be viewed using Excel or other similar programs. The file contains two columns of data and the following parameters in the rows. First, the coefficient of determination is displayed (r^2) followed by the p-value associated with the overall model (i.e. is the r^2 significantly greater than zero). The second row has the constant (mean phenotypic value). The third has the additive genetic value of allele 1 and the associated p-value. The fourth row has the additive genetic value of allele 2 and the associated p-value. The fifth row has the additive genetic value of allele 3 and the associated p-value. The sixth row has the non-additive genetic value of alleles 1 + 2 and the associated p-value. The seventh row has the non-additive genetic value of alleles 1 + 3 and the associated p-value. Finally, the eighth row has the non-additive genetic value of alleles 2 + 3 and the associated p-value.

Because of the nature of a Monte Carlo simulation (used by the program to estimate the p-values), some fluctuations can occur in the probabilities reported each time the program is run with the same data. It is recommended that the program be run at least twice to ensure that the parameter values (i.e. the estimates of s_i, d_j and I_{ij}) do not change and to confirm the p-values. If parameter values changes, this is an indication of program failure and should be reported to the authors.

When the program is run with the data file above, the following output should be obtained:

		0.4419	0.553
		0.684	
	1	0.4486	0.26
	2	-0.0064	0.497
	3	0.1286	0.045
1	2	-0.6925	0.231
1	3	-0.47	0.24
2	3	-0.4225	0.26

Note: the p-values will fluctuate marginally with each run, but the parameter values should not change.