MAN Package for pedigree analysis.

Contents.

Introduction.

The Program Package MAN, written in Visual C++, is intended to perform pedigree analysis of quantitative traits and genetic markers on human (!) pedigrees of practically any structure having no consanguinity loops.

Besides direct procedures of the analysis, the package contains a number of tools, which allow to edit in various ways the initial data, including adjustment and construction of new derived traits.

There are also several graphical options to present pedigree structures, trait scatter plots and results of analysis.

The simulation options enable to estimate the power of the analyses for samples of any definite structure and size and construct the null-distributions for cases, when the appropriate distribution is not known or sufficiently deviates from asymptotic one.

1. Operations with pedigree data

1.1 Data input options

1.1.1 Import from file.

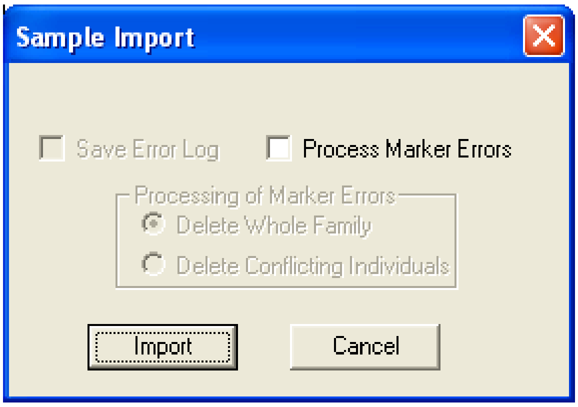
It is possible to import the pedigree data from various formats of the data file, including the text file (Tab-delimited and formatted Space-delimited), Dbase (.dbf), EXCEL (.xls). Each imported file should be organized as a table.

There are six obligatory fields(columns), defining the pedigree structure:

1) Code of pedigree;
2) Code of individual;
3) Code of his/her father;
4) Code of his/her mother;
5) Sex (1-for male, 2-for female);
6) Proband status (0-proband, 1-potential proband, 2 -for others);

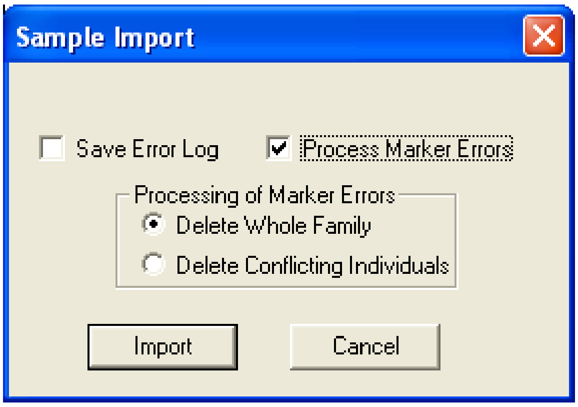
These fields can be arranged in any order. Traits have to be presented in the table as separate columns, individual data as rows. Data can be numerical or textural.

The table should be sorted according to “Code of pedigree”. If you want to include some unrelated individuals in the table, they should have the empty “Code of pedigree”-field and should be placed before all other individuals. At least one pedigree should be presented in the table. If your data contain only unrelated individuals, add after the end of data two rows for dummy father and mother of the last individual and fill correspondingly the obligatory fields for the three last rows of the table.



□ Save Error Log □ Process Marker Errors

Processing of Marker Errors
● Delete Whole Family
○ Delete Conflicting Individuals



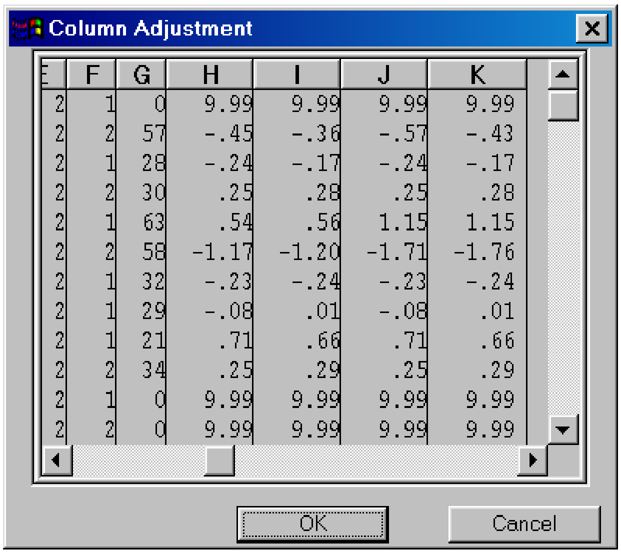
□ Save Error Log ☑ Process Marker Errors

Processing of Marker Errors
● Delete Whole Family
○ Delete Conflicting Individuals

By default the "Marker" data are strings, consisting of two allele names divided with "/"- sign, for example "A1/A2" or "258/272". This format will be recognized automatically. For dialelic markers with allele names being one symbol text values (for example “A” and “G”) the following denotations are possible: “AA”,”GG” or “A”,”G” for homozygous individuals, “AG” or “Both” for heterozygous and empty string for missing data.

There are two options concerning to Mendelian error processing on the import stage: 1) delete genotypes of the whole nuclear family having Mendelian conflicts, 2) delete only conflicting individuals. If desirable, you can create the error log file in Excel format.

If a not-empty document window is active, click on the "New" option in the "Sample" menu. Then click on the "Import" option in the "Sample" menu. The “Sample Import” dialog appears. If a part of your data contains marker genotypes, you can check the “Process Marker Error” option and choose, how to process marker errors (options 1 or 2). If you want to see the error log, check the “Save Error Log” checkbox. If you do not want to process marker errors on the stage of import, all genotypes will be included in the sample.


Column Adjustment

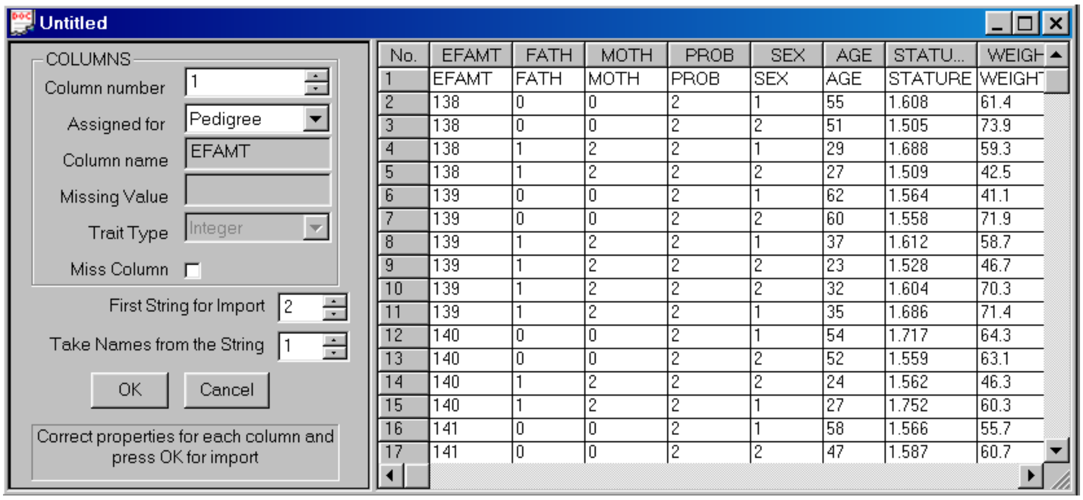
E	F	G	H	I	J	K
2	1	0	9.99	9.99	9.99	9.99
2	2	57	-.45	-.36	-.57	-.43
2	1	28	-.24	-.17	-.24	-.17
2	2	30	.25	.28	.25	.28
2	1	63	.54	.56	1.15	1.15
2	2	58	-1.17	-1.20	-1.71	-1.76
2	1	32	-.23	-.24	-.23	-.24
2	1	29	-.08	.01	-.08	.01
2	1	21	.71	.66	.71	.66
2	2	34	.25	.29	.25	.29
2	1	0	9.99	9.99	9.99	9.99
2	2	0	9.99	9.99	9.99	9.99

After selecting options, click “Import”. In the "Open file" dialog chose the file for import.

If the imported file is the fixed width field formatted Space-delimited text, the "Column Adjustment" dialog appears. You must test and adjust the width of columns stated by the program. Put the mouse pointer to the header of the column. Near the column delimiter the form of the mouse pointer will change, now move the boundary of the column by dragging. The additional column boundary can be dragged from the right boundary of the most right column.

When all columns have the correct width, click OK.

After the contents of the opened file appears in the table in the right side of the document window, you must specify the column assignment by selecting the appropriate item in the "Assigned for" combo box. Each column should be one of the obligatory fields or be a trait.


Untitled

COLUMNS

Column number: 1

Assigned for: Pedigree

Column name: EFAMT

Missing Value:

Trait Type: Integer

Miss Column ☐

First String for Import: 2

Take Names from the String: 1

Correct properties for each column and press OK for import

No.	EFAMT	FATH	MOTH	PROB	SEX	AGE	STATU...	WEIGH
1	EFAMT	FATH	MOTH	PROB	SEX	AGE	STATURE	WEIGHT
2	138	0	0	2	1	55	1.608	61.4
3	138	0	0	2	2	51	1.505	73.9
4	138	1	2	2	1	29	1.688	59.3
5	138	1	2	2	2	27	1.509	42.5
6	139	0	0	2	1	62	1.564	41.1
7	139	0	0	2	2	60	1.558	71.9
8	139	1	2	2	1	37	1.612	58.7
9	139	1	2	2	2	23	1.528	46.7
10	139	1	2	2	2	32	1.604	70.3
11	139	1	2	2	1	35	1.686	71.4
12	140	0	0	2	1	54	1.717	64.3
13	140	0	0	2	2	52	1.559	63.1
14	140	1	2	2	2	24	1.562	46.3
15	140	1	2	2	1	27	1.752	60.3
16	141	0	0	2	1	58	1.566	55.7
17	141	0	0	2	2	47	1.587	60.7

If the column is assigned for the "Trait", the column name can be changed in the "Column name" text box. If some special string code denotes the missing value, it should be specified in the "Missing Value" text box (the initial and ending spaces of the string are *ignored, by default the empty string denotes the missing value). Choose the correct item in the "Trait Type" combo box.*

Some columns can be missed during the input. If you don't need them, check the "Miss Column" check box. The "Miss Column" option can be also applied to multiple columns. The column, which properties are specified, is chosen by "Column number" field or by click on the header of the desirable column in the table. To select multiple columns click on the header of the first column, then pressing the <Shift> key click on the header of the last column.

Some strings on the beginning of the file can be ignored (if them contain some additional service information), to do this specify "First String for Import".

If some string of your file contains the names of the trait columns, specify the "Take Names from the String" field. All columns will be renamed together.

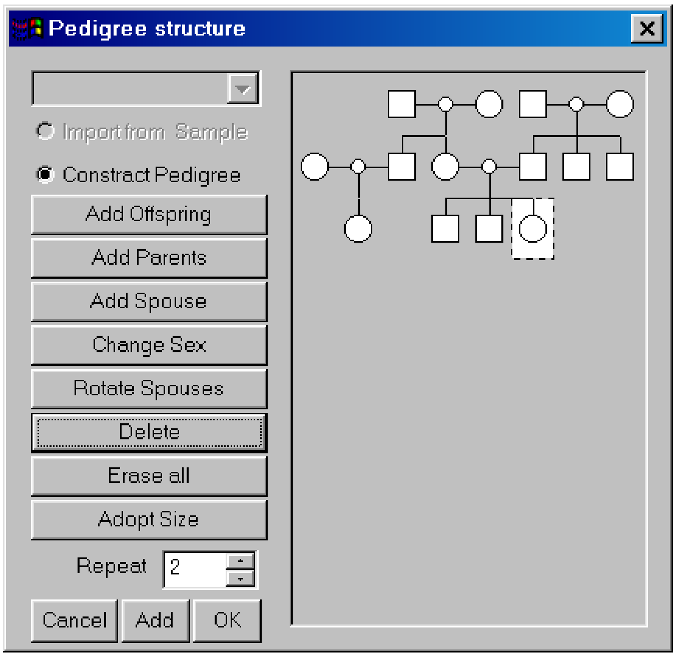
Click OK to transform data to the document form. If there are some errors or uncertainties in the data, you will get an error message. If the data are OK and you have chosen error-log, the Excel Error-log file will be opened and can be saved, if desirable.

1.1.2 Manual input.

You can construct pedigrees visually and add them sequentially to the sample.

If the not empty document is opened, click on the "New" option in the "Sample" menu. Click the menu option "Serial" - "Pedigrees" - "Construct". The "Pedigree structure" dialog appears with the minimal pedigree consisting of two parents and one offspring. Males are showed as squares and females as circles. The "marriage" is shown as small circle. Each pedigree member or "marriage" (center of nuclear pedigree) can be selected by clicking on it with left mouse button. The selected item appears on the background of white rectangular with dashed boundaries.

If the "marriage" is selected the "Add Offspring" button can be used to add offspring for nuclear family, you can do the same by double clicking on the "marriage" with left mouse button. The "Rotate Spouse" button can be used for better pedigree arrangement, the same can be done by right click on the "marriage". The "marriage" can be deleted if there is only one not married offspring and one of the spouses is the founder (has no parents in the pedigree) or if offspring is married, but two spouses are founders.



○ Import from Sample

● Constract Pedigree

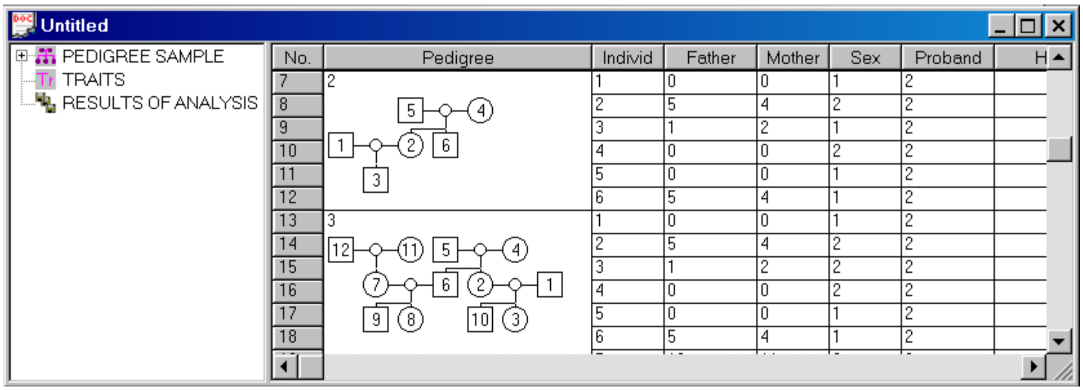
Repeat 2

If the pedigree member is chosen, and he/she is a founder (has not parents in the pedigree), the "Add Parents" button can be used, the same can be done by right click on the founder. If the pedigree member has no spouse in the pedigree (is child and not a parent) the "Change Sex" button can be used, the same can be done by right click on the child. The "Add Spouse" button can also be used. The selected member gets the new spouse and one offspring (the new "marriage"); double clicking on he/she with left mouse button can do the same.

The pedigree member can be deleted if he/she has no spouse and is not the single offspring in the pedigree.

The "Erase all" button return the drawn pedigree to the initial three-member form. The "Adopt size" button changes the picture size corresponding to window size.

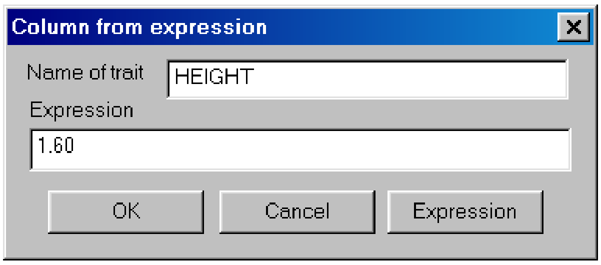
When the pedigree structure is ready, the multiplication coefficient is to be specified in the "Repeat" field. The click on the "Add" button appends the specified number of equal structure pedigrees to the sample. When all pedigrees are appended, click OK to finish the work and to see the result in the document window.



No.	Pedigree	Individ	Father	Mother	Sex	Proband	H
7	2	1	0	0	1	2	
8		2	5	4	2	2	
9		3	1	2	1	2	
10		4	0	0	2	2	
11		5	0	0	1	2	
12		6	5	4	1	2	
13	3	1	0	0	1	2	
14		2	5	4	2	2	
15		3	1	2	2	2	
16		4	0	0	2	2	
17		5	0	0	1	2	
18		6	5	4	1	2	

Having the pedigree structure, you can add some traits columns.

Click on the "Edit"-"Add Trait"-"Formula" menu item. The "Column from expression" dialog appears. Specify the "Name of trait" field. In the expression field specify the most frequent trait value. Click OK to add the new trait column. The new trait column appears in the document window.



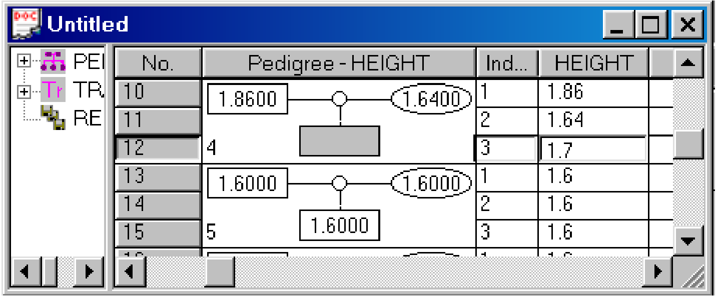
Column from expression

Name of trait: HEIGHT

Expression: 1.60

Trait value for each individual can be edited manually.

All individuals have initially equal trait values. To change the trait value for the desired individual, click on the appropriate cell of the trait column and edit the value as in the common textbox (to specify missing value clear the content of the cell). Then click the <Enter> key to save changes. Living the cell without saving (clicking the <Enter> key) cancels all changes. To see trait values in the pedigree structure scheme, click on the header of the trait column.



No.	Pedigree - HEIGHT	Ind...	HEIGHT
10		1	1.86
11		2	1.64
12	4	3	1.7
13		1	1.6
14		2	1.6
15	5	3	1.6

1.2 Drawing of the pedigree structure.

1.2.1 The scheme of each pedigree.

Drawing of each included pedigree structure is performed by default in the "Pedigree" column of document window table. By default the individual code is shown in the scheme. When some trait is selected, its values are shown for individuals. The scheme of selected pedigree can be introduced into MS Office documents using clipboard.

When the pedigree is selected the text color of all it's strings (individuals) changes. To select the pedigree double click on any of its strings. To put the scheme of the pedigree into the clipboard click "Edit"-"Copy" menu item.

1.2.2 Sorting pedigrees by their structure.

The pedigree structure of the whole sample can be drawn on the single figure. Pedigrees are sorted according to number of generations and pedigree size.

To draw the figure, click on the "Pedigree structure" item in the left side of the document window. Then click "Sample"-"Print preview" menu item to view or print the figure or "Edit"-"Copy" to put the figure into the clipboard.

1.3 Checking untypical observations

To check untypical (probably erroneous) measurements the limitation rule must be formulated. For this purpose some integral data (trait statistics) or trait transformations can be used. These options are available in the trait pop-up menu for quantitative trait columns. "Properties" item shows the dialog containing trait statistics, means and variances can be used to formulate the limitation rule. Some other options build auxiliary trait columns: "Standardize" item - the column containing the standardized version of the original trait; "Pedigree Mean" - the column containing mean trait value in the pedigree; "Pedigree Min" - the column containing minimal trait value in the pedigree; "Pedigree Max" - the column containing maximal trait value in the pedigree; "Pedigree Diff" - the column containing the mean modulus of the trait value difference between the individual and all his/she's first order relatives; "Pedigree DMax" - the column containing the maximal modulus of the trait value difference between individual and all his/she's first order relatives.

EXAMPLE.

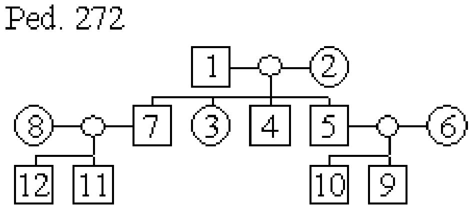
Marking pedigree members, which cause trait differences in first order relatives greater than 3.5 standard deviations for the quantitative trait "Tr_1".

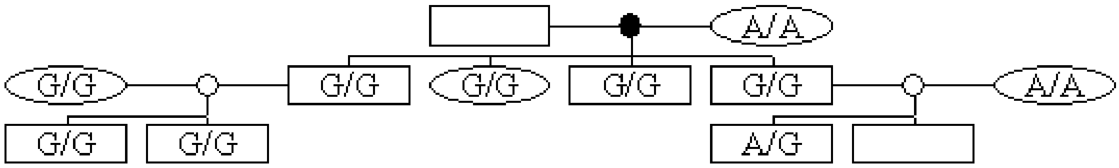
1) *Right click on the trait column, choose "Standardize" from the pop-up menu and create new trait "Tr1_S" containing standardized version of "Tr_1".*
2) *Right click on the column "Tr1_S", choose the "Pedigree Diff" option. The new trait column "Tr1_S_D" will be created.*
3) *Right click on the column "Tr1_S", choose the "Pedigree DMax" option. The new trait column "Tr1_S_DM" will appear.*
4) *Right click on the column "Tr1_S_D" and choose the "Pedigree Max" option. The new trait column "Tr1_S_D_M" will appear.*
5) *Create a new trait column using the formula (menu "Edit"-"Add trait"-"Formula"):*
   " if((Tr1_S_DM > 3.5) & (Tr1_S_D_M - Tr1_S_D<1.E-6), 1, 0) ".
   In this new column all desirable individuals will be marked with "1".
6) *To find the marked individuals click on the new trait column by right mouse button and chose "Adjustment" in the pop-up menu. In the dialog in the Covariate-combobox choose "Tr_1" and click OK. The scatter plot window will be opened. The points, corresponding to marked individuals, are in the upper part of the plot. Click on one of these points. The square cursor appears and surrounds the selected point. Now click on the "Edit"-"Find" menu item. In the document window the pedigree, which includes the selected individual, will be shown highlighted with red font and the string, corresponding to the individual appears sunken. Now you can see and edit all individual traits.*
7) *To see in the scatter-plot the pedigree, to which the individual (selected by square cursor) belong, click "Edit"-"Mark pedigree" menu item. All members of the pedigree will be surrounded with black circles.*

1.4 Checking violation from Mendelian rules for DNA marker data.

The option "Test Errors" is available in the column pop-up menu of marker trait.

Right click on the marker trait column and choose the "Test Errors" option. If some pedigrees contain violation from Mendelian rules, you will see the count of conflicting nuclear families. You can see the conflicting data in two ways: as a table (tab-delimited text) and/or graphically. If you use the first option the error table is to be seen in “Marker errors” dialog.. To copy the table into clipboard, select it in the editbox and press <Ctrl-C>. For the second option in the print preview window only pedigrees having conflicts will be drawn. The nuclear pedigrees, which have caused the contradiction, will be marked with black circle.





1.5 Finding optimal adjustment function for quantitative trait.

As a rule quantitative traits are influenced by a great number of factors (covariates). The best fitting technique (maximum likelihood parameter estimation) enables to build the population model of the trait including the covariate dependence as a function of certain type, for which only a set of coefficients is to be determined. The available function types are as follows.

For polynomial functions the whole interval of available covariate data can be divided into a number *k* (up to 10) of subintervals. For each subinterval its boundaries $\{L_0, L_1, \ldots, L_k\}$ and the power of polynomial function$\{N_1, N_2, \ldots, N_k\}$ can be determined arbitrary. For each interval $Y(X) = Y_i(X) = \sum_{n=0}^{N_i} a_{in}(X - L_{i-1})^n$, when $L_{i-1}<X<L_i$. The function values on the interval boundaries ($X=L_i$) hold equal for the left and right intervals: $Y_i(L_i) = Y_{i+1}(L_i)$. So the piece-wise polynomial continuous function occurs. The matrix of polynomial coefficients $\{a_{in}\}$(for each interval and each power) is computed using least mean square method (the maximum likelihood estimation, which suppose the normal distribution of trait around its mean, which depends on covariate).

There is also a possibility to include the boundaries of one interval (L_{i-1}, L_i) as parameters in the maximum likelihood estimation procedure. The variants of this option are two and three interval functions referred in the **Approximation-Empirical** submenu. They are: constant-linear, linear- constant and linear-constant-linear piece-wise polynomial functions. Often they are suitable as simple aging models for example for traits characterizing bone development (as fully described in [Malkin et al., 2002]). The biological rationale of these models concurs with the three recognized stages of bone ontogenesis, namely, bone acquisition in youth, stabilization at maturity and loss of bone tissue at senescence.

For one interval including the whole range of covariate additional special functions are available (denotation corresponds to **Approximation** menu items):

exponential: $Y(X)=C_0+e^{a_0+a_1(X-L_0)}$;

logarithmic: $Y(X)=\ln(a_0+a_1(X-L_0))$;

logistic: $Y(X)=a_2+\frac{a_3}{1+e^{a_0+a_1(X-L_0)}}$;

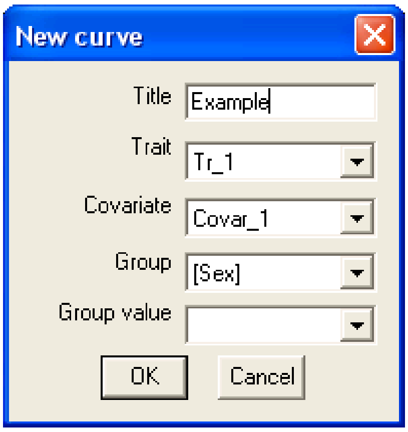
normal: $Y(X)=C_0+e^{a_0+a_1(X-L_0)+a_2(X-L_0)^2}$;

stochastic $Y(X)=B(X-t_m)+\frac{B}{q}(1-q)^{(X-t_0)}$, where $q=(t_m-t_0)^{-1}$;

growth curve $Y(X)=A+BX+C\ln(X+1)$, children growth curve for height and weight.

To compare various adjustment curves the analog of multiple regression coefficient *R*, is calculated for each curve (R^2 is the proportion of trait variance explained by the model to the whole trait variance). The log-likelihood value is also available for each adjustment function. Let us define two different piece-wise polynomial functions in such a way, that the function with lesser number of estimated parameters coincides with the other, having some parameters constrained to constant values. Then two functions can be compared using likelihood ratio test in order to choose for the investigated trait the best fitting most parsimonious one.

EXAMPLE 1.



New curve

Title: Example

Trait: Tr_1

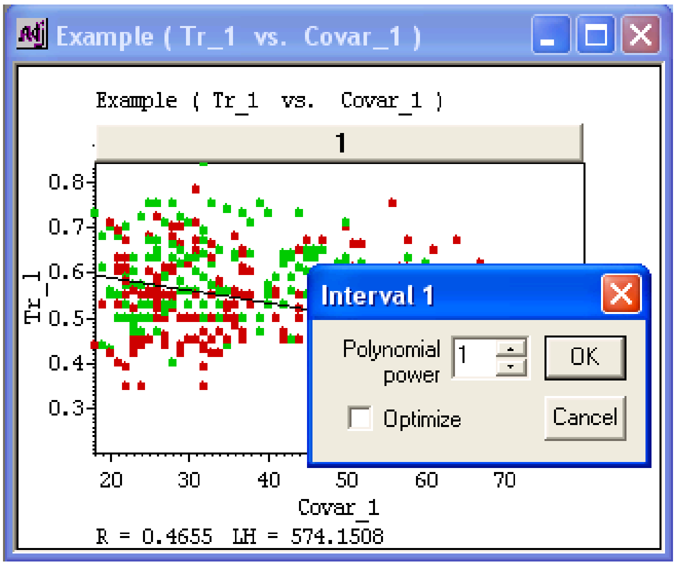
Covariate: Covar_1

Group: [Sex]

Group value:

Choosing the best fitting most parsimonious polynomial function for the quantitative trait "Tr_1" and covariate "Covar_1".

1) Right click on the trait column and choose "Adjustment" from the pop-up menu. The "New curve" dialog appears. In the "Trait" combo box the trait "Tr_1" is already selected (You also can open the "New curve" dialog from "View"-"Curve"-"New" menu item. In this case you should select the desirable trait "Tr_1" in the "Trait" combo box, where all quantitative traits, presented in the sample, are available). In the "Covariate" combo box select "Covar_1" from the list of quantitative traits. If you want to build the adjustment only for some specific group determined by separate group value in the integer or categorical (text) trait "Group_tr", specify "Group_tr" in the "Group" combo box and the desired value in "Group value" combo box. By default the "Group" is sex. If the "Group value" is not defined, both possible group values 1(male) and 2(female) will be displayed in the scatterplot. The field "Title" is optional. Click on the OK button. Now your active window is the scatter plot. The red circles correspond to sample members with group value 1(by default males) and green circles to all others.



Polynomial power 1

☐ Optimize

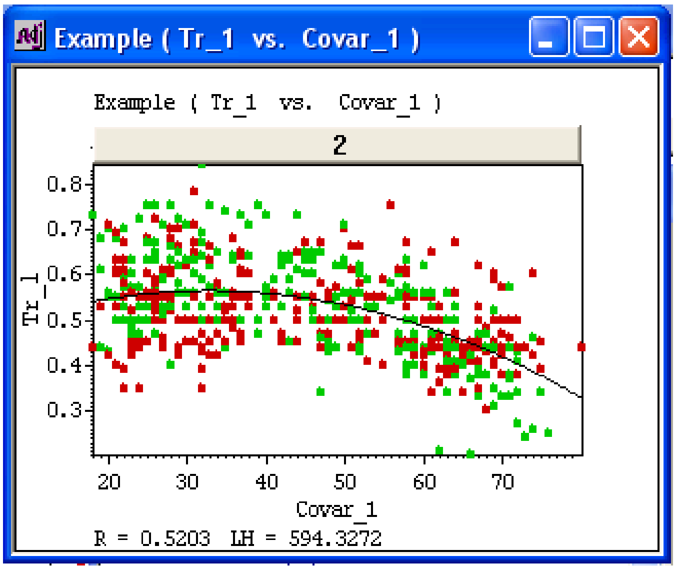
OK

Cancel

R = 0.4655 LH = 574.1508

2) Select in the "Approximation" menu the desirable function type. For this example select "Approximation"- "Polynomial" menu item. The linear function appears. In the lower part of the window R and log-likelihood value (LH = ln[probability density])) are available. Upper to the graph appears the interval header with polynomial power (1) on it.

3) To change the power of polynomial click on the interval header, the Interval dialog appears. Change the polynomial power to 2 and click OK. The new power appears on the interval header. The form of the curve changes together with R and LH values. The double difference in LH value should be distributed as χ^2distribution with 1 degree of freedom. We can see, that the quadratic polynomial function is significantly better than the linear one (LH[1]=573.1508, LH[2]=594.3272). Changing the polynomial power to 3 we find, that the likelihood ratio test for polynomials with power 3 (LH=594,7179) and 2 is not significant. So the second power polynomial is the best fitting most parsimonious model among the family of one-interval polynomial functions.



R = 0.5203 LH = 594.3272

NOTE:We have not done here comparison between polynomials with power 0 (constant) and power 1, but it should also be done if the quadratic polynomial is not significantly better than linear.

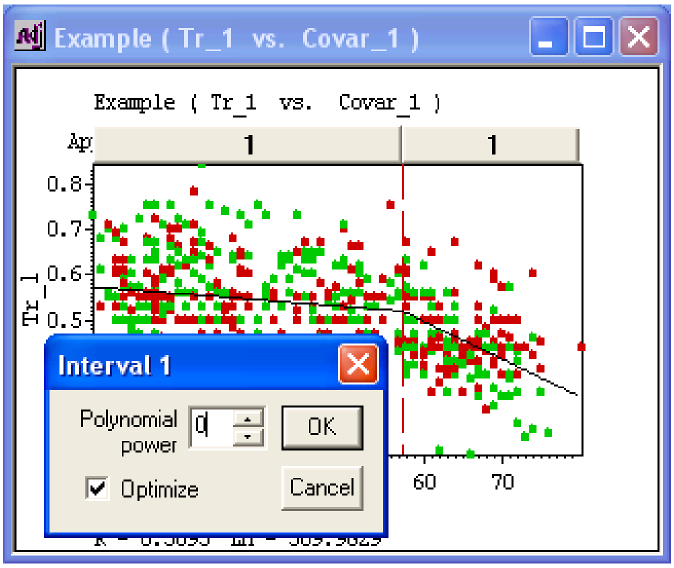
EXAMPLE 2.

Choosing the best fitting most parsimonious piece-wise linear two interval function for the quantitative trait "Tr_1" and covariate "Covar_1".

Execute steps 1) and 2)of the previous example.

3) Move the cursor to the right edge of the interval header. After the form of mouse cursor changes, drug from the right edge additional interval boundary. Now we have two interval headers and the curve is piece-wise linear.

4) Click on the left interval header. In the interval dialog change the polynomial power to 0 and check the "Optimize" check box.



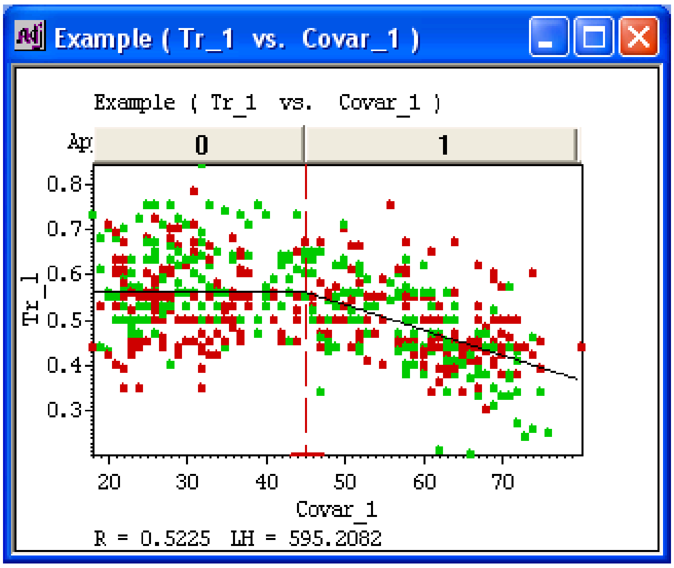
Polynomial power 0

☑ Optimize

OK

Cancel

Then click OK. The boundary between intervals moves to its optimal value for two interval constant-linear function. The read interval on the X-axis defines the error value of the optimal boundary position. The initial linear function can be assumed to be also two-interval with *first interval having zero length. So the likelihood ratio test for linear (LH = 573.1508) and constant-linear (LH = 595.2082) models should be distributed as χ^2 with 1 degree of freedom. And we see that the difference is significant. For the next comparison click on the left interval header, change the polynomial power to 1 and click optimize. The likelihood of linear-linear function (LH=595.4248) is not significantly better than of constant linear one. So the constant-linear function is best fitting most parsimonious one in the family of two-interval piece-wise linear functions.*



R = 0.5225 LH = 595.2082

1.5.1 Expression of adjustment curve, parameter values and errors.

When the scatter plot window contains adjustment curve and is the active window, the general expression of the curve and the parameter values can be displayed by clicking on the "View"-"Curve"-"Info" menu item. In the edit box the parameter values are presented as tab-delimited text. To copy the parameters values into the clipboard, select the contents of the edit box by the mouse and click <Ctrl-C>.

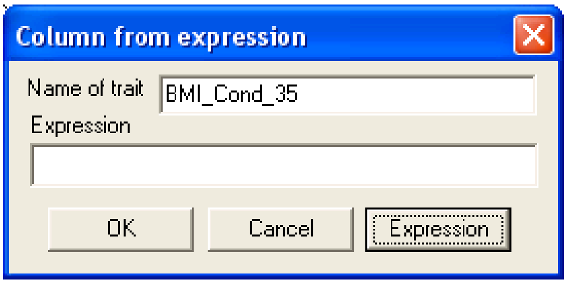
1.5.2 Saving the adjustment curve.

When the scatter plot window contains adjustment curve and is the active window, click the "Edit"-"Add curve" menu item. The curve icon appears in the tree in the left side of the document window (root item "Curves"). You need to save curve, if you plane to create new trait, adjusted in accordance with the current curve. Only curves included in the "Curves" item of the tree can be used to create and edit adjusted traits.

1.6 Creating new derived traits.

1.6.1 Creating a trait by formula.

The new trait can be created as mathematical functions on any combination of already existing traits, having discrete and continuous range. The interface support simple mathematical operators +,-,*,/, comparison operators =,<,<=,>,>=, logical operators **&** (AND), **|** (OR), **NOT(**<expression>**)**, conditional function **if(**<logical expression>,<expression if TRUE(1)>,<expression if FALSE(0)>**)**, information function **IsMis**(<expression>), which return 1(TRUE) if the expression has missing value, and a few of mathematical functions **int(**<expression>**)**, **abs(**<expression>**)**, **sqrt(**<expression>**)**, **exp(**<expression>**)**, **ln(**<expression>**)**. Additionally there are two function **F(**<expression>**), M(**<expression>**)**, and **S(**<expression>**)**, which return the expression value of individuals father, mother and sibling (nearest next) correspondingly. Expression can be any combination of operators, functions, digital values and trait names, which are presented in the sample. Constant string values should appear in the expression in quotation marks. The + operator is also used for



Name of trait: BMI_Cond_35

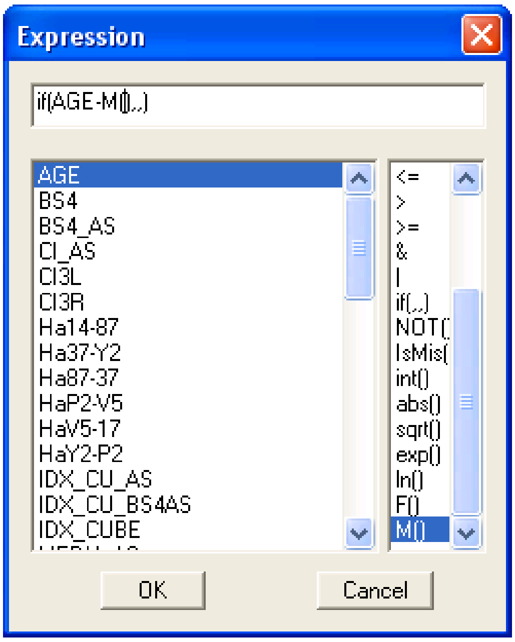
Expression:

concatenation; if at least one of the operands is the string, the other is also converted to string. So to convert a numerical trait to the text trait you can use the following: <numerical trait>+"". The "internal" variables [Pedigree], [Individual], [Father], [Mother] can also be used and are proceeded as text values. Variables [Proband] and [Sex] have integer values ([Sex]=1 for males and 2 for females). [MissVal] should be used when some of individuals don't get values if some condition is not fulfilled. Don't use [MissVal] in the comparison operation, use the **IsMis**(<expression>) function instead. Round brackets can be used to specify the order of operations.

EXAMPLE.

Building the BMI (Body Mass Index) trait only for individuals, which were born from the mothers older than 35 years.

1) *Click the menu item "Edit"-"Add trait"-"Formula". The Column-from-expression dialog appears. In the Name-of-trait edit box type the name of new trait: **BMI_Cond_35**. The expression in the Expression edit box can be formed directly using keyboard or in the special expression wizard.*
2) *Click the "Expression" button to open the expression wizard window. The upper edit box is assigned for the expression string. The left list box contains the list of all trait names of the sample and internal variables (enclosed in square brackets). The right list box contains the list of possible operators and functions. Double click on the list string place the appropriate variable, operator or function in the expression window to the cursor position.*



if(AGE-M(),,)

3) *For our trait we should form the conditional expression. Double click on the **if(,,)** function. The string* "if(,,)" *appears in the upper edit box. The cursor position is placed before the first comma to get the condition expression. In this example the condition demands that the difference between the mother and offspring age should be greater than 35, in functional form: AGE-M(AGE)>35. To build it, double click on the AGE trait in the left list box. Now the expression in the edit box should be* "if(AGE,,)" *with cursor position after "AGE". Double click on the minus operator, then on the **M()** function, then once more on the "AGE". Now the expression is* "if(AGE-M(AGE),,)". *Move the cursor position after the first comma and form the TRUE-case expression:* WEIGHT/(STATURE* STATURE), *then move the cursor position after the second comma and form the FALSE-case expression:* [MissVal]. *The final expression should be as follows:*
"if(AGE-M(AGE)>35, WEIGHT/(STATURE* STATURE), [MissVal])"
4) *Click OK. If there are some errors in the expression, you will get an error message, otherwise the wizard dialog closes and the expression appears in the edit box of the Column-from-expression dialog.*
5) *Click OK. If there are no errors in the expression, the dialog closes and the new trait appears as the last column in the data table (right part of the sample window.)*

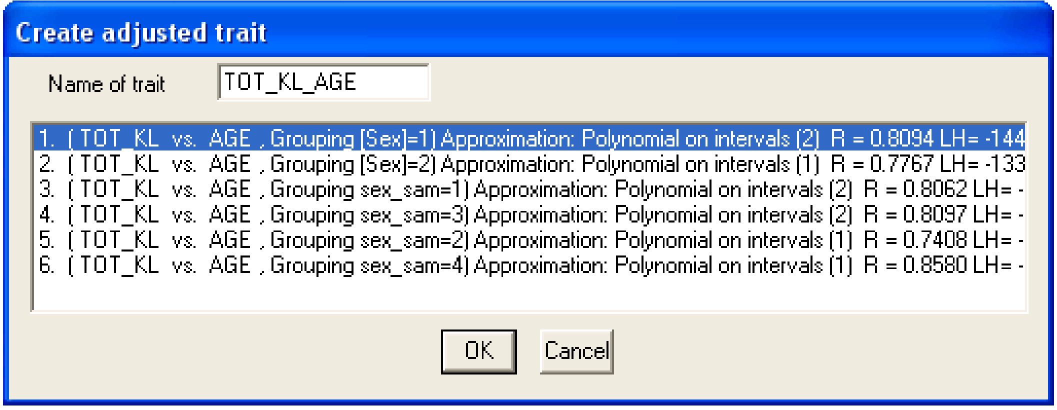
1.6.2 Creating a trait adjusted for the covariate.

You can create a new adjusted trait as a difference between the initial trait value of each individual and its predicted value, based on the adjustment curve for selected covariate. The procedure permits grouping for categorical traits - sex, age group, location and so on. The appropriate adjustment curve should be created and saved (see 1.5, 1.5.2). If for different groups you need different adjustment (for example in accordance to [sex]=1 and [sex]=2), save the adjustment curve for each group value.

Example.

Building the trait adjusted on age separately for males and females.

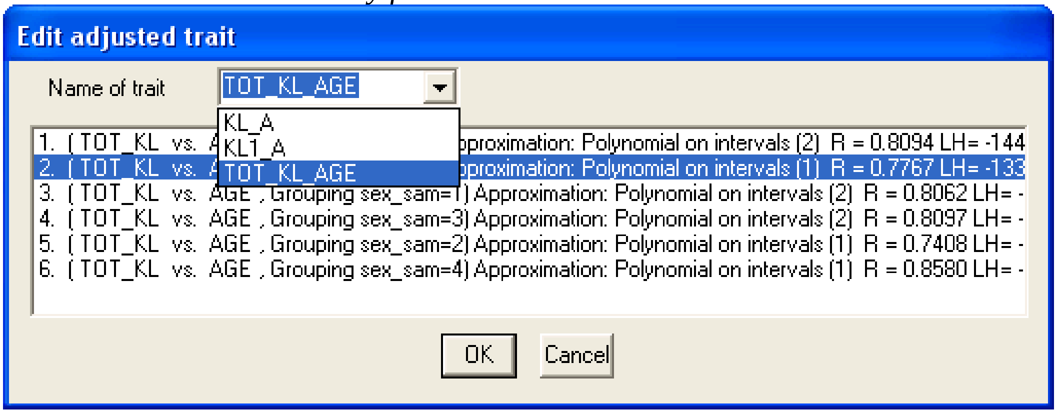
1) Create and save adjustment curves for the trait ***TOT_KL*** *vs. AGE for group values [sex]=1 and [sex]=2* (see 1.5, 1.5.2)*.*



Name of trait: TOT_KL_AGE

1. (TOT_KL vs. AGE , Grouping [Sex]=1) Approximation: Polynomial on intervals (2) R = 0.8094 LH= -144
2. (TOT_KL vs. AGE , Grouping [Sex]=2) Approximation: Polynomial on intervals (1) R = 0.7767 LH= -133
3. (TOT_KL vs. AGE , Grouping sex_sam=1) Approximation: Polynomial on intervals (2) R = 0.8062 LH= -
4. (TOT_KL vs. AGE , Grouping sex_sam=3) Approximation: Polynomial on intervals (2) R = 0.8097 LH= -
5. (TOT_KL vs. AGE , Grouping sex_sam=2) Approximation: Polynomial on intervals (1) R = 0.7408 LH= -
6. (TOT_KL vs. AGE , Grouping sex_sam=4) Approximation: Polynomial on intervals (1) R = 0.8580 LH= -

2) Click the menu item "Edit"-"Add trait"-"Adjusted". The Create-adjusted-trait-dialog appears. In the Name-of-trait edit box type the name of new trait: ***TOT_KL_AGE****. In the list box below you see the list of all adjustment curves, select the appropriate curve for [sex]=1 and click <OK>. The new trait* ***TOT_KL_AGE*** *appears as the last column in the data table. This trait contains values only for males.*



Name of trait: TOT_KL_AGE

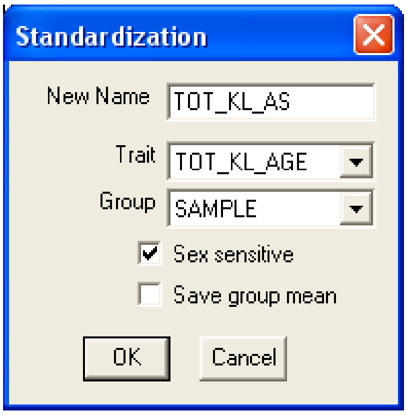
1. (TOT_KL vs. A pproximation: Polynomial on intervals (2) R = 0.8094 LH= -144
2. (TOT_KL vs. A pproximation: Polynomial on intervals (1) R = 0.7767 LH= -133
3. (TOT_KL vs. AGE , Grouping sex_sam=1) Approximation: Polynomial on intervals (2) R = 0.8062 LH= -
4. (TOT_KL vs. AGE , Grouping sex_sam=3) Approximation: Polynomial on intervals (2) R = 0.8097 LH= -
5. (TOT_KL vs. AGE , Grouping sex_sam=2) Approximation: Polynomial on intervals (1) R = 0.7408 LH= -
6. (TOT_KL vs. AGE , Grouping sex_sam=4) Approximation: Polynomial on intervals (1) R = 0.8580 LH= -

3) Click the menu item "Edit"-"Edit trait"-"Adjusted". The Edit-adjusted-trait-dialog appears. Select in the Name-of-trait combo box the name of trait, which was created on the previous step: ***TOT_KL_AGE****. In the list box below select the adjustment curve appropriate to female ([sex]=2) and click <OK>. Now trait contains values for both sexes.*

1.6.3 Standardized trait.

The standardized trait is a trait with zero mean and variance equal to 1. Such traits are often used for statistical applications. The standardized trait can be produced on the base of each quantitative trait. The procedure permits grouping for categorical traits - sex, age group and so on.

Click right mouse button on the column of trait, which should be transformed. In the dropdown menu select “Standardize”. The standardization dialog appears. You also can view this dialog by clicking “Edit”-“Add trait”-“Standardized” menu. If the name of initial trait for transformation is not selected in the ***Trait*** *combo box, select it. Type in the “New Name”-editbox the name for the new standardized trait. If the standardization should be made for each group value of some categorical trait, select it in the* ***Group*** *combo box. If the transformation should be made for each sex separately, check the* ***Sex-sensitive*** *checkbox. If you want to save group means and to turn only the overall mean to zero check* ***Save-group-mean*** *checkbox.*



New Name: TOT_KL_AS

Trait: TOT_KL_AGE

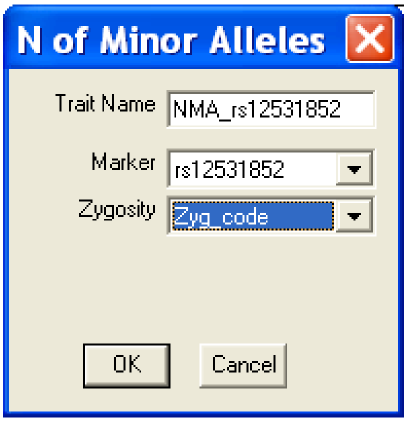
Group: SAMPLE

☑ Sex sensitive

□ Save group mean

1.6.4 Creating an integer trait corresponding to number of minor alleles in selected diallelic marker (SNP).

The marker trait contains data of chromosomal marker genotype. Each trait value should contain a pair of alleles. For diallelic markers there are a number of analyses. which use in some way regression of a quantitative trait on the number of definite alleles in individual genotype. To use them you can produce an integer trait containing the number of minor alleles in individual marker genotype. If the sample contains twin pairs, for MZ twins as a rule only one of the twins is genotyped. If you have an integer trait which denote twin sets in the pedigree with the same value, (positive for MZ and negative for DZ twins), it can be used to give to the MZ twin with missing marker genotype the number of minor alleles of the genotyped co-twin. So the power of analysis can be increased.



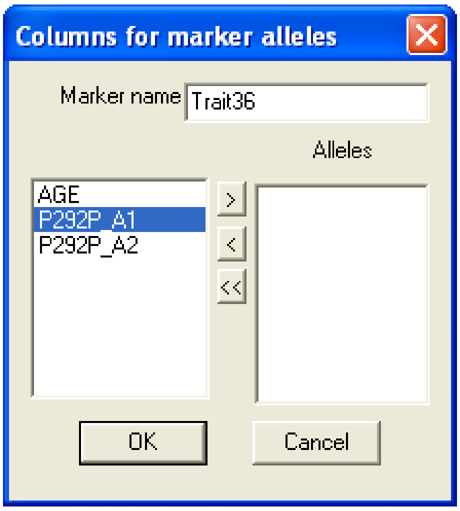
Trait Name: NMA_rs12531852

Marker: rs12531852

Zygosity: Zyg_code

Click the menu item “Edit”-“Add trait”-“ Number of Minor Allele”. The Number-of- Minor-Allele-dialog appears. The left listbox contains the list of all integer or categorical traits, which are available. Select in the "Marker" combobox the dialelic marker from the list. The name of new integer trait by default is NMA_<marker name>, but it can be changed in the "Trait Name" editbox. If the sample contains twin pairs, select integer trait defining zygosity in the appropriate combobox. Click “OK”. The new integer trait appears as the last column in the data table.

1.6.5 Creating a marker trait from two integer or categorical trait columns.

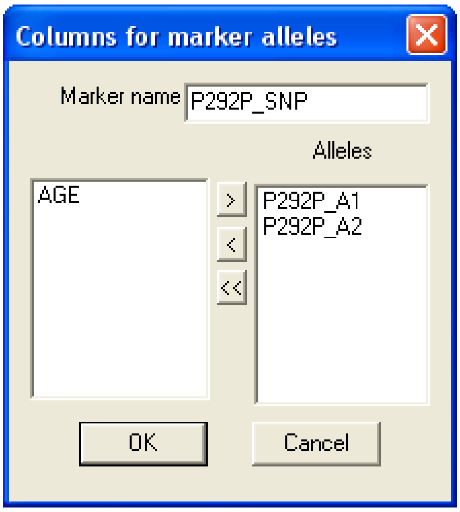


Marker name: Trait36

AGE
P292P_A1
P292P_A2

The marker trait contains data of chromosomal marker genotype. Each trait value should contain a pair of alleles. If in your input file alleles were included as separate columns, the program considers them as separate integer or categorical traits. To use the data, you should convert these traits into united marker trait.

Click the menu item "Edit"-"Add trait"-"Marker from Alleles". The Columns-for-marker-alleles-dialog appears. The left listbox contains the list of all integer or categorical traits, which are available. Select in the listbox the trait, corresponding to the first marker allele, and click the upper button marked with ">". The selected column name will be moved to the right list box named "Alleles". Analogously move the trait, corresponding to the second marker allele to the right list box. Type the name of the new marker in the "Marker Name"-editbox. Click "OK". The new marker appears as the last column in the data table.



Marker name: P292P_SNP

AGE

Alleles: P292P_A1, P292P_A2

1.6.6 Creating a new marker trait, containing factorized data of an existing marker.

If the marker contains a great number of alleles, some of them having very low frequency, you can factorize the marker. The factorization procedure consider selected group of different marker alleles as one allele of the mew marker. In general, you can design up to 5 arbitrary groups of alleles to produce new marker having up to 5 alleles. Let the number of alleles for new marker was selected to be N. There are two automated procedures to distribute the old marker alleles through the new marker alleles.

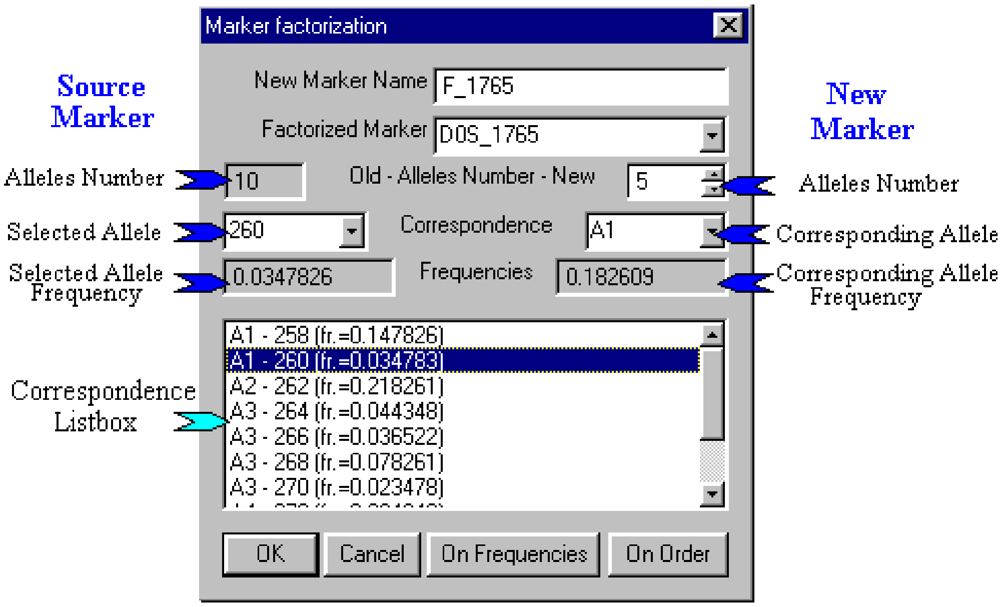
First variant - according to frequency.

The 1 allele of the new marker coincides with the most frequent allele of the old markers. The next allele is chosen as most frequent from the rest and so on up to N-1. The N-th allele united all alleles from the old marker, which were not chosen to be separate alleles (rare alleles).

Second variant - according to order.

The procedure ordered all allele names of the old marker alphabetically. Then N allele groups are designed from the closest alleles, to form a new marker with as equal as possible allele frequencies. This can be useful for the STR markers factorization. STR marker alleles differ from each other with the number of repetitions of small DNA fragments. Names of alleles correspond as a rule to the number of repetition or to the weight of the DNA segment proportional to the number of repetitions.

It is also possible to construct groups of alleles in an arbitrary way.



New Marker Name: F_1765

Factorized Marker: DOS_1765

Old - Alleles Number - New: 10 | 5

Correspondence: 260 | A1

Frequencies: 0.0347826 | 0.182609

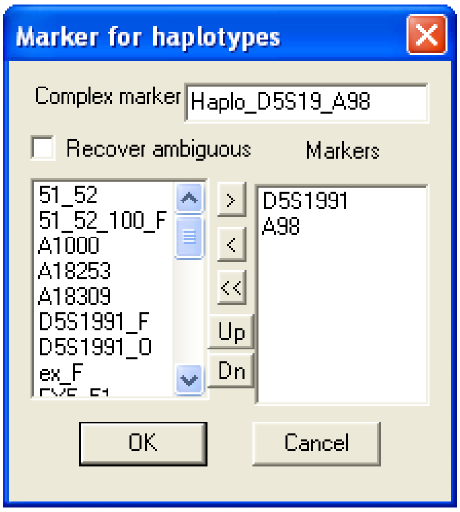
```
A1 - 258 (fr.=0.147826)
A1 - 260 (fr.=0.034783)
A2 - 262 (fr.=0.218261)
A3 - 264 (fr.=0.044348)
A3 - 266 (fr.=0.036522)
A3 - 268 (fr.=0.078261)
A3 - 270 (fr.=0.023478)
```

Do right click on the marker column in the data table or on the marker name in the tree in the left side of the document window (root item "Traits"). In the pop-up menu select the item "Factorize". The Factorization-dialog appears. Type in the "New Marker Name" editbox the corresponding new name. The field "Factorized Marker" contains the name of column with old marker data. There are two fields of "Alleles Number". The left value corresponds to the old marker and cannot be changed. In the right editbox the number of alleles of the new marker should be stated (2÷5). For each allele of the old marker define the allele of the new marker in which you want to include it. The frequency of allele appears under the appropriate allele ID. In the listbox in the bottom of the dialog you can see all correspondences together, sorted in accordance to new marker alleles. Double click on the string in this listbox provide the selected correspondence for editing: the old marker allele appears in the left combobox, the new marker allele in the right. Changing the allele selection in the right combobox you edit the correspondence. At the same moment, changes the list of correspondences in the bottom of the dialog. Clicking on the "On Frequencies" or "On Order" button, you get the correspondence list for two automatic variants described above. Then you can continue with editing, up to the correspondence will be appropriate. "OK" button creates the new marker with defined correspondence of alleles. In the properties of the new marker (Column pop-up menu, item "Properties") you can see how the marker was produced.

1.6.7 Creating a new marker trait, containing haplotype data restored from genotypes of two tightly linked markers.

The procedure reconstructs all possible haplotype combinations for a whole pedigree, in accordance to the pedigree structure. For this purpose it utilizes individuals, who have both unphased genotypes and are not double heterozygotes, and thus for whom the haplotype pair is known exactly. Because the marker pair is supposed to be tightly linked, we reconstruct haplotypes assuming minimal number of recombinations between initial markers, which can explain the existing joint genotype data for these two markers. There are two options available. 1) You can reconstruct only those individuals, whose haplotypes were uniquely identified with the assumption of minimal number of recombination events. In this case, the proportion of individuals having uncertainty of haplotype phasing, connected with “non-recombination” assumption, is less than a half of recombination probability. 2) You can reconstruct all individuals with both marker genotypes defined (ambiguous genotypes are chosen as most likely with haplotype frequencies observed in unambiguous sub-sample). The resulting marker can be considered as N*M -allelic locus with each allele corresponding to the respective haplotype variant (N, M are the numbers of alleles in initial markers).

*Click the menu item “Edit”-“Add trait”-“Haplotype Recover”. The "Markers for haplotype"-dialog appears. Type in the “Complex marker”-editbox the name of new marker. Define the pair of markers forming haplotypes as follows. Select from the list of all markers presented in the sample the first marker to include in the pair. Move it to the right listbox using “>”-button. Select the second marker and move it to the right. The order of selected markers defines the order of alleles in the haplotype denotation and can be changed with “Up” and “Dn” buttons. In a new marker, the haplotypes will be denoted as: <allele of the upper marker>|<allele of the lower marker>. If you enable ambiguous reconstruction of haplotypes, check the “Recover ambiguous”-checkbox. Click “OK”. The new marker appears in the data table as last trait column. This marker treated as N*M -allelic locus can contain Mendelian errors if a minimal possible number of recombination events is not zero.*



Marker for haplotypes

Complex marker: Haplo_D5S19_A98

□ Recover ambiguous

51_52
51_52_100_F
A1000
A18253
A18309
D5S1991_F
D5S1991_0
ex_F

Markers:
D5S1991
A98

1.6.8. Including traits from the external table.

If you want to include some additional traits in the existing sample, the information should be organized as a table in Excel, DBF or tab-delimited text format.

There are two variants of table organization:

1) Each column of the table corresponds to separate trait, each row to separate individual. The names of traits can be included in the first row of the table. To match the values of the included trait to the individuals included in the sample you can use pedigree and individual names (or numbers). Appropriate columns (pedigree, individual) should be included in the imported table. If in the sample there is a trait, which contains a unique text key, the same key can be included in the imported table to match the trait values to individuals.

2) This variant is available only if there is a unique text key in the sample. The genotyping data, got from the genotyping provider, are as usual organized as a table having 3 obligatory columns: Individual Key, Marker ID, Genotype,- and probably some other columns concerning genotyping quality and so on. After importing, each marker forms a separate trait.

For importing marker-traits in both variants, there are the same coding options and error- processing options as for sample import (see 1.1.1).

Click the menu item “Edit”-“Add trait”-“From Table”. The “Trait Import”- dialog appears. In the “Relation Key” combobox list you can see and choose all of the unique text keys, available in the sample. If there are no unique keys in the sample, only one option is available: “Sample key (ped,ind)”. If desirable, select options of marker errors processing and error-log saving, and then click “Import”. In the "Open file" dialog chose the file containing *traits for import. The further actions are similar to those described in (1.1.1). The only difference is the assignment of columns for obligatory fields. If you have chosen "Sample key (ped,ind)", you should assign columns for Pedigree and Individual. If your choice was some unique key, you should assign column for Key. If your table has the second type of organization, you should assign columns for MarkerID and Genotype. In this case all other columns, besides Key, MarkerID and Genotype, should be marked with "Miss Column" checkbox (dialog in the left side of the document window). Click "OK" to finish import and to view error-log, if you have chosen this option.*

1.6.8. Format of the import error-log.

If you selected the unique key for matching trait values from the importing table to individuals in the sample, the first worksheet in the log will be "Missing Values". It contain the table with the most left column being individuals key and upper row containing all imported trait names. If the crossing sell corresponding to individual A and trait B is equal to 1, this means that the appropriate trait value for individual A was missing in the imported table. The lowest row shows the number of missing individuals for each trait. The most right column- the number of missing traits for an individual. Only individuals having at least one missing trait value are presented in the table.

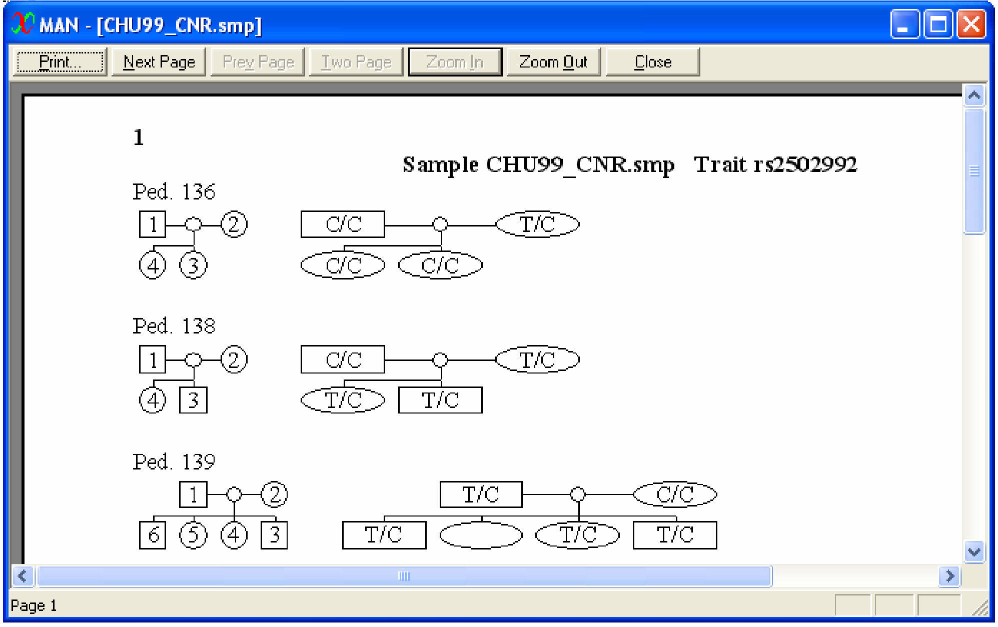
If you have chosen processing of marker traits, you get worksheet "Deleted Genotypes", containing analogous table, in which 1 denotes deleted genotype. Each marker, for which at least one Mendelian conflict was found, gets its own worksheet, where genotypes of conflicting nuclear families are presented.

1.7 Column menu (operations with a trait).

When you click with right mouse button on the trait column the pop-up menu appears. It contents depend on the type of trait.

1.7.1 Options shared by all trait types.

Delete- menu item demands confirmation of the operation. No *Undo*-option is available for this operation.

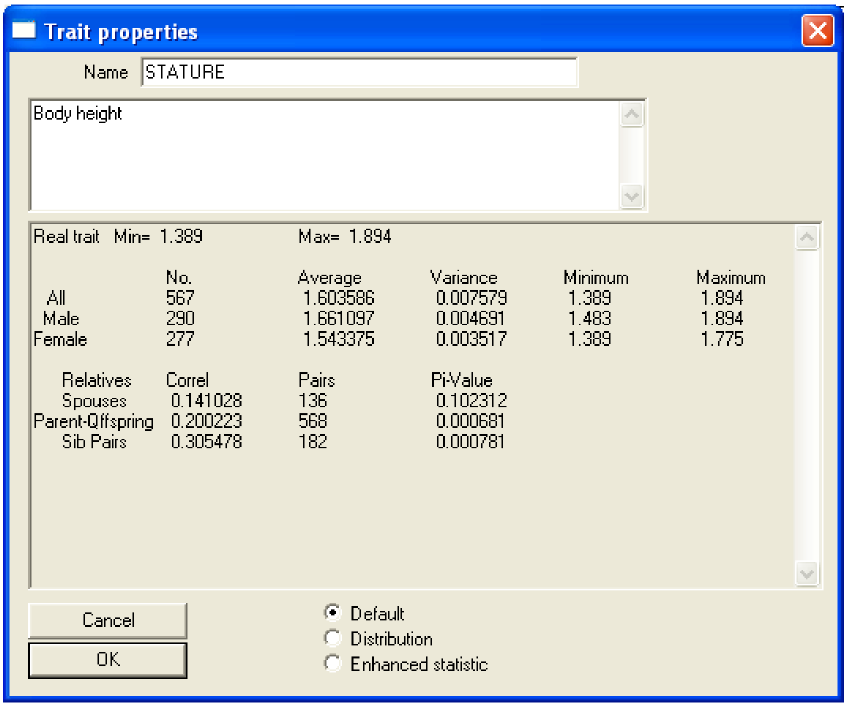


Print- shows the print preview window, where pedigrees are drown twice with individual codes and with trait values.

Properties-item displays a Property-dialog, which enables to edit trait name and notes. To save changes click OK. The Property-dialog shows also the information about trait creation history and trait statistics, corresponding to type of trait. The radio buttons in the upper part of the dialog determine three variants of information displayed.

1) *Default.*

Name and notes edit boxes are available for edition. For quantitative traits descriptive statistics (number of measured individuals, average, variance, minimum, maximum) is displayed for the whole sample and for each sex separately. Familial correlations for nuclear family member pairs are displayed with their significance. For markers and categorical traits, the statistics includes the number of measured individuals, number of phenotypes (categories) and phenotype frequencies. For genetic markers only you can see the number of pedigrees and nuclear families, having at least one measured individual; number of measured individuals being parents; number of heterozygous parent and their proportion in parent sample (heterozygosity); number of spouses pairs informative for linkage analysis (at least one of spouses should be heterozygous); number of alleles and their frequencies .

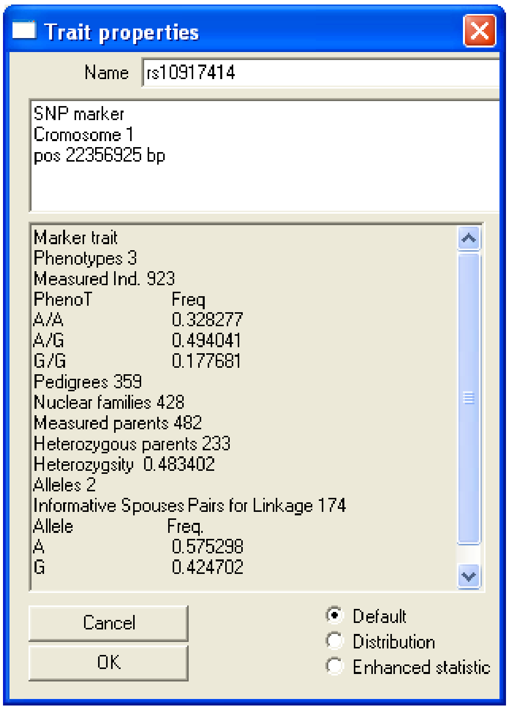


Name: STATURE

Body height

```
Real trait   Min=  1.389              Max=  1.894

                No.            Average      Variance     Minimum     Maximum
 All            567             1.603586     0.007579    1.389       1.894
 Male           290             1.661097     0.004691    1.483       1.894
Female          277             1.543375     0.003517    1.389       1.775

   Relatives    Correl         Pairs        Pi-Value
   Spouses       0.141028      136           0.102312
Parent-Qffspring 0.200223      568           0.000681
   Sib Pairs     0.305478      182           0.000781
```

● Default
○ Distribution
○ Enhanced statistic



Name: rs10917414

SNP marker
Cromosome 1
pos 22356925 bp

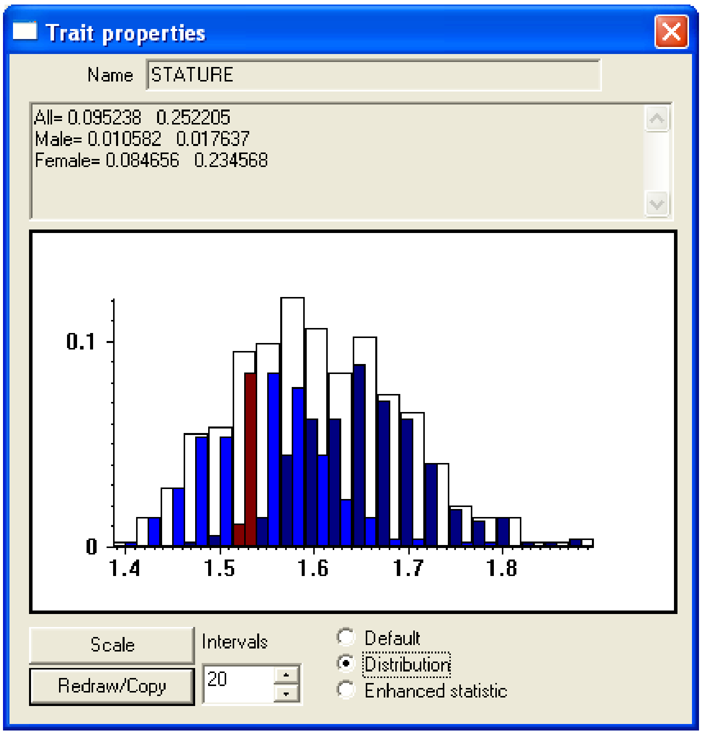
```
Marker trait
Phenotypes 3
Measured Ind. 923
PhenoT              Freq
A/A                 0.328277
A/G                 0.494041
G/G                 0.177681
Pedigrees 359
Nuclear families 428
Measured parents 482
Heterozygous parents 233
Heterozygsity  0.483402
Alleles 2
Informative Spouses Pairs for Linkage 174
Allele             Freq.
A                   0.575298
G                   0.424702
```

● Default
○ Distribution
○ Enhanced statistic

2) *Distribution.*

If you choose the appropriate radio button, the graph of the trait distribution appears. For markers the number of distribution intervals is fixed and correspond to number of phenotypes. For quantitative traits the number of distribution intervals is 20 by defaults, but can be changed. For this purpose change a number in Intervals-edit box and click on the"Copy/Redraw"-button. You can also edit the scale of vertical and horizontal axes by clicking "Scale"-button. In the displayed dialog edit appropriate scales and click OK.

To change the horizontal size of the graph, move the dialog right edge. The distribution of the whole sample is drawn in white, for males in dark-blue and for females in blue. Each time *when you click on the "Copy / Redraw"-button, the graph is copying into clipboard and you can use the "Paste"-option to put the graph into Microsoft Office applications. If you need to know numerically the proportion of individuals included in the distribution interval, click on it by left mouse button. The appropriate male and female pillars appear in red and in the upper edit box you will see the interval content proportion and accumulated proportion of individuals having trait value less then right interval edge*



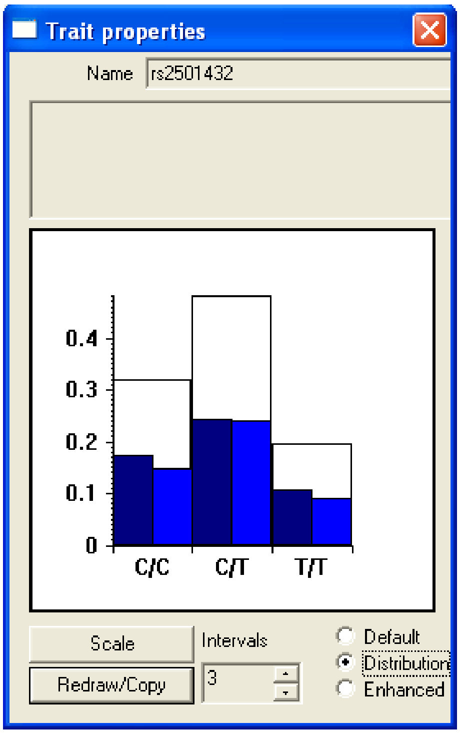
Trait properties

Name: STATURE

```
All= 0.095238   0.252205
Male= 0.010582   0.017637
Female= 0.084656   0.234568
```

Intervals: 20

○ Default
● Distribution
○ Enhanced statistic



Trait properties

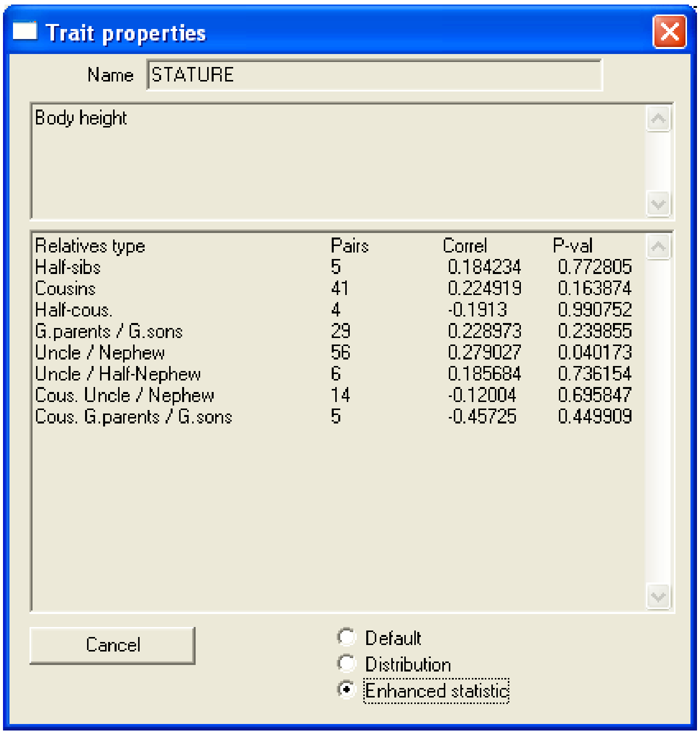
Name: rs2501432

Intervals: 3

○ Default
● Distribution
○ Enhanced

3) *Enhanced statistic.*

This option displays the number of pairs and correlations for all types of non-first order relatives, which have the measured trait values in the sample. For makers only number of relatives' pairs is shown.



Trait properties

Name: STATURE

Body height

Relatives type	Pairs	Correl	P-val
Half-sibs	5	0.184234	0.772805
Cousins	41	0.224919	0.163874
Half-cous.	4	-0.1913	0.990752
G.parents / G.sons	29	0.228973	0.239855
Uncle / Nephew	56	0.279027	0.040173
Uncle / Half-Nephew	6	0.185684	0.736154
Cous. Uncle / Nephew	14	-0.12004	0.695847
Cous. G.parents / G.sons	5	-0.45725	0.449909

○ Default
○ Distribution
● Enhanced statistic

1.7.2 Options for quantitative traits.

The options available for quantitative traits in the pop-up column menu were have already described in section 1.5 (Adjustment), section 1.6.3 (Standardize) and all the rest in section 1.3.

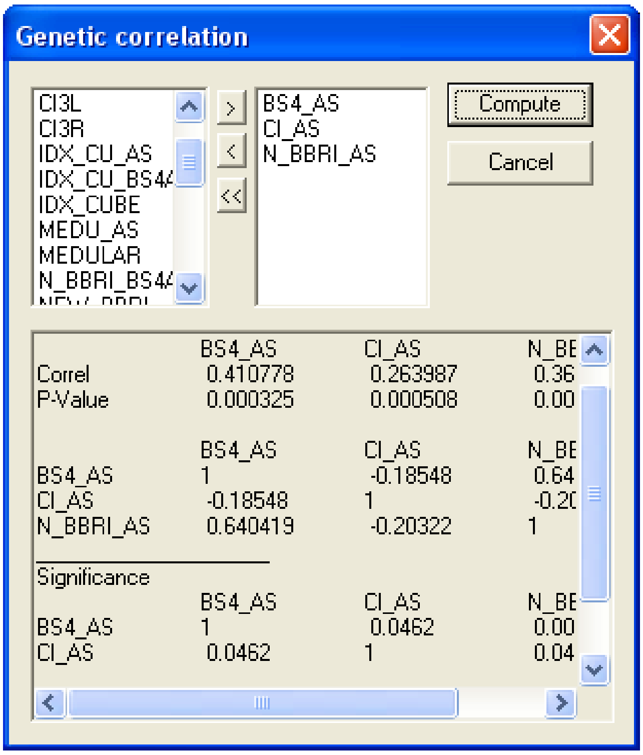
1.7.3 Options for marker traits.

The options available for marker traits in the pop-up column menu were already described in section 1.4 (Test Errors), section 1.6.5 (Factorize), 1.6.4 (N of MinAllele).

1.8 Joint statistics for multiple traits.

1.8.1 Genetic correlation.

The estimates of heritability and genetic correlation in pedigree sample in general can be computed for four research designs (the regression of offspring on mid-parent values, the regression of offspring on single-parent values, the intraclass correlation of full sibs; and the intraclass correlation of half sibs). In our output we use offspring on single-parent correlations. This design includes the most part of phenotypicaly-measured individuals in the case, when the proportion of one-offspring nuclear families is significant and some parents (but not both in the nuclear pedigree) are missing. The meaning of genetic correlation between traits is based on the assumption, that relatives share a definite part of trait variance due to correlated genotypes in a great number of regulating genes. Therefore, if for two different traits we suppose, that some proportion of regulating genes is mutual for both traits, we should expect significant cross-correlation between relatives. This cross-correlation should be normalized on the square root of the product of appropriate relatives' correlation in both traits (for equal or proportional traits we should have genetic correlation 1). Therefore, we should not use for computation traits, if their own parent-offspring correlations (including the shared genetic variance) are non-significant. Therefore, the first table in the output shows parent-offspring correlation and its significance (p-value) separately for each of traits selected for analysis. If N of the selected traits satisfy the condition p<0.05, two N×N tables are built. The first includes elements (i,j) being the parent-offspring cross-correlation of traits i and j, divided on square root of the product of both traits, i and j, own parent-offspring correlations. So, all diagonal elements of the table are equal to 1. The non-diagonal elements (i≠j) of the second N×N table present significance (p-value) of parent-offspring cross-correlation between traits i and j. In this table the diagonal elements are not used and are displayed as 1.

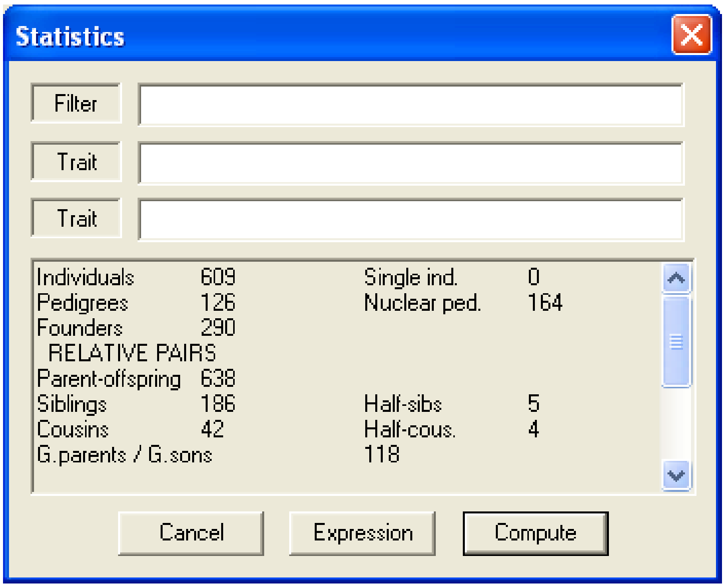


```
             BS4_AS      CI_AS       N_BE
Correl       0.410778    0.263987    0.36
P-Value      0.000325    0.000508    0.00

             BS4_AS      CI_AS       N_BE
BS4_AS       1           -0.18548    0.64
CI_AS        -0.18548    1           -0.2(
N_BBRI_AS    0.640419    -0.20322    1

Significance
             BS4_AS      CI_AS       N_BE
BS4_AS       1           0.0462      0.00
CI_AS        0.0462      1           0.04
```

Click the menu item "View"-"Genetic correlation", the appropriate dialog will be displayed. In the left listbox you can see the list of all quantitative traits. Select by mouse-click the trait, which you are going to use and move it to the right listbox using the ">"-button. Then in such a way move to the right list box other traits, which you need. Click the "Compute"-button. In the lower editbox you will see the output as described above. If you need to analyze some other trait group, change the list of selected traits (right listbox) by means of buttons: "<<"-moves all items to the left and clears the right list of traits; "<"-moves the trait selected by cursor in the right list to the left, ">" vice versa. Clicking on "Compute" will add the new output to the previous one. You can select the desirable part of the output by mouse and copy it into clipboard by <Ctrl-C> hotkey.

1.8.2 Statistics for derived expressions.

It is possible to see descriptive statistics, familial correlations and cross-correlations for virtual traits that are not columns in the data table. They should be defined by the formula. In the same way, you can select a sub-sample of individuals, for which the statistics will be computed, by defining filter as logical expression. Only cases with TRUE expression value will be included in the computation.



```
Individuals      609          Single ind.     0
Pedigrees        126          Nuclear ped.    164
Founders         290
 RELATIVE PAIRS
Parent-offspring 638
Siblings         186          Half-sibs       5
Cousins          42           Half-cous.      4
G.parents / G.sons            118
```

Click the menu item "View"-"Statistics", the Statistics-dialog will be displayed.

If the number of pedigrees in the sample is greater than one, the output editbox (the multi-line editbox) displays the sample structure statistics. The statistics includes the number of pedigrees, nuclear families, unrelated individuals and counts of relatives' pairs of each type, presented in the sample.

There are three fields: one for a filter and two for traits, which can include expressions. Click on the desirable field and then click on the "Expression"-button. Use the Expression editor dialog to build the formula as described in 1.6.1. Then click the "Compute"-button to get the output. If only a filter is defined, the output returns the number of cases with TRUE value of the filter expression. If additionally one trait is defined, the descriptive statistics (number of measured, mean, variance) and familial correlations for the trait will be displayed. If two traits are defined, you get statistics for both traits and additionally self-correlation coefficients and familial cross-correlations.

You can change the definition of trait and filter fields and then click "Compute" once more. Each time as you click "Compute" the syntax and type of non-empty expressions is checked. If there are some uncertainties, you will get the error message; otherwise, the new output will be added to the previous contents of the output-editbox. The current definition of traits and filter is displayed at the beginning of each output. You can select the desirable part of the output by mouse and copy it into clipboard by <Ctrl-C> hotkey.

1.9 Options available in the scatter plot window.

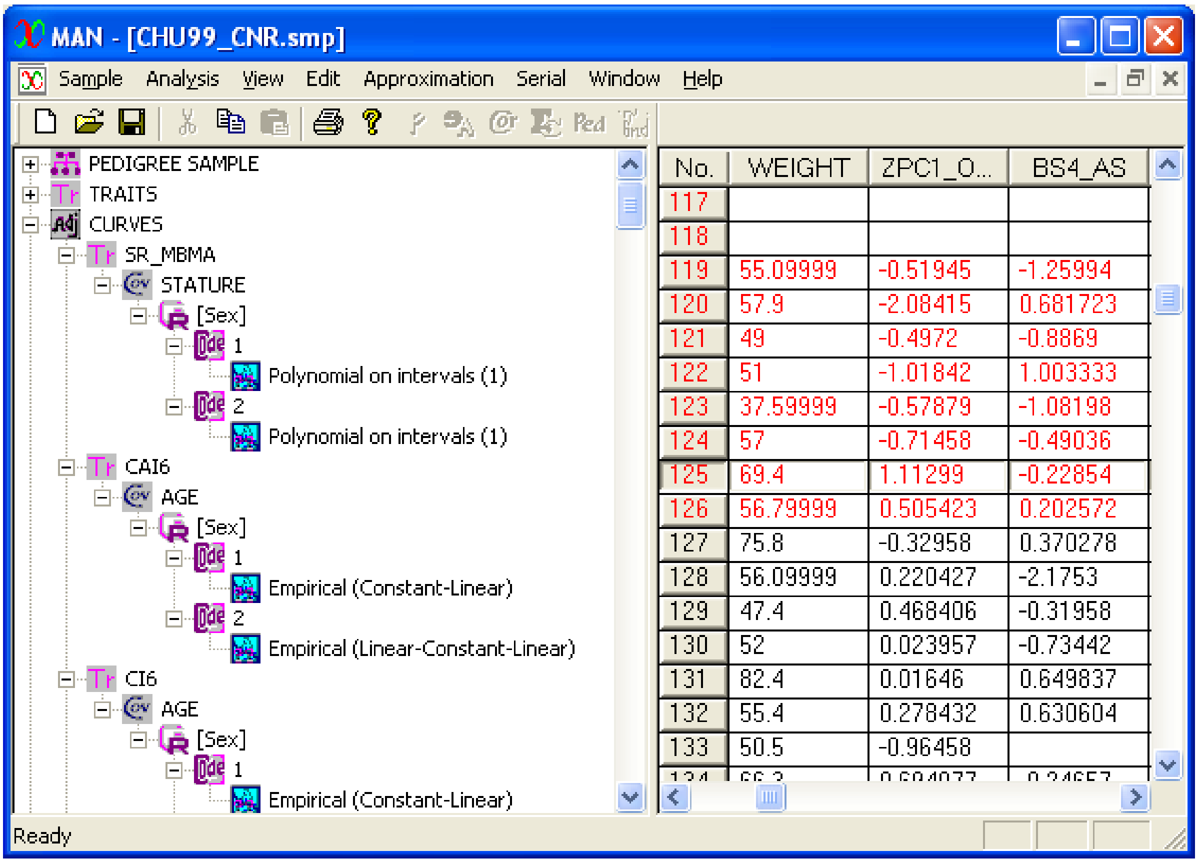
There are two ways to open the new scatter plot window. Click menu item "View"-"Curves"-"New" or make right mouse click on the trait column (real or integer) and select from the popup menu "Adjustment". The *New curve* dialog appears. Select the scatter plot options. *Trait* and *Covariate* can be selected from the list of all quantitative traits, available in the sample. The *Group* can be [Sex](default) or can be selected from the list of integer or categorical traits (text traits or markers). The *Group value* is selected from the list of all possible discreet values of the selected group-trait (for text traits and markers the category number is presented). If the *Group value* is empty or zero, all individuals having non-missing trait and covariate values will be presented in the scatter plot, if the non-zero *Group value* was selected only individuals having this value of the group-trait will be presented. When you click OK the new scatter plot window opens. Then you can use the items of "Approximation" menu to build the best fitting curve of selected type as described in 1.5(EXAMLES).

If you have saved a curve by click on “Edit”-“Add curve” menu item, the curve appears in the document tree as sub-item in CURVES item (1.5.2). The curves for the same trait are united in the <trait_name> child item of the CURVES item. Each <trait_name>-item includes covariates as child items, which in turn include groups and those include curves. To open the saved curve in the scatter plot window you should do double click on the appropriate curve item in the document tree. Otherwise you can click menu item “View”-“Curves”-“Load”, then in the opened dialog select the desired curve in the listbox and click “QK”-button.

When the scatter plot window is active you can use the following options.

1.9.1 Finding of the selected individual.

In the scatter plot window each individual is presented as a circle with vertical coordinate equal to trait and horizontal to covariate (in accordance to scales). When you click on some circle, it appears selected by square cursor. If you click on “Edit”-“Find” menu, the document window becomes active and in the trait table the string, appropriate to selected individual will be visible and highlighted with sunken “No.” button. All strings corresponding to the members of the pedigree, to which the selected individual belongs, appear in red font color.

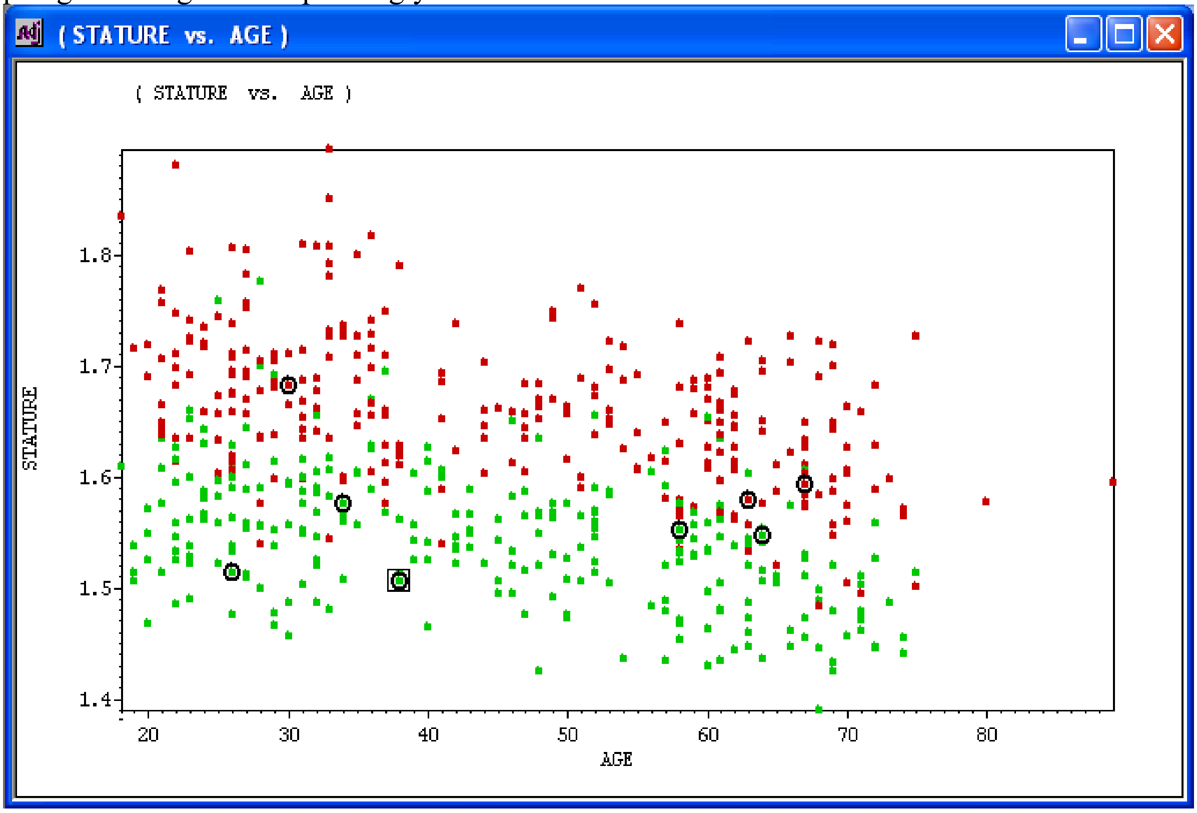


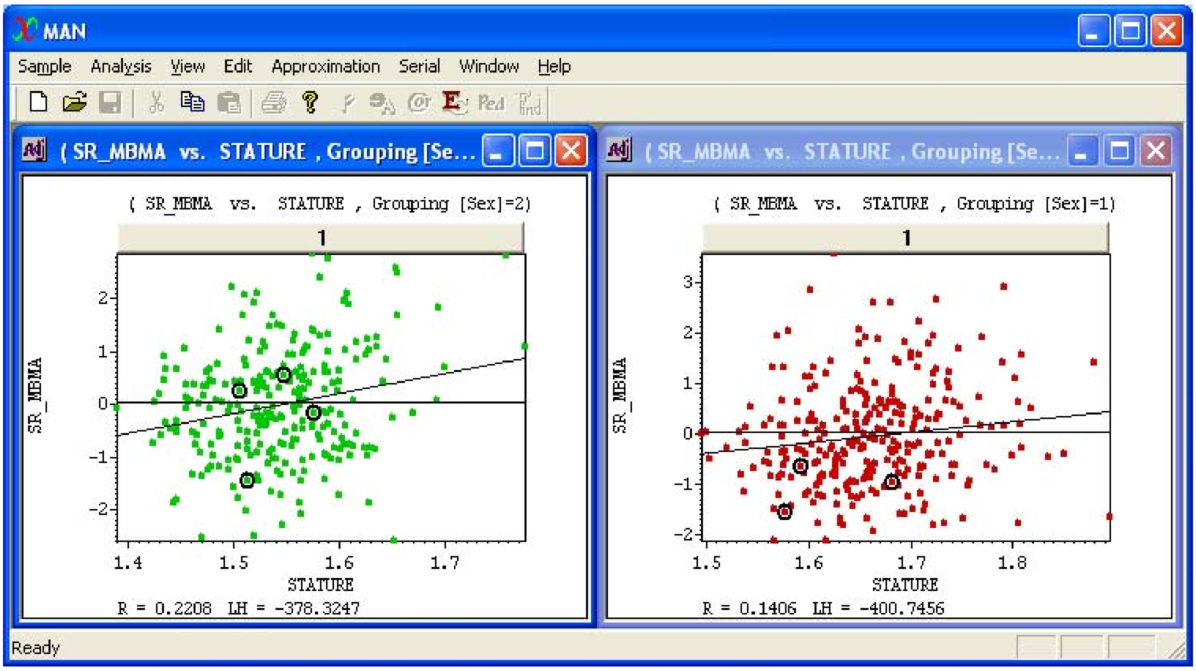
- PEDIGREE SAMPLE
- TRAITS
- CURVES
  - SR_MBMA
    - STATURE
      - [Sex]
        - 1
          - Polynomial on intervals (1)
        - 2
          - Polynomial on intervals (1)
  - CAI6
    - AGE
      - [Sex]
        - 1
          - Empirical (Constant-Linear)
        - 2
          - Empirical (Linear-Constant-Linear)
  - CI6
    - AGE
      - [Sex]
        - 1
          - Empirical (Constant-Linear)

No.	WEIGHT	ZPC1_O...	BS4_AS
117			
118			
119	55.09999	-0.51945	-1.25994
120	57.9	-2.08415	0.681723
121	49	-0.4972	-0.8869
122	51	-1.01842	1.003333
123	37.59999	-0.57879	-1.08198
124	57	-0.71458	-0.49036
125	69.4	1.11299	-0.22854
126	56.79999	0.505423	0.202572
127	75.8	-0.32958	0.370278
128	56.09999	0.220427	-2.1753
129	47.4	0.468406	-0.31958
130	52	0.023957	-0.73442
131	82.4	0.01646	0.649837
132	55.4	0.278432	0.630604
133	50.5	-0.96458	
134	66.3	0.604077	0.24657

1.9.2 Marking of the pedigree members.

If you have clicked on the individuals circle and it appears surrounded with square cursor, you can highlight all individuals belonging to the same pedigree. To do this click “Edit”-“Mark Pedigree” menu item. The members of the pedigree will appear surrounded with black circles. After that, the “View”-“Selected Pedigree“-menu item will be checked. In this regime, if a number of scatter plot windows with different traits are opened, the selected pedigree will be highlighted in all scatter plots. If you double click on some other pedigree in the trait table, the pedigree appear in red and in all opened scatter plots the highlighted pedigree changes correspondingly.




MAN
Sample
Analysis
View
Edit
Approximation
Serial
Window
Help
(SR_MBMA vs. STATURE , Grouping [Se...
(SR_MBMA vs. STATURE , Grouping [Sex]=2)
1
SR_MBMA
2
1
0
-1
-2
1.4
1.5
1.6
1.7
STATURE
(SR_MBMA vs. STATURE , Grouping [Se...
(SR_MBMA vs. STATURE , Grouping [Sex]=1)
1
SR_MBMA
3
2
1
0
-1
-2
1.5
1.6
1.7
1.8
STATURE
Ready

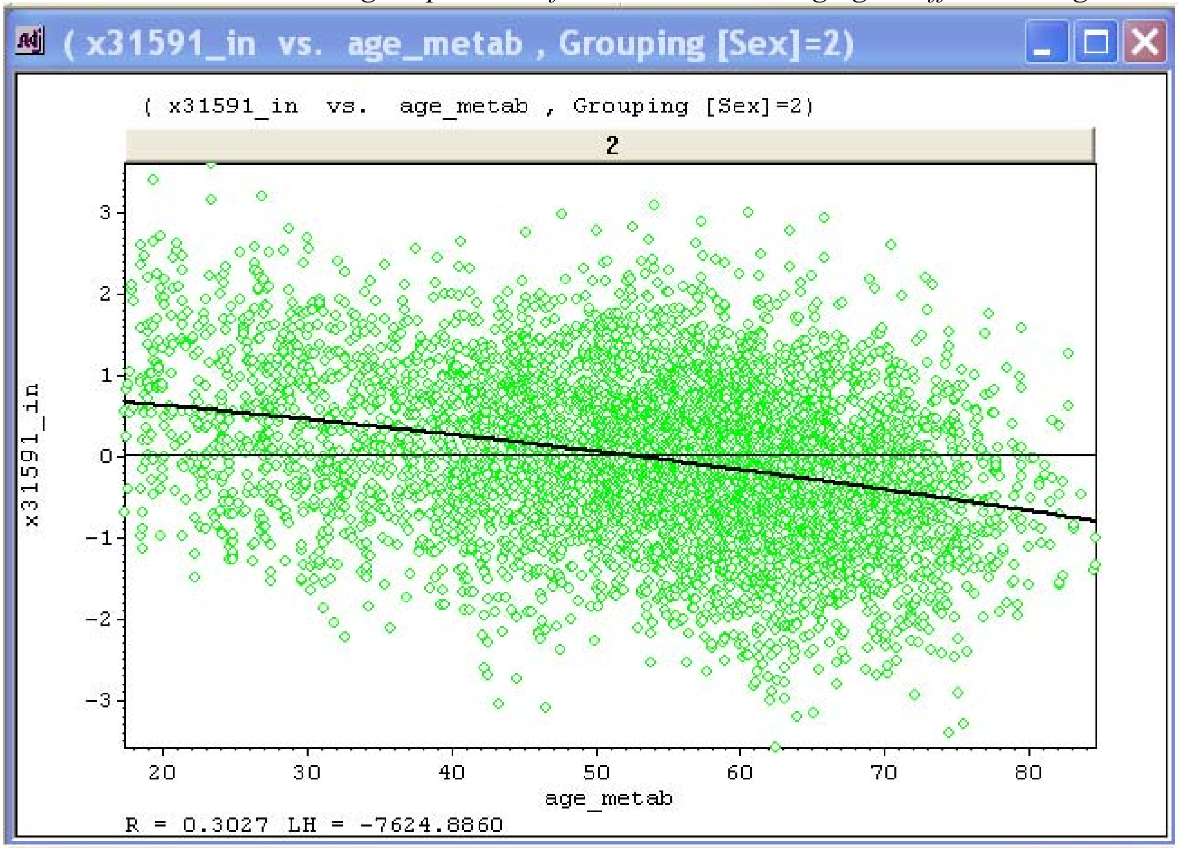
R = 0.2208 LH = -378.3247

R = 0.1406 LH = -400.7456

1.9.3 Working with subgroups.

If the scatter plot window is active and selected approximation curve is one-interval polynomial or growth curve, the presented data can be divided into subgroups and for each subgroup the approximation curve parameters will be computed separately. If the likelihood (LH) of the approximation curve for the whole data has N parameters, the LH accounting for subgroups specific features will have N*K parameters, where K is the number of subgroups. The LRT can show significance of taking into account belonging to the definite subgroup.

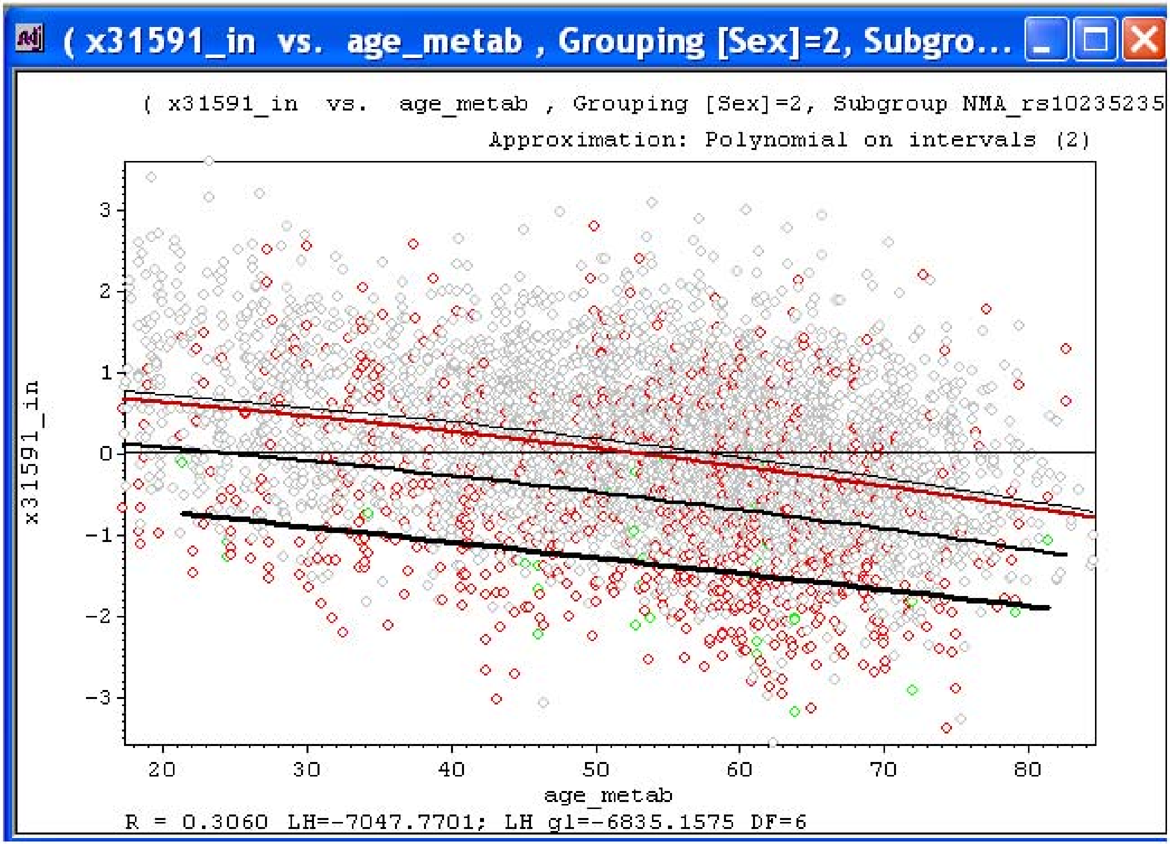
EXAMPLE 1. Trait-age dependence for individuals belonging to different categories.



(x31591_in vs. age_metab , Grouping [Sex]=2)

(x31591_in vs. age_metab , Grouping [Sex]=2)

R = 0.3027 LH = -7624.8860

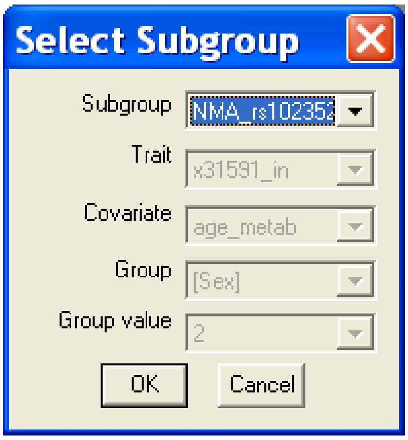


(x31591_in vs. age_metab , Grouping [Sex]=2, Subgro...

(x31591_in vs. age_metab , Grouping [Sex]=2, Subgroup NMA_rs10235235

Approximation: Polynomial on intervals (2)

R = 0.3060 LH=-7047.7701; LH gl=-6835.1575 DF=6



Select Subgroup

Subgroup: NMA_rs102352

Trait: x31591_in

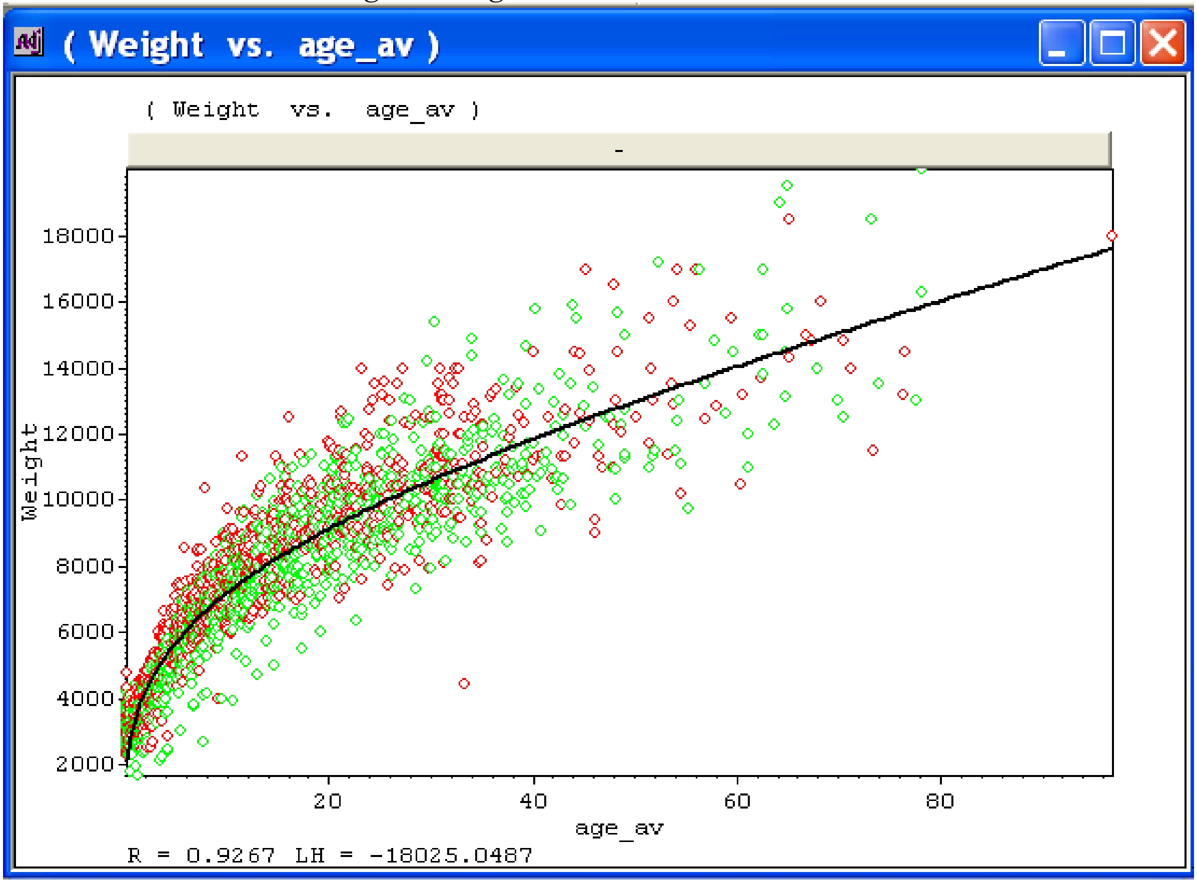
Covariate: age_metab

Group: [Sex]

Group value: 2

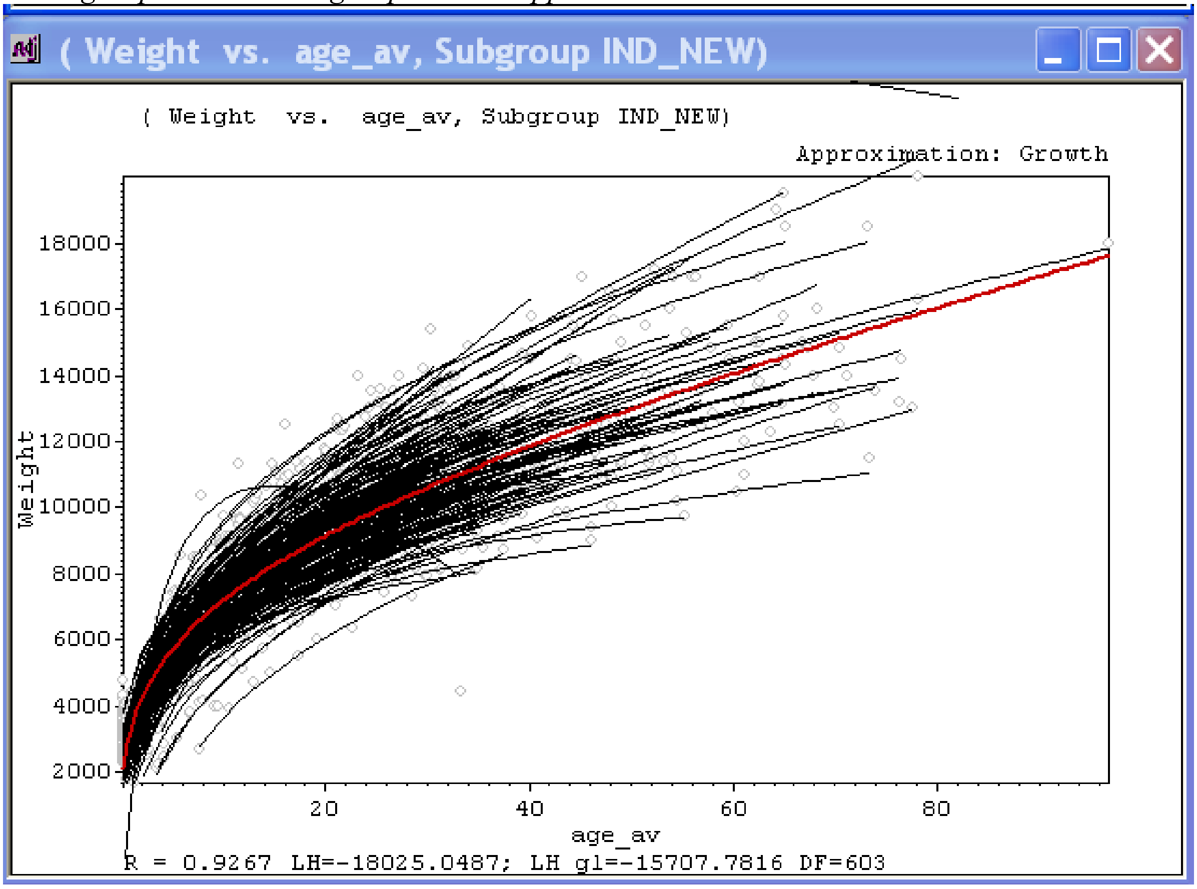
The upper graph presents the dependence of the normalized metabolite (x31591) concentration in female on the age of individual (age_metab). The SNP rs10225235 was shown to affect metabolite concentration. To see the age dependence in female for different SNP genotypes create the integer trait NMA_rs10225235, containing the number of minor alleles of the SNP (see section 1.6.4)make the initial scatterplot window (without subgroups) active and then click View-Curve- Subgroup menu item. The "Select Subgroup" dialog opens. "Trait", "Covariate", "Group" and "Group value" fields determine the data, presented on the initial scatter plot window, and cannot be changed. In the "Subgroup" combobox all integer and text traits of the sample are available, excluding the trait, which was a Group-trait of the initial graph (sex). Select from the list the desirable trait (NMA_rs10225235) and click OK-button. If the data included the initial graph, have less than two subgroup values, you will get an error message, otherwise the Subgroup Graph Window opens. It enables to view a number of curves simultaneously. Separate curve for each subgroup value (black) and the general curve for all data together (red). Likelihood values for general curve and separate parameters for each curve together with DF(degrees of freedom) difference are presented in the bottom of the window. To view the parameters for all subgroup curves click View-Curve-Info menu item.

EXAMPLE 2. Working with longitudinal data.

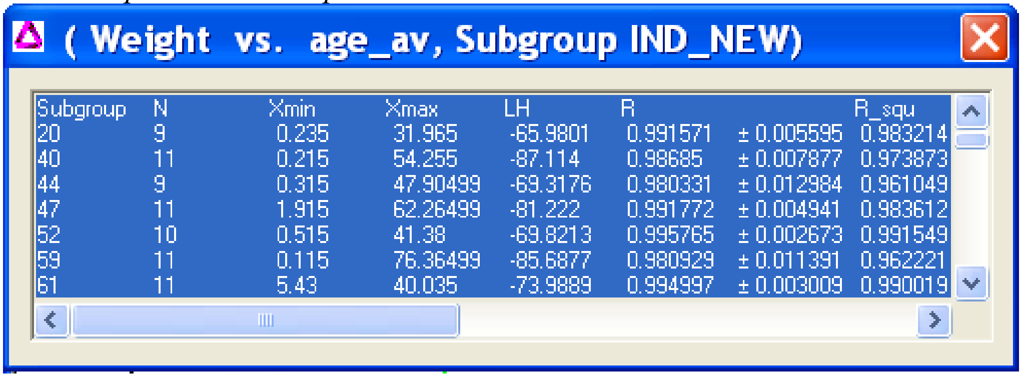


R = 0.9267 LH = -18025.0487

Sometimes longitudinal data for individuals in the sample are sets with different number of mesured points and different intervals of measurement . To use this data for genetic analysis we can built for each individual approximation curve of the same type and use individual *parameters of the curve in the further genetic analysis. Example sample presents children weight for the initial weeks of life. The measurement were taken with different time intervals. All measurement belonging to the same individual have the same ID (IND_NEW)The growth curve (*$Y(X) = A + BX + C\ln(X+1)$*) can be used as approximation. First open scatterplot for Weight on age for all the data. Select Approximation-Growth menu item, the approximation curve appears. Then click View-Curve- Subgroup menu item. Select in the dialog IND_NEW as subgroup trait. The subgroup window appears wit individual curves.*

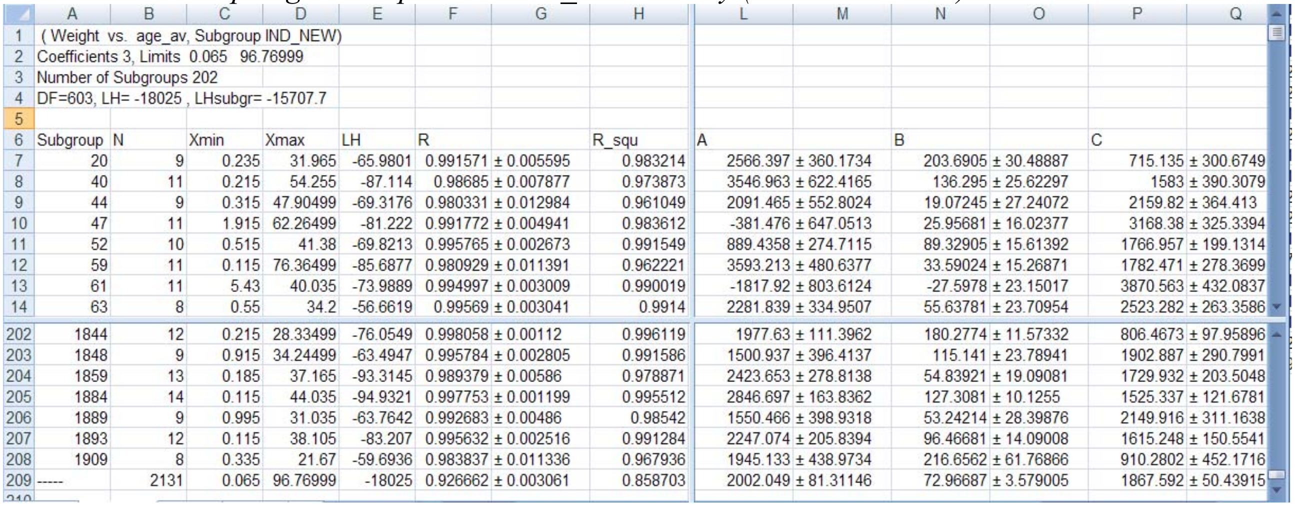

(Weight vs. age_av, Subgroup IND_NEW)
(Weight vs. age_av, Subgroup IND_NEW)
Approximation: Growth
Weight
18000
16000
14000
12000
10000
8000
6000
4000
2000
20
40
60
80
age_av
R = 0.9267 LH=-18025.0487; LH gl=-15707.7816 DF=603

To see curve parameters click View-Curve-Info menu item. In the info dialog editbox all individual curve parameters are presented as table.



Subgroup	N	Xmin	Xmax	LH	R		R_squ
20	9	0.235	31.965	-65.9801	0.991571	± 0.005595	0.983214
40	11	0.215	54.255	-87.114	0.98685	± 0.007877	0.973873
44	9	0.315	47.90499	-69.3176	0.980331	± 0.012984	0.961049
47	11	1.915	62.26499	-81.222	0.991772	± 0.004941	0.983612
52	10	0.515	41.38	-69.8213	0.995765	± 0.002673	0.991549
59	11	0.115	76.36499	-85.6877	0.980929	± 0.011391	0.962221
61	11	5.43	40.035	-73.9889	0.994997	± 0.003009	0.990019

The contents of the edit box can be copied with <Ctrl-C> key and pasted for editing to any program which uses tab-delimited text. The column of the table are: Subgroup value; N-number of measurements; Xmin - left side of the interval; Xmax - right side of the interval; LH-individual likelihood for min square method; R, R_square - multiple regression coefficient the other columns are curve parameters with standard errors. The last string in the table *correspond to general curve for all measurement together. The data in the form of table can be included in the pedigree sample with IND_NEW as key (see section 1.6.8)*

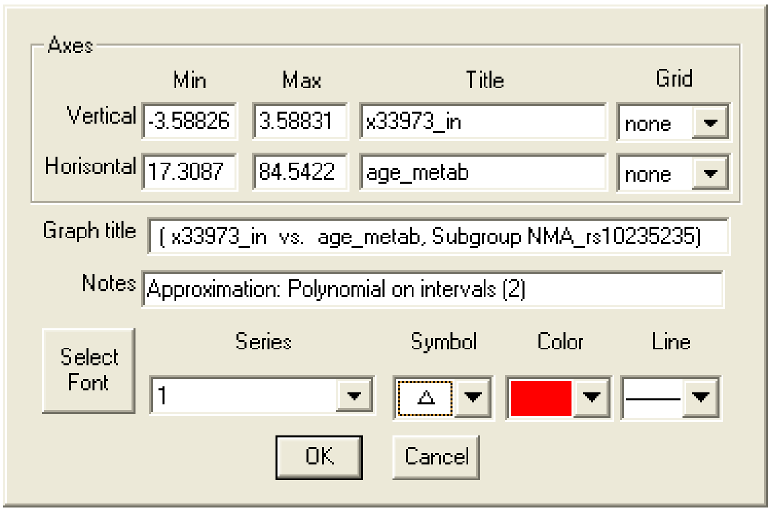


	A	B	C	D	E	F	G	H	L	M	N	O	P	Q
1	(Weight vs. age_av, Subgroup IND_NEW)													
2	Coefficients 3, Limits 0.065 96.76999													
3	Number of Subgroups 202													
4	DF=603, LH= -18025 , LHsubgr= -15707.7													
5														
6	Subgroup	N	Xmin	Xmax	LH	R		R_squ	A		B		C	
7	20	9	0.235	31.965	-65.9801	0.991571	± 0.005595	0.983214	2566.397	± 360.1734	203.6905	± 30.48887	715.135	± 300.6749
8	40	11	0.215	54.255	-87.114	0.98685	± 0.007877	0.973873	3546.963	± 622.4165	136.295	± 25.62297	1583	± 390.3079
9	44	9	0.315	47.90499	-69.3176	0.980331	± 0.012984	0.961049	2091.465	± 552.8024	19.07245	± 27.24072	2159.82	± 364.413
10	47	11	1.915	62.26499	-81.222	0.991772	± 0.004941	0.983612	-381.476	± 647.0513	25.95681	± 16.02377	3168.38	± 325.3394
11	52	10	0.515	41.38	-69.8213	0.995765	± 0.002673	0.991549	889.4358	± 274.7115	89.32905	± 15.61392	1766.957	± 199.1314
12	59	11	0.115	76.36499	-85.6877	0.980929	± 0.011391	0.962221	3593.213	± 480.6377	33.59024	± 15.26871	1782.471	± 278.3699
13	61	11	5.43	40.035	-73.9889	0.994997	± 0.003009	0.990019	-1817.92	± 803.6124	-27.5978	± 23.15017	3870.563	± 432.0837
14	63	8	0.55	34.2	-56.6619	0.99569	± 0.003041	0.9914	2281.839	± 334.9507	55.63781	± 23.70954	2523.282	± 263.3586
202	1844	12	0.215	28.33499	-76.0549	0.998058	± 0.00112	0.996119	1977.63	± 111.3962	180.2774	± 11.57332	806.4673	± 97.95896
203	1848	9	0.915	34.24499	-63.4947	0.995784	± 0.002805	0.991586	1500.937	± 396.4137	115.141	± 23.78941	1902.887	± 290.7991
204	1859	13	0.185	37.165	-93.3145	0.989379	± 0.00586	0.978871	2423.653	± 278.8138	54.83921	± 19.09081	1729.932	± 203.5048
205	1884	14	0.115	44.035	-94.9321	0.997753	± 0.001199	0.995512	2846.697	± 163.8362	127.3081	± 10.1255	1525.337	± 121.6781
206	1889	9	0.995	31.035	-63.7642	0.992683	± 0.00486	0.98542	1550.466	± 398.9318	53.24214	± 28.39876	2149.916	± 311.1638
207	1893	12	0.115	38.105	-83.207	0.995632	± 0.002516	0.991284	2247.074	± 205.8394	96.46681	± 14.09008	1615.248	± 150.5541
208	1909	8	0.335	21.67	-59.6936	0.983837	± 0.011336	0.967936	1945.133	± 438.9734	216.6562	± 61.76866	910.2802	± 452.1716
209	-----	2131	0.065	96.76999	-18025	0.926662	± 0.003061	0.858703	2002.049	± 81.31146	72.96687	± 3.579005	1867.592	± 50.43915

If polynomial approximation curve is used, you can compare goodness of approximation with different polynomial powers for each individual separately using LH columns and LRT.

1.9.4 Editing of the scatter plot.

When the scatter plot window opens, scales for horizontal and vertical axes are defined automatically to represent all available data. The axis titles are also automatically taken as trait and covariate name. The title of the graph includes the title, specified by the user, and also trait, covariate and grouping information in brackets. If some approximation curve was chosen, below the graph the R coefficient (square root from proportion of explained variance, for linear approximation this is a correlation coefficient) and log-likelihood value, LH, are presented. If you compare visually graphs for the same trait, but for different groups, it is suitable to equalize the trait scales. You can do it by editing of the graph in the scatter plot window. To do this, click on the scatter plot window, which you want to edit, to make it active. Then click on the menu item “Edit”-“Edit Curve”. The editing dialog opens. In the *Axes* control group you can specify minimal (*Min*) and maximal (*Max*) values of the axes scales, axes titles (*Title*) and grid line options (*Grid*). The *Graph title* and *Notes* also can be edited. By default the *Notes* field contains the type of approximation curve. If you make the *Notes* field empty the values of R and LH will be not presented in the graph.



Axes

	Min	Max	Title	Grid
Vertical	-3.58826	3.58831	x33973_in	none
Horisontal	17.3087	84.5422	age_metab	none

Graph title: (x33973_in vs. age_metab, Subgroup NMA_rs10235235)

Notes: Approximation: Polynomial on intervals (2)

Series: 1

Symbol: △

Color: 🟥

Line: —

If a number of group (group value was not selected) or subgroup values are presented, for each series its own marker form and color can be selected. By default group value 1 have red color, group value 2 - green and all others gray. Select group value in the "Series" combobox and then select for this set the marker form in the "Symbol" combobox and marker color in the "Color" combobox. If the subgroup window is edited in the "Line" combobox the line weight can be choosen.

After clicking “OK”-button you will see the corrected scatter plot window. The changes are applied only for that window, which was active before you have opened the curve editing dialog. If you have a number of scatter plot windows opened, each window should be edited independently.

1.9.5 Copying of the scatter plots into external documents.

To copy the scatter plot into clipboard, make the appropriate window active and then click the menu item “Edit”-“Copy”. The graph is copied into clipboard as enhanced metafile and then can be included by “Paste” option in Office documents or other programs.

2. Pedigree based analysis.

Pedigree analysis is the study of trait inheritance. The inheritance of the studied trait is made explicit by collecting and analyzing a sample of pedigrees. Most widely used is a term *genetic pedigree analysis*, in which the analytical description of the trait inheritance assumes that the main factors underlying this inheritance are genes – the DNA segments, positioned on the chromosomes and transmitted from parent to offspring in accordance with Mendelian laws.

Pedigree analysis is performed on a *sample* of pedigrees $\{(\boldsymbol{X}_k,\boldsymbol{C}_k)\}$, where is $\boldsymbol{X}_k$ the set of phenotypes observed in k-th pedigree (including probably genotyping results of some chromosomal markers), and $\boldsymbol{C}_k$ is the structure of relations of pedigree members. Elston (1998) distinguished model-based and model-free pedigree analyses. In the first, we formulate model of the trait inheritance and the sampling procedure that are used, while in the second the analysis proceeds without such explicit models. Transmission disequilibrium test, discussed later, is an example of model-free analysis. The most part of analyses implemented in MAN are model-based.

2.1. Genetic model.

Let $\mathbf{X} = \{x\}$ be a set of possible phenotypic characteristics and $\mathbf{G}$ - the set of genotypes that are involved in controlling the trait being studied. Denote by $X_n = \{x_1,...,x_n\}$, $x_i \in \mathbf{X}$, the set of phenotypes observed on the n members of a pedigree and by $G_n = \{g_1,...,g_n\}$ $g_i \in \mathbf{G}$, - their given set of genotypes. Define a *genetic model* for the inheritance of a trait as the following three distributions determined on two sets, $\mathbf{X}$ and $\mathbf{G}$:

$$\theta = \{\, p(g_1,g_2), P(g\,|\,g_1,g_2), f(X_n\,|\,G_n)\,|\,\mathbf{X},\mathbf{G}\}, \tag{1.1}$$

Where $p(g_1,g_2)$ is the joint population distribution of genotypes in spouse pairs determined by the population mating structure. $P(g|g_1,g_2)$ is the conditional probability that an offspring receives genotype g from parents having genotypes g_1 and g_2 - the core of the genetic inheritance. $f(X_n|G_n)$ is the joint distribution of the phenotypes of the n pedigree members given their set of genotypes G_n.

The probability $P(\mathbf{X},C|\theta)$ of a sampled pedigree having the structure C and phenotypic content $\mathbf{X}$, formulated based on genetic model, θ, is called the pedigree likelihood . If, as is usually the case, the pedigrees are sampled independently of one another, then the sample likelihood is simply the product of the likelihoods for the pedigrees included in the analysis (or equivalently the sum of their log-likelihoods). The explicit formulation of genetic model components and their parameters can be different reflecting differences in population properties, specific phenotype features and mathematical form of distribution, used to describe genotype- phenotype correspondence in the pedigree. The parameters of genetic model are estimated at the point of the sample likelihood maximum in the multi-dimensional parameter space. The likelihood ratio test (LRT) is used to compare different models, in order to test definite hypotheses and to build most parsimonious model of trait inheritance.

2.2. Analysis window.

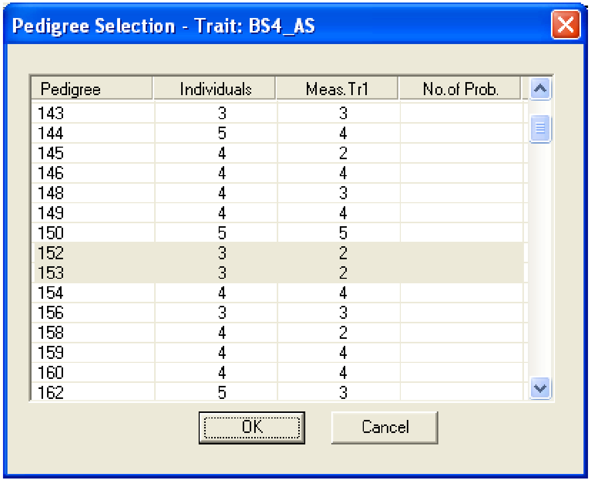
Analysis windows are available from the "Analysis" menu. All analysis windows have the same structure. The left side of the window contains three control groups: TRAITS, PEDIGREE DATA and MODEL FORMULATION. The right side presents the table of model parameters.

2.2.1. Traits control group.

The traits, used in the particular analysis, are specified in the first control group. For segregation (SA) and variance-component (VCA) analyses they are: the investigated inherited traits and their covariates. For linkage and disequilibrium analyses additionally genetic markers (traits containing genotyping results for persons) are specified here. The count of traits to be specified varies depending on the analysis. Each trait is specified in the appropriate combobox (or listbox for multiple covariates in VCA) and should be selected from the list of possible traits of appropriate type presented in the sample.

2.2.2. Pedigree Data control group.

Information of analyzed and excluded from the analysis pedigrees and individuals counts is displayed in the second control group. Pedigrees, which have less then two measured individuals or only two measured individuals, which are spouses, are excluded automatically. You can additionally exclude some other pedigrees by clicking "List of pedigrees"-button and checking the desirable pedigrees in **Pedigree Selection** dialog. *All highlighted pedigree strings in the list will be excluded from the analysis if you finish the editing with "OK". To select multiple strings press the <Ctrl>-key and click the desirable additional pedigree string in the list.*



Pedigree Selection - Trait: BS4_AS

Pedigree	Individuals	Meas.Tr1	No.of Prob.
143	3	3	
144	5	4	
145	4	2	
146	4	4	
148	4	3	
149	4	4	
150	5	5	
152	3	2	
153	3	2	
154	4	4	
156	3	3	
158	4	2	
159	4	4	
160	4	4	
162	5	3	

2.2.3. Model Formulation control group.

Basic model or design options of analysis performed are stated in the third control group. The available options should be selected in appropriate comboboxes. When you change some model formulation option, the availability of other options can change.

2.2.4 Table of model parameters

Purpose. This table contains a list of numerical model parameters, their steps for maximization, upper and lower limits of possible values and constraints applied. When the mouse cursor stopped on the button with parameter denotation, tip with parameter meaning is displayed.

Constraints. The parameter constraints can be introduced, changed or canceled. When you click on parameter button, if constraint button is empty and disabled, the menu with all possible constraint options appears. If some constraint from the menu was selected, it appears as constraint-button in the parameter row. The click on the constraint-button cancels the chosen constraint.

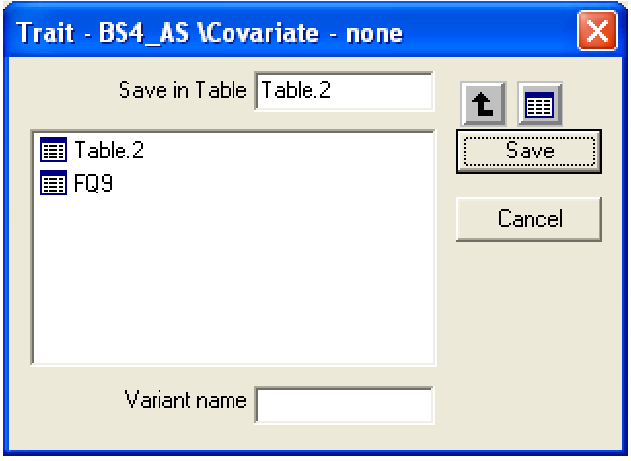
Editing values. Initial parameter value in the parameter row may be edited immediately in the table, clicking by mouse on the value starts editing. Steps, left and right parameter limits can be also edited in such a way.

Displaying intermediate results. For model based analyses, when maximization process is in progress, current parameter values and step values are shown in the table. In addition, the number of current iteration and log-likelihood level are displayed below the table.

Sequential maximizations. For model-based analyses, it is possible to use parameter estimates in the previous maximization as initial values for the next one. The new maximization should begin with sufficiently large step values. To set step values you can edit them, but for most part of analyses click on the "Step" header in the parameter table returns all steps to the initial (large) values and you can run the new maximization.

2.2.5 Performing analysis and saving results.

When the analysis is fully specified, click the "Run"-button. When the maximization of the model starts, parameters and steps will change simultaneously with the number of iteration, displayed below the parameter table. Only single maximization is possible at a time. You can stop the maximization by "Cancel"-button. When the maximization is finished, you get a message. If an error message appears, this means that initial parameters give a negligible likelihood for the specified model. You can change (edit) the initial values of parameters and start the maximization once more. When the maximization finished normally, the "Save"-button will be available (for segregation analyses the "Run" button will be disabled to prevent you from running the new maximization without saving the results). You can save the analysis result (parameter values and LH) as a column in a results table. Click the "Save"-button to display the Save-dialog. On the title bar of the dialog, the traits taking part in the analysis are displayed. If there are already some tables, having the same type of analysis and the same traits taking part in it, you will see the list of these tables. You can chose one of the available tables or create a new table by clicking on the button with table-icon. Double click on the table in the list displays the column (variant) names in this table. You can return to the list of tables by clicking button with "LevelUp"-icon. Specify table and variant name and click "Save". The dialog closes and you return to the model window. Now the "Run" button is enabled and you can use the displayed

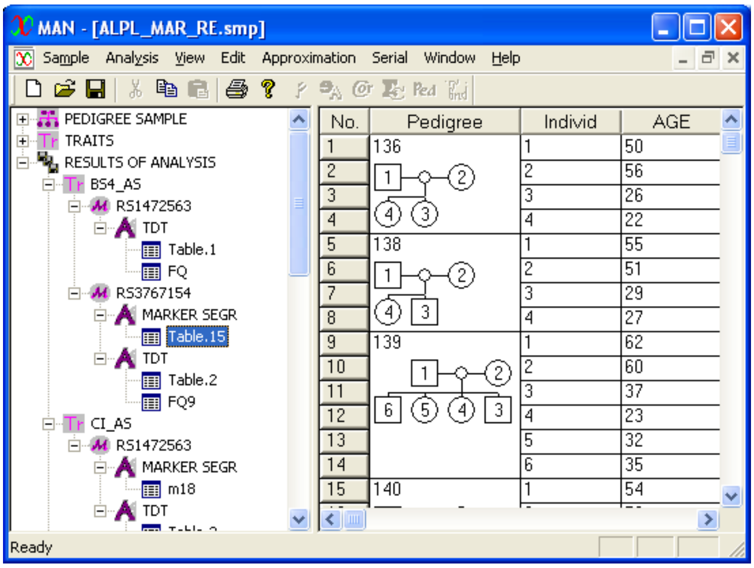


Trait - BS4_AS \Covariate - none

Save in Table: Table.2

Table.2
FQ9

Variant name:



MAN - [ALPL_MAR_RE.smp]

PEDIGREE SAMPLE
TRAITS
RESULTS OF ANALYSIS
BS4_AS
RS1472563
TDT
Table.1
FQ
RS3767154
MARKER SEGR
Table.15
TDT
Table.2
FQ9
CI_AS
RS1472563
MARKER SEGR
m18
TDT

No.	Pedigree	Individ	AGE
1	136	1	50
2		2	56
3		3	26
4		4	22
5	138	1	55
6		2	51
7		3	29
8		4	27
9	139	1	62
10		2	60
11		3	37
12		4	23
13		5	32
14		6	35
15	140	1	54

Ready

parameter values as initial for a new maximization. Specify, if you need, new constraints and then click on the "Step" header in the parameter table to set sufficiently large steps.

2.3. Analysis result tables.

When you have saved some analysis results, you can see the list of result tables in the document tree in the left side of the document window. All tables are included in the RESULTS OF ANALYSIS item. They are arranged hierarchically in accordance with traits taking part in the analysis and analysis type. Double click on the table name opens the appropriate table in a separate window. If not the whole table contents can be seen in the table window, you can scroll it horizontally or vertically.

2.3.1. Tools menu.

"Tools"- menu is available when the table window is active.

Export to Excel. This item in the "Tools"-menu opens the excel file containing the current table. Here the table can be edited.

Delete Column. To delete some column from the table, select the column by left mouse click and then click on "Tools"-"Delete Column"-item.

Error Function. If the analysis enables the standard error (SE) computation of the model parameters, you can select by click a column of the table and then click "Tools"-"Error Function". Appears the analysis window, in which error computation for selected model has started. When the computation finishes, you get a message and should save parameters with their SE in a new column of the table (using "Save"button).

Error Correlation. As a rule, parameters of the model are not orthogonal to each other. If you have computed and saved parameters with SE, you can also see the error correlation table. Select the appropriate column with SE in it and click "Tools"-"Error Correlation"-item. The Correlation table opens. You can export it to Excel by "Tools"-"Export to Excel"-menu item.

Diagram. This "Tools"-menu item is enabled only for specific analyses results and present them graphically.

2.3.2. Using parameters from the result table for other analysis.

When the results table window is active, some models (columns of the table) with their parameter values and constraint can be loaded into analysis window. For example, results of segregation analysis can be loaded in Segregation, or Bivariate, or Model Based Linkage windows to perform analysis with model parameters as initial values. To do this select the desirable table column and then click on the appropriate item in the "Analysis"-menu.

3. Segregation analysis.

The simplest version of the genetic model is the major gene (MG) model. This model explicitly includes two kinds of effects. First, the effect of a diallelic gene called MG, which forms a three-component genotypic set G; for two alleles, A_1 and A_2, the genotypes are A_1A_1, A_1A_2 and A_2A_2, which we can number g = 1, 2 and 3, respectively. Second, all the other effects involved in the trait control that determine the joint phenotypic distribution among members of the sampled pedigree.

3.1. Model parameterization.

3.1.1. Population characteristics.

The first distribution determining the genetic model (1.1) is the genotypic distribution of pairs of spouses. This is a characteristic of the population. The distribution $p(g_1,g_2)$, where g_1 and g_2 are the genotypes of two spouses, is determined by the genotype frequencies and by the type of assortative mating occurring with respect to the trait under consideration. Under panmixia, p(g$_1$,g$_2$) = p$_{g1}$×p$_{g2}$, where p$_g$ is the population frequency of genotype g. For the Hardy-Weinberg equilibrium genotype frequencies are given by: p$_g$ ={ p^2, 2p(1-p), (1-p)^2} for g = 1, 2 and 3, respectively (here p is the population frequency of allele A$_1$).

Assortative mating is a very complex process governed by a number of physical and social characteristics of the mates. Formally, any assortative mating effect is expressed in genetic model terms as the inequality: p(g$_1$, g$_2$) ≠ p$_{g1}$×p$_{g2}$. Assume that the probability of mating between a pair of individuals with genotypes g_i and g_j (g_i , g_j ∈ **G**) is proportional to a factor $q(g_i)q(g_j)\Psi(g_i,g_j)$. Where $q(g_i)$ = p_{g_i}, if the i-th spouse has no parents included in the analyzed pedigree (founder), and $q(g_i) = P(g_i|g_m,g_f)$, if g_m and g_f are genotypes of the parents of the i-th member. The assortative mating factor is of the form: $\Psi(g_i,g_j)$ = $exp[\varphi\times(\mu_{g_i} - \mu_{g_j})^2/\sigma^2_\mu]$, where φ is the coefficient of non-random mating; μ_g is the genotypic value of genotype g, and $\sigma^2_\mu = \sum p_g\mu^2_g$ - $[\sum p_g\mu_g]^2$ is the genotypic variance. The normalizing factor for pedigree likelihood we calculate for the whole pedigree.

3.1.2. Transmission probabilities.

Each offspring genotype g is formed from two parental haplotypes (ξ,η) or (η,ξ), where ξ,η ∈ {h} – the set of haplotypes from which offspring genotypes can be formed. We now introduce the *transmission probability*, P(ξ | g), the probability that an individual having genotype g ∈ G produces the haplotype ξ∈ {h} in the formation of his/her offspring genotypes (Elston and Stewart, 1971). Using these transmission probabilities, we can express the second distribution defining the genetic model (1.1), i.e., the distribution of offspring genotypes given the genotypes of their parents, or transition probabilities, $P(g|g_1g_2)$, as follows:

$$P(g \mid g_1,g_2)=[\, P(\xi \mid g_1)\, P(\eta \mid g_2)+(1-\delta_{\xi\eta})\, P(\eta \mid g_1)\, P(\xi \mid g_2)].$$

Where g = (ξ,η) = (η,ξ) and $\delta_{\xi\eta}$ =1 if ξ=η and $\delta_{\xi\eta}$ =0 if ξ≠η (the Kronecker symbol). By definition, the equality $\sum_{\xi\in\{h\}}P(\xi\,|\,g)=1$ is true for any possible g. Therefore, to parameterize the MG model we need only 3 independent transmission probability parameters τ_1, τ_2, τ_3 , which are probabilities to transmit to offspring allele A$_1$ from parent having genotype g=1,2,3 correspondingly.

3.1.3. Genotype-phenotype correspondence.

For a continuous quantitative trait, the conditional distribution of trait values among individuals having the same genotype g, $f(x|g)$, is usually assumed (after transformation, if necessary) to be normal with expectation μ_g (genotypic value) and residual variance σ^2_g. This residual phenotypic variation is caused by all the factors (genetic and environmental), other than the genes defined in the model **G.** Two groups of factors cause the residual co-segregation between trait values in relatives. First are potential minor-genes that are involved in the trait control, but have not been identified in the model explicitly. Second are common familial environmental (household) factors.

Usually, the n-variable normal approximation is used as the joint distribution of trait residuals $f(X_n|G_n)$ of the n members of a pedigree, the residual of the i-th individual being defined as x_i - μ_{g_i}, where x_i is his/her trait value and μ_{g_i} is the genotypic value of his/her genotype g_i ∈ **G.** This distribution is determined by an n×n symmetric covariance matrix $\{\sigma_{g_i}\sigma_{g_j}r_{ij}\}$, where $\sigma^2_{g_i}$, $\sigma^2_{g_j}$ are the residual variances of genotypes g_i, g_j ∈ **G** and r_{ij} is the pairwise correlation coefficient between the residuals of relatives i and j. The model parameterization of the n-variable normal distribution can be made in various ways. Ginsburg (1997) proposed a parameterization in which the *partial correlations* (elements of a matrix inverse to correlation matrix R) between pedigree members are used as model parameters with following three assumptions:

1) Three partial correlation coefficients are introduced as the model parameters: ρ - between spouses, β - between parents and offspring and ε - between sibs.

2) The partial correlation between any given pair of relatives (spouses, siblings, etc.) is the same regardless of either the particular position of this pair in the pedigree under consideration or of the pedigree structure.

3) The partial correlation between any pair of pedigree members not belonging to the same nuclear family equals zero.

3.1.4. Trait covariates.

The simplest way to account for age and sex (probably some other covariates) effects on inter-individual trait variation is the regression adjustment of the trait values for age and sex effects made prior to any pedigree analysis. The better way is to incorporate the age and sex effects explicitly into a penetrance function determining the MG effect.

Let x_{gst} be the trait value in an individual having MG genotype g, sex s (m - male, f - female) and age t ($t > 0$) . The following linear model is assumed for this trait:

$$x_{gst} = \mu_{gst} + \xi = \mu_{gs} + \varphi_{gs}(t) + \xi.$$

Where μ_{gst} is the expected trait value of individuals having the same g, s and t. ξ is the trait residual not affected by the MG, sex or age. μ_{gs} is the expected trait value of individuals having genotype g and sex s. $\varphi_{gs}(t)$ is a function describing the genotype-sex specific age dependence of the trait value of these individuals. As implemented in MAN, $\varphi_{gs}(t)$ can be linear or quadratic function of age t with parameters a_{gs}-linear coefficient and b_{gs}- quadratic. Two additional formulations of $\varphi_{gs}(t)$ constant-linear and linear-constant, that are of special anthropological interest and can be used to approximate age dependent changes for example in bone anatomy. The parameters are T_{gs} - a genotype-sex specific age threshold, and a_{gs} - a slope coefficient, measuring the rate of change in the trait per year. The constant component of $\varphi_{gs}(t)$ is always computed from the condition $\sum_t \nu_t \varphi_{gs}(t) = 0$, where v_t is the frequency of individuals having age t.

3.2. Transmission probability tests.

Elston and Stewart (1971) introduced a transmission probability model, under which to test hypotheses of inheritance in pedigree segregation analysis. To test for a MG model, where the trait is under the control of a single locus with two alleles, A_1 and A_2, their statistical model can be outlined as follows. The transmission probabilities, τ_g = Pr($A_1|g$), that a parent with genotype g (g = 1, 2 and 3 for genotypes A_1A_1, A_1A_2 and A_2A_2, respectively) transmits allele A_1 to his/her offspring, are estimated from the pedigree sample to have arbitrary values, together with other parameters specified by the model. Now two specific genetic hypotheses can be tested. 1) The transmission is Mendelian, then τ_g = 1.0, 0.5 and 0.0 for g = 1, 2 and 3, respectively. 2) All three τ_g are equal to the same value, $\bar{\tau}$ (which means that the offspring genotype is independent of his/her parental genotypes). Accordingly, two likelihood ratio tests (LRT) are introduced ("transmission probability tests") for monogenic model θ_i:

$$\lambda_{i1} = 2\sum_k \ln[P(X_k, C_k \mid \hat{\theta}_i, \hat{\tau}_g) / P(X_k, C_k \mid \hat{\theta}_i, \tau_M)]$$

and

$\lambda_{i2} = 2\sum_k \ln[P(X_k, C_k \mid \hat{\theta}_i, \hat{\tau}_g) / P(X_k, C_k \mid \hat{\theta}_i, \bar{\tau})]$.

Here τ_M denotes the Mendelian transmission probabilities, and $\hat{\tau}_g$ and $\bar{\tau}$ are estimates of the transmission probabilities found together with other model parameters ($\hat{\tau}_g$-three arbitrary unconstrained parameters, $\bar{\tau}$ - equal τ_g).

For the pedigree sample $\{(X_k,C_k)\}$, the null hypothesis, H_0: $\tau_g = \tau_M$, that the trait is really controlled by one Mendelian diallelic locus, can be tested by the LRT statistic which, if the estimate $\hat{\tau}_g$ is unconstrained, is distributed asymptotically as the central χ^2 with df = 3. The distribution of the second test statistic was assumed asymptotically to be χ^2 with 2 df, if the offspring genotype is really independent of the parental genotypes. Using these tests, the MG trait description is accepted if Mendelian model is not rejected by the test λi_1 and, at the same time, equal τ_g model is rejected by the test λi_2.

3.3. Most parsimonious models.

Any accepted parameter constraint leads to a particular simplified model of inheritance (additive trait control, equal residual variances for the three major genotypes, etc). By making the description of the mode of inheritance more “economical”, such a constraint clearly decreases the maximum of pedigree sample likelihood, in comparison with that of a model not having that constraint. If this decrease is found to be statistically non-significant by LRT, the simpler model would be preferred. The *most parsimonious* (MP) model is defined as the one, for which any further parameter constraint is rejected statistically (with predefined type I error and power that depends on the amount of information in the sample).

3.4. Operating in Segregation analysis window.

Click on “Analysis”-“Segregation”-“Conventional” menu item. The Segregation analysis window opens. Fist of all select in “Analyzed trait”- combobox the trait from the list of all real and integer traits available in the sample. Preferable is to standardize the trait before the analysis to avoid troubles during maximization. In “Covariate”-combobox you can select a covariate from the list of integer traits. After selection of traits you can see what number of pedigrees and observed individuals are excluded from the analysis automatically (see 2.1.2 section).

3.4.1. Model formulation.

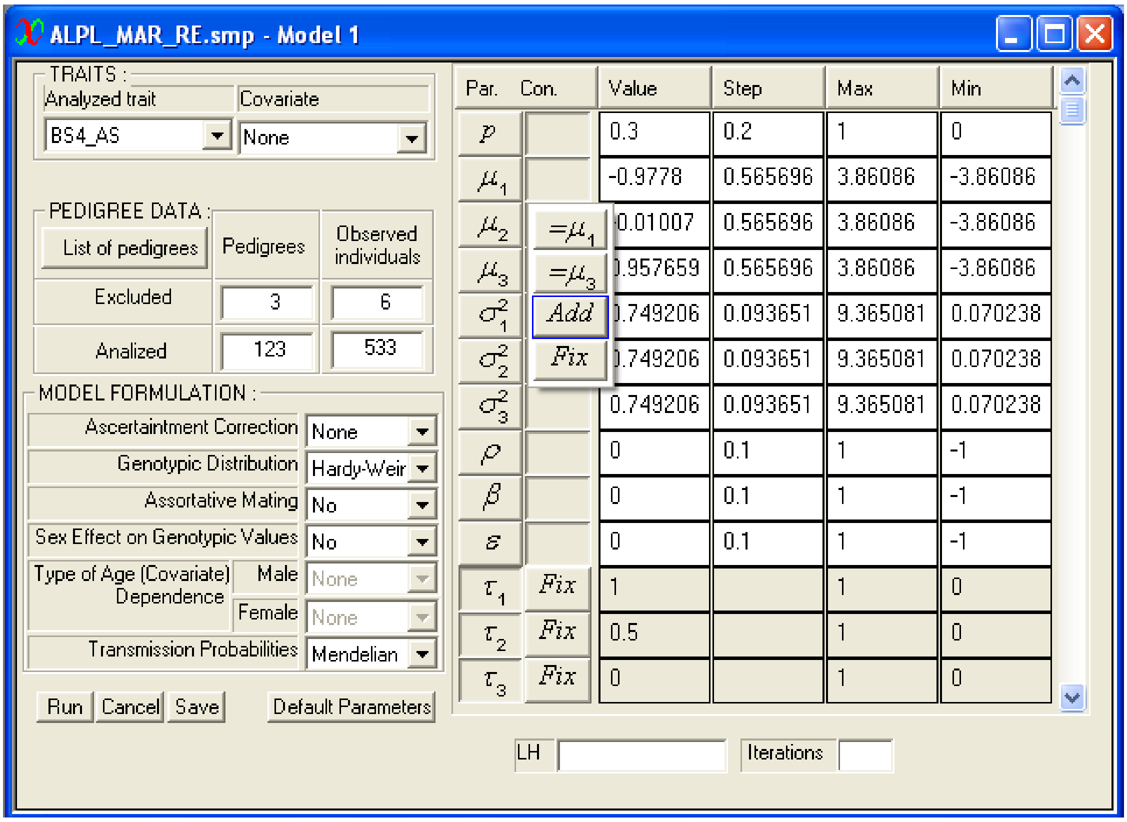
Ascertainment Correction. If you data was sampled through proband and proband-individuals are marked with 1 or 0 in “Proband”-column, you can chouse AMF (ascertainment model free) correction for Segregation analysis. In this case the likelihood conditional on the proband trait values is computed. If there are no probands in the sample, this choice has no effect.

Genotypic Distribution. By default Hardy-Weinberg equilibrium is chosen and single parameter p denotes a frequency of allele A_1 in the population. Genotype frequencies are computed as given in 3.1.1. For “Arbitrary” genotype distribution in place of p two parameters p_1, p_2 appear in the parameter table. They are population frequencies of the first (A_1A_1) and second (A_1A_2) genotypes correspondingly.

Assortative Mating. If you chose “Yes” the assortative mating parameter φ (see 3.1.1) is included in the model and in the parameter table.

Sex Effect on Genotypic Values. By default “No”, sex effect is not included in the model. Shared by both sexes genotypic values μ$_1$, μ$_2$, μ$_3$ and corresponding shared parameters of covariate dependence *a, b* or *a, T* (see 3.1.4), if actual covariate and type of dependence were chosen, are present in the parameter table. If “Yes” is chosen, the double set of parameters μ$_{m1}$, μ$_{m2}$, μ$_{m3}$ for males and μ$_{f1}$, μ$_{f2}$, μ$_{f3}$ for females appears in parameter table. If actual covariate and type of covariate dependence were chosen, the covariate dependence parameters appear for each sex with subscripts *m* and *f*.

Type of Age (Covariate) Dependence. This choice is available only if an actual covariate was chosen. Without sex effect, the covariate dependence chosen in the “Male”-combobox is used for both sexes (“Female”-combobox is disabled) and parameters, appearing in parameter table, have only genotype subscript (1, 2, 3). If sex effect option was chosen, both “Male”- and “Female”-comboboxes are enabled. You can choose different types of dependence for different sexes. Parameters, appearing in parameter table, have two subscripts: sex (*m* or *f*) and genotype (1, 2, 3).
The list of covariate dependence types is as follows (see 3.1.4):
None – no covariate dependence means constant genotypic values.
Linear – linear polynomial covariate dependence; parameters a_{sg}
Quadratic - second power polynomial covariate dependence; parameters a_{sg} , b_{sg}.
Linear+None – two-interval linear –constant dependence; parameters a_{sg} , T_{sg}.
None+Linear – two-interval constant-linear dependence; parameters a_{sg} , T_{sg}.



ALPL_MAR_RE.smp - Model 1

TRAITS :

Analyzed trait	Covariate
BS4_AS	None

PEDIGREE DATA :

List of pedigrees	Pedigrees	Observed individuals
Excluded	3	6
Analized	123	533

MODEL FORMULATION :

Ascertaintment Correction		None
Genotypic Distribution		Hardy-Weir
Assortative Mating		No
Sex Effect on Genotypic Values		No
Type of Age (Covariate) Dependence	Male	None
	Female	None
Transmission Probabilities		Mendelian

Par.	Con.	Value	Step	Max	Min
p		0.3	0.2	1	0
μ_1		-0.9778	0.565696	3.86086	-3.86086
μ_2		0.01007	0.565696	3.86086	-3.86086
μ_3		0.957659	0.565696	3.86086	-3.86086
σ_1^2		0.749206	0.093651	9.365081	0.070238
σ_2^2		0.749206	0.093651	9.365081	0.070238
σ_3^2		0.749206	0.093651	9.365081	0.070238
ρ		0	0.1	1	-1
β		0	0.1	1	-1
ε		0	0.1	1	-1
τ_1	Fix	1		1	0
τ_2	Fix	0.5		1	0
τ_3	Fix	0		1	0

Transmission Probabilities. Three initial options (Mendelian, Arbitrary and Equal) are described in 3.2. Two additional options ([*Arbitrary] and [*Equal]) have in comparison with Arbitrary and Equal one more constraint: transmission probabilities should correspond to the condition, that the genotype frequencies in the next generation coincide with those of the previous generation. Without assortative mating, this condition provides connection between transmission probabilities and population genotype frequencies. In particular, for Hardy-Weinberg equilibrium [*Equal] means that all τ -s should be not only equal, but also equal to the population frequency of the allele $A_{1.}$ Number of degrees of freedom for LRT between [*Arbitrary] and Mendelian models is 2, between [*Arbitrary] and [*Equal] models also 2.

3.4.2. Specifying parameter table.

You can edit the initial values, steps and limits of parameters immediately in the parameter table. You can copy all parameter values together by clicking “Edit”-“Copy” menu item, and then load them into other similar analysis window by “Edit”-“Paste”.

For each type of parameters its own set of constraints can be introduced (see in 2.1.4 how to introduce and cancel constraints).The possible constraints for Segregation analysis are as follows:

Fixed- parameter has a constant value and should not change in maximization process. Constraint is applicable to all types of parameters.

Equal to other parameter – parameter is not maximized independently but together with one other. Constraint is applicable to genotypic values, their covariate dependence parameters and genotypic variances. It can be used to simulate dominance effect (equal values for heterozygote and homozygote of specific type). To state equality in genotypic values between sexes, you can introduce constraint for specific female parameter to be equal to corresponding male parameter.

Additive – constraint can be introduced for genotypic values, their covariate dependence parameters and genotypic variances, it is available in constraint menu for genotype 2 (A_1A_2- heterozygotes) and means that parameter is an average value between corresponding parameters for genotypes 1(A_1A_1) and 3(A_2A_2), different homozygotes’.

3.4.3. Trait expectation, predicted by the model.

When the maximization procedure is over or you have loaded some model into the analysis window from the results table (see 2.3.2), you can create the trait expectation, predicted by the model as separate trait column. When segregation analysis window is active, do click on “Edit”-“Add Trait”-“Prediction”-menu item. After the computation is over, you see a message box. The new column, containing trait expectation prediction, will be the last column in the document trait table. Its name will be “Pred+<number>”. You can rename it using “Properties”-item of the column pop-up menu. In the *Properties*-dialog of this new trait you can see all parameters of the model, trough which the expectation was estimated, and also genetic-statistical properties of the model. They are as follows. Components of phenotypic variance attributed to: Gen, Sex, Age (or other covariate), joint Gen-Sex, Gen-Age, Sex-Age, Gen-Sex-Age, Familial (trough relatives’ residual correlations), Residual and Total trait variance. In the second string of the table proportions of variance components to the total variance are given. Additionally you can see heritability H^2, proportion of variance described together with gene, sex and age factors (GSA), proportion of variance attributed to all model parameters (Total). The next are the estimated pairwise residual correlation coefficients between relatives: spouses, siblings and parent-offspring. Descriptive statistics usual to all quantitative traits is also available (see 1.7.1).

3.4.4. Most likely MG genotype, predicted by the model.

When the maximization procedure is over or you have loaded some model into the analysis window from the results table (see 2.3.2), you can create the new marker-trait containing most likely genotype set of major gene predicted for each pedigree of the sample by the model. The set having maximal likelihood is chosen for the whole pedigree rather than for each pedigree member. When segregation analysis window is active, do click on "Edit"-"Add Trait"-"Prediction GT"-menu item. After the computation is over, you see a message box. The new column, containing major gene most likely genotype prediction, will be the last column in the document trait table. In the properties-dialog of this new marker-trait you can see all parameters of the model, trough which the genotype set was chosen and also descriptive statistics usual to all marker traits (see 1.7.1).

3.4.5. Partial pedigree likelihood.

When the maximization procedure is over or you have loaded some model into the analysis window from the results table (see 2.3.2), you can create a new separate trait column, containing the partial pedigree likelihood (pedigree log-likelihood deleted on the number of measured pedigree members) corresponding the estimated model parameters. When segregation analysis window is active, do click on "Edit"-"Add Trait"-"Pedigree LH"-menu item. After the computation is over, you see a message box. The new column, containing partial pedigree log-LH, will be the last column in the document trait table. The estimated values are useful for selecting pedigrees maximally and minimally corresponding to selected model. When the sample heterogeneity is supposed, you can select more homogeneous in inheritance type pedigree sub-samples. You also can select pedigree having most positive or negative difference between LH for equal and Mendelian models.

3.5. Bi-Cross-sectional Segregation Analysis.

This variant of segregation analysis was designed to analyze inherited traits having repeated measurements. Parameters of the model are the same as for conventional analysis. The only additional parameter is ω - partial correlation between residuals of sequential measurements of an individual. Because this partial correlation in general should depend on the time difference between measurements, this simple variant of parameterization is more applicable when the difference in time between sequential measurements is similar for all individuals. This means that most suitable data are two cross-sectional measurements of the same population. Two repeated measurements should present two trait columns. If covariate dependence is assumed, both covariate measurements should be also different trait columns (for example Age1-of the first measurement and Age2-of the second). This analysis is more informative, if for example the age dependence of the trait is assumed to be inherited.

Click on "Analysis"-"Segregation"-"Bi-Cross-sectional" menu item to see the analysis window. The analysis window is similar to that of segregation analysis. Only in the TRAIT control group, you should specify two columns for sequential measurements of the trait and two columns for sequential covariate measurements.

4. Bivariate analysis.

A major gene model can be applied to a pair of inherited traits simultaneously. As a rule, the major gene model should be previously tested for each trait separately. If for both traits the model was accepted, we can test the hypothesis that both traits are under control of the same gene or by two linked genes, which are in linkage disequilibrium (LD) with each other.

The first variant of this analysis suppose that both traits are controlled by the same gene, but genotypic values, dominance and covariate dependence are specific for each trait (the most parsimonious model of inheritance may differ for the traits). So for this analyses we have a double parameter set of genotypic values, μ_{gs}, their age and sex dependence, a_{gs}, b_{gs} or T_{gs}, genotypic variances, σ_g, and partial correlations of trait residuals between specific relatives, $\{\rho, \beta, \varepsilon\}$. The shared by two traits parameters are the allele frequency, p (for Hardy-Weinberg equilibrium), or genotype frequencies, $\{p_1, p_{2\}}$, the partial self correlation between trait values of one individual, ω, and transmission probabilities τ_g. We can estimate transmission probabilities and construct transmission probability (TP) test $\{\lambda_1 , \lambda_2 \}$ as described in 3.2. Some times, the MG effect for one of two traits is so significant, that bivarite LRTs support the hypothesis of same MG for two traits. Therefore, if the MG effect of the same gene was accepted in $\{\lambda_1, \lambda_2\}$, we should additionally test two following parameter constraints for Mendelian model. Let's do the genotype specific parameters (μ_g, a_g, b_g ,T_g) equal for all genotypes ($\mu_1 =\mu_2 =\mu_3$ and so on) formerly only for parameter set of the first trait and then in turn only for the second. The Mendelian model tested in λ_1 (having some different parameters for different genotypes simultaneously in parameter sets for both traits) should be significantly better than each model of these two additional is.

For the second implementation (LD-model) of bivariate analysis the shared parameters are: frequencies of A_1 alleles of two linked genes, p and q, partial self correlation, ω, recombination fraction, θ, and Lewontin's disequilibrium parameter δ.

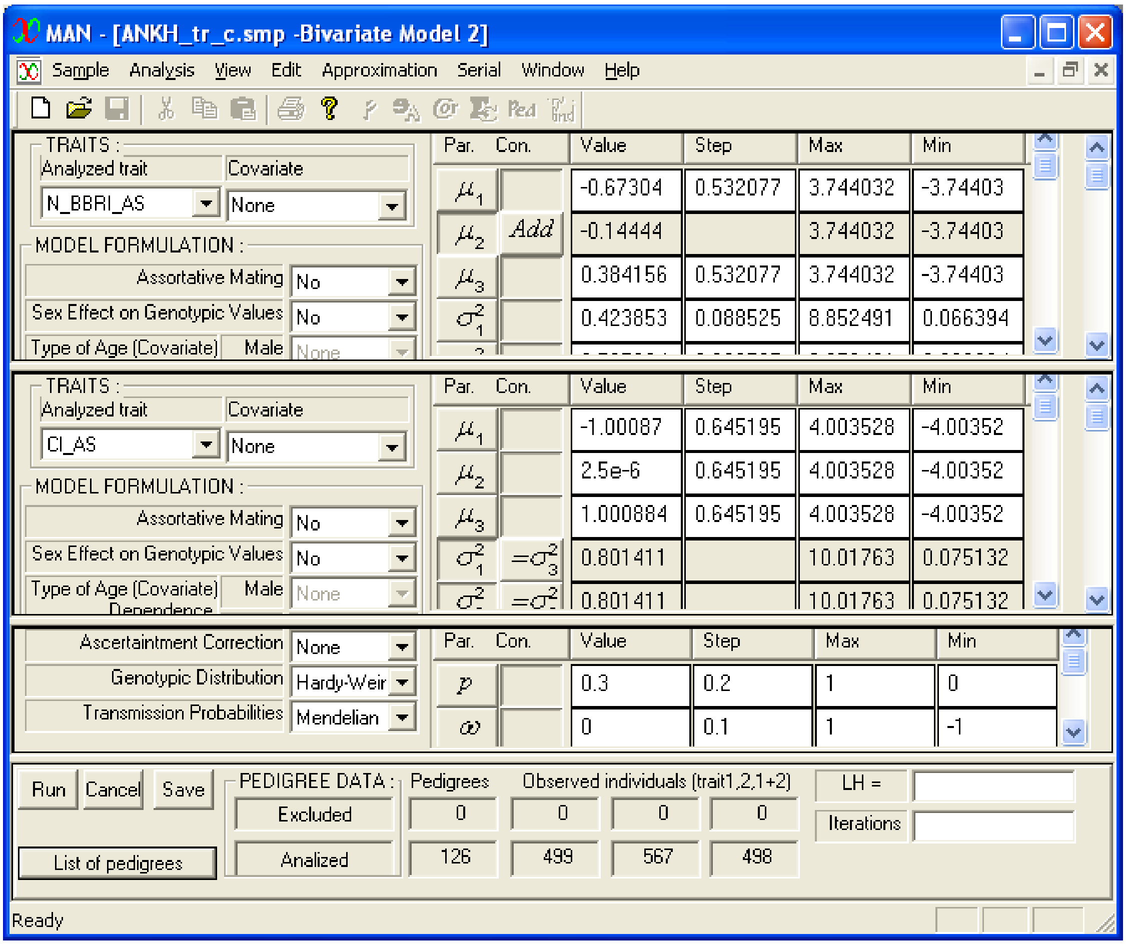
This model can be compared with "Hardy-Weinberg" "Mendelian" model of the first implementation, corresponding to θ=0, δ=1, q=p. The null hypothesis, which we should reject is no linkage (θ=0.5) or no disequilibrium (δ=0) between to major genes. The "Hardy-Weinberg" "Mendelian" model, corresponding to θ=0, δ=1, q=p, can be compared by LRT with LD-model having free θ, δ, q, p parameters. This test the hypothesis that two gene model is significantly better than single gene.

4.1. User interface for bivariate analysis.

Click on "Analysis"-"Bivariate"-menu item to open the analysis window. Only one bivariate analysis window can be opened at a time. The window consists of three sub-windows with moving horizontal splitters between them. The two upper sub-windows are similar to window of segregation analysis. Both of them are intended for model formulation, parameters and constraints for two analyzing traits separately. For each trait, its own covariate or no covariate can be defined. Model formulation options: sex dependence, assortative mating and type of covariate dependence - can be chosen independently for both traits (as described in 3.4.1).

The rest of model parameters relating to both traits are presented in the third sub-window. Here are also the rest of model formulation options. They are used as described in 3.4.1, except for **Genotypic Distribution** (GD) option, which differs slightly from described in 3.4.1 section.

The two first options in GD-combobox: "Hardy-Weinberg" and "Arbitrary" are the same as in 3.41; if one of them is chosen the **Transmission Probabilities**(TP) – combobox is enabled.



TRAITS :

Analyzed trait	Covariate
N_BBRI_AS	None

MODEL FORMULATION :

Assortative Mating: No

Sex Effect on Genotypic Values: No

Type of Age (Covariate) Male: None

Par.	Con.	Value	Step	Max	Min
μ_1		-0.67304	0.532077	3.744032	-3.74403
μ_2	Add	-0.14444		3.744032	-3.74403
μ_3		0.384156	0.532077	3.744032	-3.74403
σ_1^2		0.423853	0.088525	8.852491	0.066394

TRAITS :

Analyzed trait	Covariate
CI_AS	None

MODEL FORMULATION :

Assortative Mating: No

Sex Effect on Genotypic Values: No

Type of Age (Covariate) Dependence Male: None

Par.	Con.	Value	Step	Max	Min
μ_1		-1.00087	0.645195	4.003528	-4.00352
μ_2		2.5e-6	0.645195	4.003528	-4.00352
μ_3		1.000884	0.645195	4.003528	-4.00352
σ_1^2	$=\sigma_3^2$	0.801411		10.01763	0.075132
σ_2^2	$=\sigma_2^2$	0.801411		10.01763	0.075132

Ascertainment Correction: None

Genotypic Distribution: Hardy-Weir

Transmission Probabilities: Mendelian

Par.	Con.	Value	Step	Max	Min
p		0.3	0.2	1	0
ω		0	0.1	1	-1

PEDIGREE DATA :

	Pedigrees	Observed individuals (trait1,2,1+2)		
Excluded	0	0	0	0
Analized	126	499	567	498

LH =

Iterations

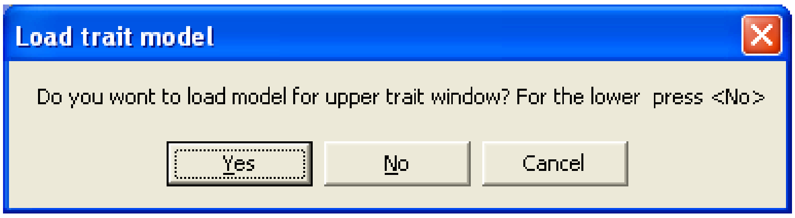
Ready

The “L-Disequilibrium” option of GD-combobox describe the model where each trait has its own major gene with allele frequencies *p* and *q* for traits defined in uppermost and middle sub-windows correspondingly. The TP-combobox is disabled, but two genes are supposed to be in linkage disequilibrium. The recombination fraction,θ, and Lewontin’s disequilibrium parameter δ between to major genes appear in the lowest parameter table and can be estimated. All these parameters can be constrained to constant, additionally *q*=*p* constraint can be introduced.

4.1.1. Using results of single trait segregation analysis in bivariate.

If one or both traits, which you are going to investigate in bivariate analysis, were formerly tested in segregation analysis and most parsimonious (MP) model was built, you can load it in bivariate analysis window with its parameter values as initial for bivariate maximization. To do this, open the appropriate result table by double click on it (2.3). Select by mouse click the column with desirable model and click “Analysis”-“Bivariate” menu item.

If no Bivariate analysis window was already opened, it opens with the selected segregation trait model in the upper sub-window. The trait in the lower sub-window will be selected casually. To load the second trait with its MP model, open the corresponding table and select the desirable column. Then click “Analysis”-“Bivariate” menu item. The Yes-No- message box with the question appear, which enables to determine upper (Yes) or lower (No) sub-window. Click on Yes or No to load the selected trait and model into the appropriate sub-window.



Load trait model

Do you wont to load model for upper trait window? For the lower press <No>

5. Variance component (VC) analysis.

This analysis assumes that the individual trait value is a function of covariates plus a random trait residual partially correlated between relatives.

$$x_i = \varphi(y_{i1}, y_{i2},.., y_{iK}) + \xi_i,$$

where x_i is the individual trait value, $y_{i1}, y_{i2},.., y_{iK}$ – individuals covariate values and ξ_i - residual.

For a single integer covariate (the main example is age in years), four types of covariate dependence $\varphi(y)$ are available: *linear* (parameters: μ-mean trait value, a –linear coefficients); *quadratic* (parameters: μ, a, b – second power coefficient); two-interval *linear-constant* (μ, a-linear slope, T- breaking point); two-interval *constant-linear* (μ, a, T) (see 3.1.4 for more detailed parameter description). The type of covariate dependence can be introduced for the whole sample (without sex effect) or for each sex separately. The type of dependence can be different for males and females.

For multiple covariates (real or integer) linear multiple regression model is estimated. You can select up to ten covariates.

$$\varphi(y_{i1}, y_{i2},.., y_{iK}) = \alpha_0+\beta_1\cdot y_{i1}+\beta_2\cdot y_{i2}+\ldots+\beta_K\cdot y_{iK}$$

Parameters are α_0 - intercept, β_1, β_2,…, β_K - regression coefficients, where K is the number of covariates. The multiple regression coefficients can be estimated for the whole sample(without sex effect) or for each sex separately (α_{0m}, α_{0f}, β_{1m}, β_{1f}, β_{2m}, β_{2f}, …, $\beta_{Km,}\beta_{Km}$).

The trait residual ξ in a pedigree is assumed to have n-variable normal distribution, where n is a number of pedigree member. The components of variance-covariance matrix, determining this distribution, we parameterize as follows. Let assume, that the trait residual is a result of the simultaneous influence of a number of orthogonal factors genetic or environmental. Genetic variance is assumed to be a result of joint effect of a great number of genes, each having small input. We should distinguish additive and dominant genetic variance. The shared by two relatives additive genetic variance σ^2_{AD} is approximately proportional to the number of coinciding gene alleles: the full σ^2_{AD} is shared only for mono-zygotic twins, 0.5 σ^2_{AD} for parent-offspring and siblings or di-zygotic twins, 0.25 σ^2_{AD} for half-sibs and uncle-nephew and so on. The dominant genetic variance σ^2_{DO} shared by two relatives is approximately proportional to the number of coinciding genotypes (combinations of two alleles); it is assumed to be nonzero only for monozygotic twins (σ^2_{DO}) and for siblings or di-zygotic twins ($0.25\sigma^2_{DO}$). For each environmental factor we define the group of pedigree members, sharing it. Possible environment components are: σ^2_{HS} – household component, shared by members of the same nuclear family; σ^2_{SP} additional environment component shared by spouses; σ^2_{SB} additional component shared by siblings; σ^2_{TW} additional environment factors shared by twins; σ^2_{RS} residual individual variance. Because we assume that, all factors are orthogonal to each other, the covariance matrix component for each pair of individuals is a sum of genetic and environment effect variances shared by the relatives i and k. The matrix diagonal components (total individual variances) can be modeled in two ways.

The first way, all diagonal elements are equal to each other. This means that the total value of individual variance does not depend on its position in the pedigree. For this case, we cannot distinguish between genetic dominant component and sibling specific environment effects (if there are no monozygotic twins). Therefore, σ^2_{DO} is not estimated separately. Sibling specific variance σ^2_{SB} may include both environmental and dominant genetic effects.

The second way, diagonal elements are a sum of individual genetic and residual variances (these components are the same for all pedigree members) plus variances of environmental effects, which the individual shares with his relatives. For example, if individual has no siblings, the sibling specific variance σ^2_{SB} is not included in his total variance. In this case, dominant genetic variance σ^2_{DO} can be estimated independently from sibling specific environmental variance.

Both described above variants of model formulation are available for computation.

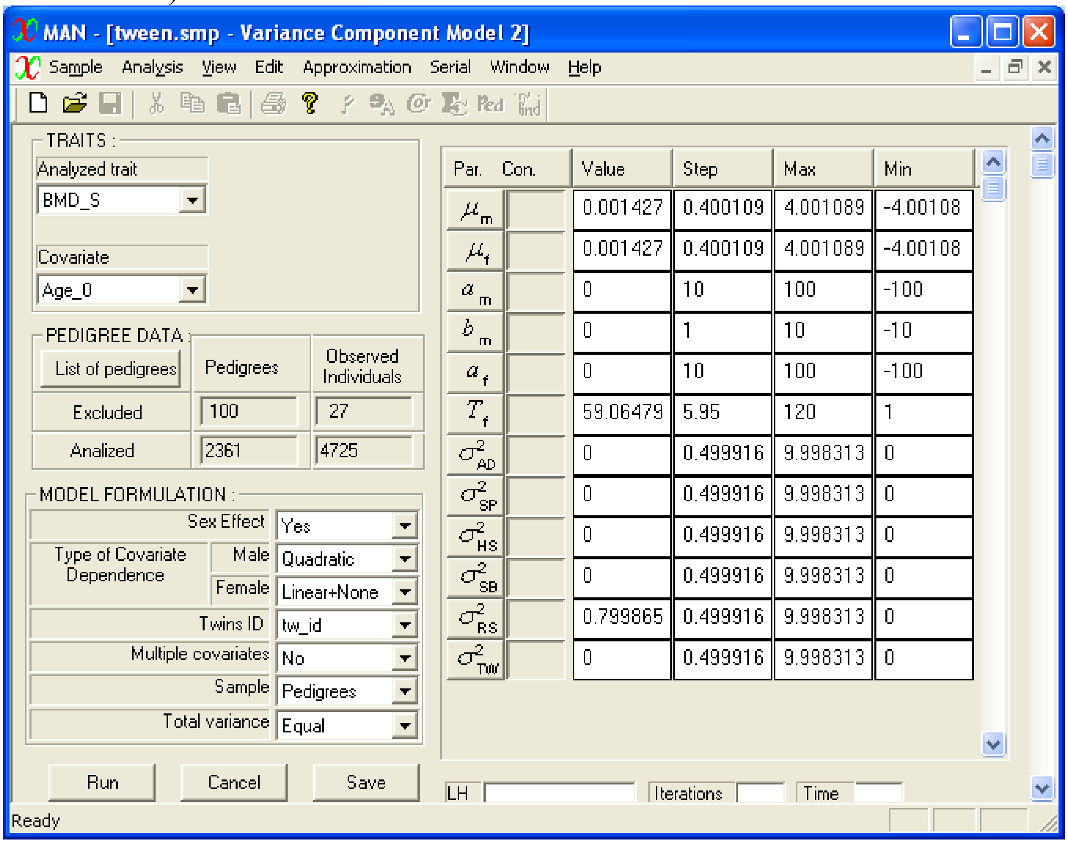
The purpose of analysis is to find the most parsimonious model by means of LRT, comparing general and constraint models. Then, you can estimate the input of significant covariates and significant components of residual variance.

5.1. User interface for variance component analysis.

Click on "Analysis"-"Variance Component"-"Univariate"-menu item to open the analysis window.

5.1.1. Options of model formulation group.

Sex effect. By default "No". The model parameters characterizing the non-random part of the trait are the same for males and females. If "Yes", two different sets of parameters (for males and females) are estimated.


MAN - [tween.smp - Variance Component Model 2]
Sample Analysis View Edit Approximation Serial Window Help
Run
Cancel
Save
LH
Iterations
Time
Ready

TRAITS :
Analyzed trait: BMD_S
Covariate: Age_0

PEDIGREE DATA :

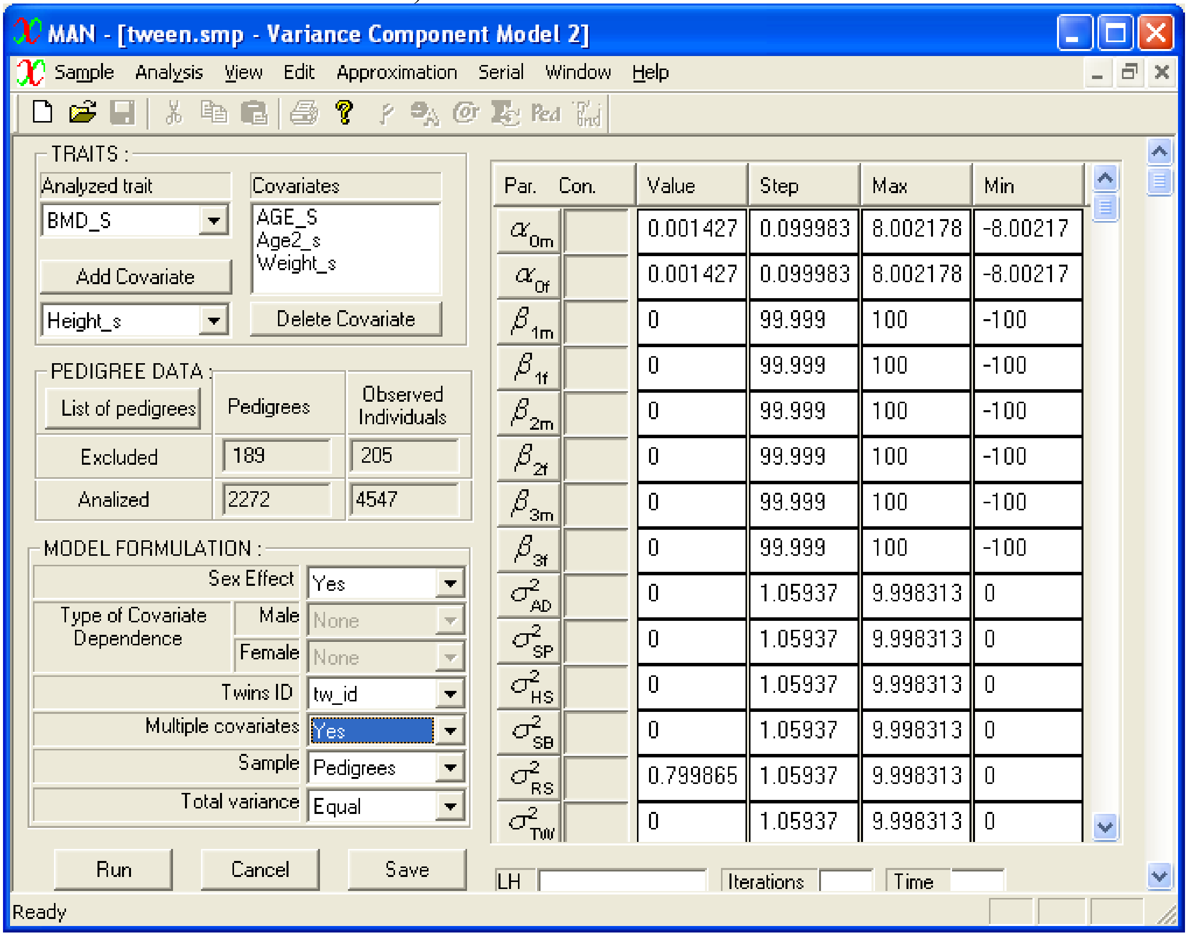
List of pedigrees	Pedigrees	Observed Individuals
Excluded	100	27
Analized	2361	4725

MODEL FORMULATION :
Sex Effect: Yes
Type of Covariate Dependence: Male: Quadratic; Female: Linear+None
Twins ID: tw_id
Multiple covariates: No
Sample: Pedigrees
Total variance: Equal

Par.	Con.	Value	Step	Max	Min
μ_m		0.001427	0.400109	4.001089	-4.00108
μ_f		0.001427	0.400109	4.001089	-4.00108
a_m		0	10	100	-100
b_m		0	1	10	-10
a_f		0	10	100	-100
T_f		59.06479	5.95	120	1
σ^2_{AD}		0	0.499916	9.998313	0
σ^2_{SP}		0	0.499916	9.998313	0
σ^2_{HS}		0	0.499916	9.998313	0
σ^2_{SB}		0	0.499916	9.998313	0
σ^2_{RS}		0.799865	0.499916	9.998313	0
σ^2_{TW}		0	0.499916	9.998313	0

Multiple covariates. By default “No”. In the TRAITS group only one covariate can be selected in the Covariate-combobox from the list of integer traits available in the sample.
If an actual covariate was selected, the “type of covariate dependence” can be chosen.

If “Yes”, in TRAITS group appears an empty “Covariate”-list box and two additional buttons “Add covariate” and “Delete covariate” to enable the formation of covariate list (more details see below in 5.1.2 section).


MAN - [tween.smp - Variance Component Model 2]
Sample Analysis View Edit Approximation Serial Window Help
Run
Cancel
Save
Ready

TRAITS :

Analyzed trait	Covariates
BMD_S	AGE_S Age2_s Weight_s
Add Covariate	
Height_s	Delete Covariate

PEDIGREE DATA :

List of pedigrees	Pedigrees	Observed Individuals
Excluded	189	205
Analized	2272	4547

MODEL FORMULATION :

Sex Effect		Yes
Type of Covariate Dependence	Male	None
	Female	None
Twins ID		tw_id
Multiple covariates		Yes
Sample		Pedigrees
Total variance		Equal

Par.	Con.	Value	Step	Max	Min
α_{0m}		0.001427	0.099983	8.002178	-8.00217
α_{0f}		0.001427	0.099983	8.002178	-8.00217
β_{1m}		0	99.999	100	-100
β_{1f}		0	99.999	100	-100
β_{2m}		0	99.999	100	-100
β_{2f}		0	99.999	100	-100
β_{3m}		0	99.999	100	-100
β_{3f}		0	99.999	100	-100
σ^2_{AD}		0	1.05937	9.998313	0
σ^2_{SP}		0	1.05937	9.998313	0
σ^2_{HS}		0	1.05937	9.998313	0
σ^2_{SB}		0	1.05937	9.998313	0
σ^2_{RS}		0.799865	1.05937	9.998313	0
σ^2_{TW}		0	1.05937	9.998313	0

LH   Iterations   Time

Type of covariate dependence. The option is enabled only for one integer covariate regime (“Multiple covariate”- combobox set to “No”). If “sex effect”- combobox was set to “No”, the type of covariate dependence selected for male is applied to both sexes. If “sex effect” was set to “Yes”, the “female”-combobox is also enabled.

Sample. By default - “Pedigrees”. The model will be estimated on the sample containing complex pedigrees. If “Nucl.Fam.”(nuclear families) was selected. The analysis is done on the sample of nuclear families. Each complex pedigree is divided into nuclear families. So complex pedigree members, who are members, for example, of two nuclear families appear in the sample twice. This variant is less precise, but was included to compare results sensitivity to complex pedigree structure and non-independence of nuclear families.

If "Incl.Singles" is selected the model will be estimated on pedigrees, but pedigrees containing only one measured individual are not excluded from the analysis. This option should be used only when significance of regression on the covariate is tested.

Total variance. By default “equal”. The variance-covariance matrix has equal diagonal elements. If “Additive effects” was selected the diagonal elements are sum of individual components (equal by value for all) and environment components, which individual shares with its specific relatives (see the last paragraph of section 5).

Twins ID. By default “none” (twins are not identified). If the sample contains twins, you should identify them with some integer variable, which for each twin set should have the same value and should be unique for each separate set within nuclear family. The monozygotic twins should have positive ID, the dizygotic twins – negative. The simplest way to create such a variable is to give positive integer age values to monozygotic and negative age values to dizygotic twins, because age should be equal and unique in the nuclear family for each twin set (you can use “Edit”-“Add trait”-“Formula” menu item to create such trait, see 1.6.1). Choose this specific variable from the list of all integer variables, which are available in “Twins ID”-combobox to account for twins.

5.1.2. Selection of covariates in TRAIT control group.

If “Multiple covariates”- combobox is set to “No”, the covariate is selected from the list of all integer traits of the sample. This list is available in “Covariate”- combobox. As a rule, the single integer covariate used for this analysis is age. For many human traits, the age dependence is strong and often nonlinear. For this reason four type of covariate dependence described in 3.1.4 and 5 are available and can be chosen in the “Model formulation” group.

If “Multiple covariates”- combobox is set to “Yes”, the appearance of TRAITS-control group changes. The empty listbox, named “Covariates”, appear in the right side of the control group. Two buttons “Add covariate” and “Delete covariate” appear to enable the formation of the covariate list. In the combobox below the “Add covariate”-button all quantitative traits of the sample (real and integer) are available.

To add the trait to covariates list, select the desirable trait in the combobox and click the “Add covariate”-button. The selected trait moves to the Covariates-listbox. Simultaneously the new parameters α_0-intercept and β_1 appear in the parameter table. When you select and add the next covariate, the additional β-coefficient appear id the parameter table. To delete some covariate from the covariate list, click on it by the mouse cursor in the Covariates-listbox. When the trait to delete will be selected (blue string in the listbox), click the “Delete covariate” button.

5.1.3. Possible constraints in the parameter table and its special features.

Constraints can be introduced for a model as described in 2.1.4. All parameters can be constrained to fixed value. When parameters of covariate dependence are sex specific, each of female parameters can be constrained to be equal to male parameter.

If you use multiple covariates, coefficient β_1 corresponds to the upper covariate in the “Covariates”-listbox, β_2 – to the next covariate and so on (covariates in the alphabetic order).

Click on the “Step”-header in the parameter table increases all steps to enable the next maximization.

5.1.4. Additional option for trait adjustment, using VC-model.

If the model includes non-trivial covariate dependence, you can save the trait residual (trait adjusted for covariates accounting for pedigree structure) as new separate trait column. When the window of variance component analysis is active, the click on “Edit”-“Add trait”-“Adjusted” menu item produces the new trait column containing appropriate trait residuals. The adjusted trait appears as last column in the trait table of the document window. Its name will be “Adjustment+<number>”. You can rename it using “Properties”-item of the column pop-up menu. In the properties- dialog of this new trait you can see the name of trait, which was adjusted, the list of covariates and all parameters of the model, trough that the residual values were estimated.

5.2. Notes for performing variance component analysis.

5.2.1. Standardized traits and covariates.

It is a good idea, to standardize the analyzed trait and all multiple covariates before the analysis. In this case the maximization procedure don’t have difficulties with very small and very great values, which cannot pass the intermediate maximization limitation conditions and can likely result in maximization ending with error message.

Additionally, if you want to know the proportion of different variance components in the total trait variance you should divide them on the total trait variance. When the trait was previously standardized, the estimated components show directly proportions of the total variance.

If all covariates, taking part in multiple linear regression model, are standardized, the estimated β coefficients coincide with appropriate Wilkinson Betas’ and can be compared by value.

5.2.2. Using VC-analysis for testing disequilibrium between trait and genetic markers.

To do this you should produce for each genetic marker a new trait containing the number of copies of selected allele in the genotype. If the marker has more then two alleles you can factorize it as described in 1.6.5 and then using “Edit”-“Add trait”-“Formula”-option (1.6.1) you can create new trait containing the desirable trait. For example, if factorized marker name is “MarF” and its alleles are “1” and “2”, you can create the trait using formula

if(MarF=”1/1”, 2, if(MarF=”2/2”,0,1))

For dialelic markers you can also use “Edit”-“Add trait”-" Number of Minor Allele" menu item to produce the desired integer trait. This new trait you can use as covariate for VC analysis and test the significance of its β coefficient by LRT. To test for dominant effect, you can use additional variable, which is non-zero only for heterozygous individuals. You can try to test, which factorizations of the multi-allelic marker influence the trait significantly, using all factorizations together in multiple regression model. Analogously, different transformed markers can be used as covariates simultaneously.

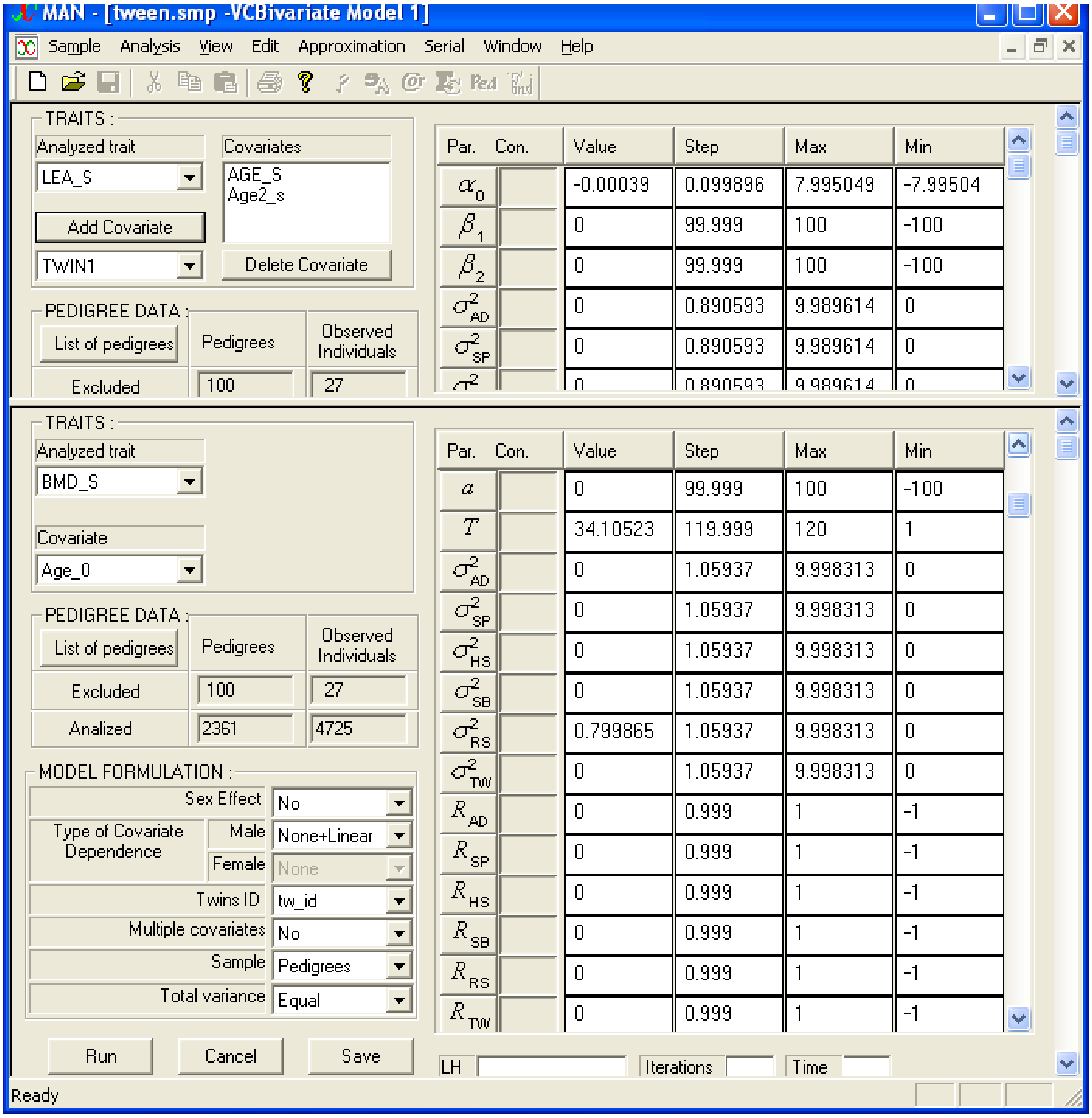
5.2.3. Computation of parameter standard errors.

To compute parameter standard errors (SE) you should save the model maximization results as column in the results tables. The procedure is described in section 2.3, paragraph *Error Function.* It is preferably to compute parameters SE for most parsimonious model, because otherwise (for many non-significant model parameters) the likelihood maximum is not expressed sufficiently for some coordinates to enable estimation of SE.

5.3. Bivariate variance component analysis.

If we are interested in how genetic and environment variance components of two different traits are correlated, we can use bivariate VC analysis. The model formulation is independent for each trait and its covariates. The only shared model formulation options are *Sample* (pedigrees or nuclear families) and *Total variance* (way of modeling of covariance matrix diagonal).

For each trait the same set of parameters as in univariate analysis is to be estimated. The additional parameters are the components correlation coefficients: R_{HS} – for household components; R_{SP} – for spouses’ specific environments; R_{SB} –for siblings’ specific components; R_{TW} – for twins’ specific environments; R_{RS} –for individual residuals of two traits. The significance of correlation can be tested with LRT.



TRAITS :

Analyzed trait	Covariates
LEA_S	AGE_S Age2_s
TWIN1	

PEDIGREE DATA :

	Pedigrees	Observed Individuals
Excluded	100	27

Par.	Con.	Value	Step	Max	Min
α_0		-0.00039	0.099896	7.995049	-7.99504
β_1		0	99.999	100	-100
β_2		0	99.999	100	-100
σ^2_{AD}		0	0.890593	9.989614	0
σ^2_{SP}		0	0.890593	9.989614	0
σ^2		0	0.890593	9.989614	0

TRAITS :

Analyzed trait: BMD_S

Covariate: Age_0

PEDIGREE DATA :

	Pedigrees	Observed Individuals
Excluded	100	27
Analized	2361	4725

MODEL FORMULATION :

Sex Effect		No
Type of Covariate Dependence	Male	None+Linear
	Female	None
Twins ID		tw_id
Multiple covariates		No
Sample		Pedigrees
Total variance		Equal

Par.	Con.	Value	Step	Max	Min
α		0	99.999	100	-100
T		34.10523	119.999	120	1
σ^2_{AD}		0	1.05937	9.998313	0
σ^2_{SP}		0	1.05937	9.998313	0
σ^2_{HS}		0	1.05937	9.998313	0
σ^2_{SB}		0	1.05937	9.998313	0
σ^2_{RS}		0.799865	1.05937	9.998313	0
σ^2_{TW}		0	1.05937	9.998313	0
R_{AD}		0	0.999	1	-1
R_{SP}		0	0.999	1	-1
R_{HS}		0	0.999	1	-1
R_{SB}		0	0.999	1	-1
R_{RS}		0	0.999	1	-1
R_{TW}		0	0.999	1	-1

LH | Iterations | Time

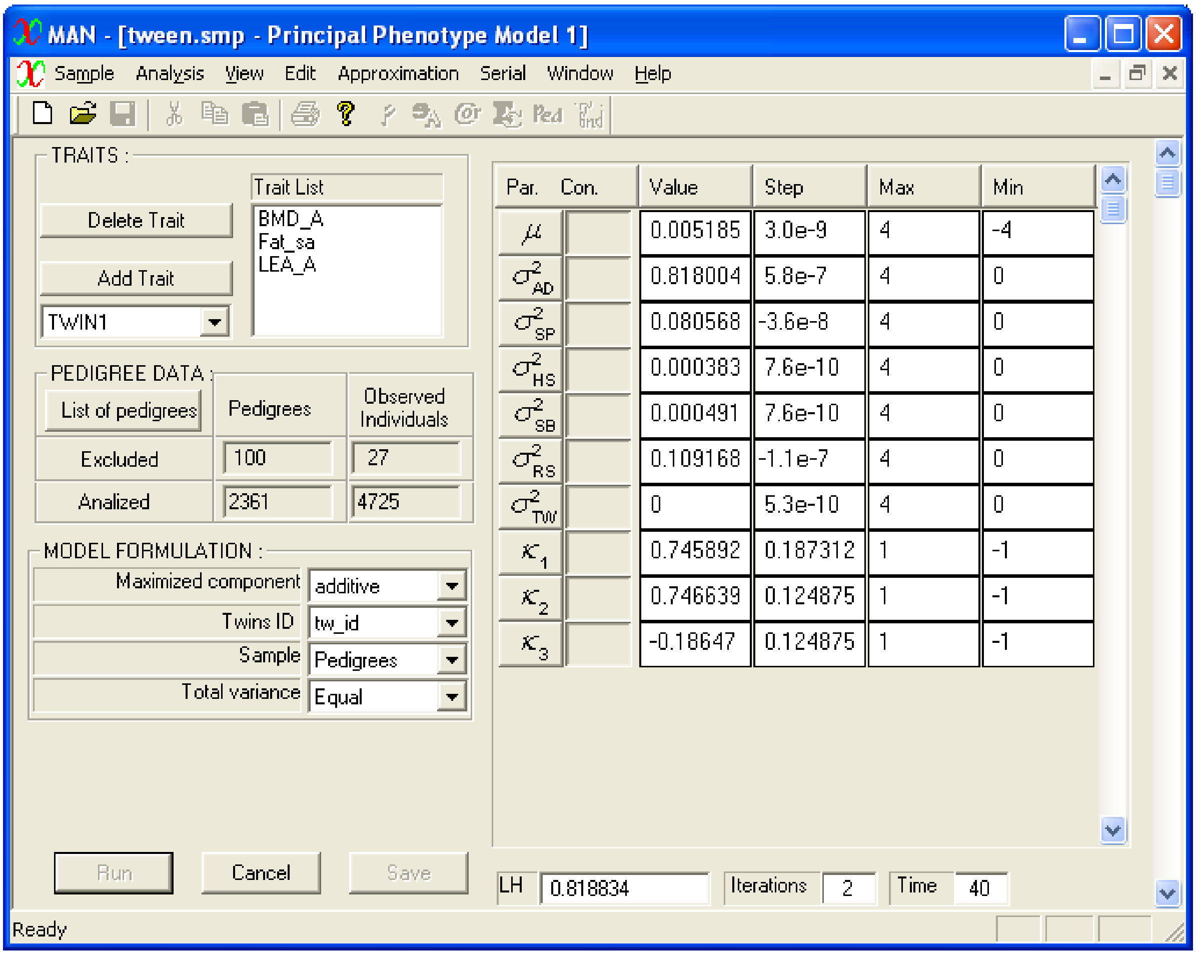
5.3.1. User interface for bivariate variance component analysis.

Click “Analysis”-“Variance Component”-“Bivariate” menu item to open the window of bivariate VC analysis. The window consists of two sub-windows similar to those of univariate VC analysis. In each sub-window you should select trait, its covariates, independent model formulation options and parameter constraints. The three shared model formulation options for two traits are *Twin ID* (the trait which identifies twin sets or “None”), *Sample* (pedigrees or nuclear families) and *Total variance* (way of modeling of covariance matrix diagonal, see 5.1.1). The appropriate comboboxes are presented only in the lower sub-window. The specific for bivariate analysis correlation parameters R are presented in the parameter table of lower sub-window. The computation procedure initially estimates model parameters and their SE for each trait separately and then uses these parameter values to estimate correlation coefficients and joint likelihood. It is preferable to use most parsimonious model for each trait. You can load this model from the result table of an univariate analysis. To do this, open the appropriate table, select the column containing the most parsimonious model and click “Analysis”-“Variance Component”-“Bivariate” menu item. If the bivariate analysis window was not opened previously, it opens with the selected trait model in the upper sub-window. If the bivariate analysis window was already opened, you will be asked to what sub-window (upper or lower) the selected model should be loaded. Only one bivariate VC analysis window exists at a time. When you run the maximization, the standard errors of parameters are estimated automatically. After saving results, you can see both parameters and their SE in the result table. For each bivariate maximization, the result table contains three columns. The first two include parameters of univariate trait models with univariate log-likelihoods' below them. The third column contains correlation parameters and the joint bivariate log-likelihood. This log-likelihood should be used for testing significance of component correlation coefficients.

5.4. Principal phenotype.

This analysis was named so analogously to principal component analysis, which designs factors being linear combinations of initial traits and meeting specific maximization conditions. The purpose of *Principal phenotype* analysis is to built the trait (phenotype), being linear combination of a number of initial traits in such a way, that it should have as maximal as possible proportion of definite variance component - additive genetic, or sibling specific, or spouses specific, or so on. For example if you have a number of traits, measured on different human body sites, characterizing adiposity, you can try to build mostly heritable trait characterizing the adiposity in total. It is also possible to try to combine in such a way traits of different nature, sharing probably definite part of genetic or environment factors.

5.4.1. User interface for principal phenotype analysis.



Trait List: BMD_A
Fat_sa
LEA_A

TWIN1

	Pedigrees	Observed Individuals
Excluded	100	27
Analized	2361	4725

Maximized component: additive

Twins ID: tw_id

Sample: Pedigrees

Total variance: Equal

Par.	Con.	Value	Step	Max	Min
μ		0.005185	3.0e-9	4	-4
σ^2_{AD}		0.818004	5.8e-7	4	0
σ^2_{SP}		0.080568	-3.6e-8	4	0
σ^2_{HS}		0.000383	7.6e-10	4	0
σ^2_{SB}		0.000491	7.6e-10	4	0
σ^2_{RS}		0.109168	-1.1e-7	4	0
σ^2_{TW}		0	5.3e-10	4	0
κ_1		0.745892	0.187312	1	-1
κ_2		0.746639	0.124875	1	-1
κ_3		-0.18647	0.124875	1	-1

LH 0.818834 Iterations 2 Time 40

Analysis enables to combine up to 10 traits. You should create the trait list for the analysis in the TRAIT control group by means of “Add trait” and “Delete trait” buttons. The “Trait list”- listbox is placed in the right side of TRAIT control group. To add the trait, select its name in the combobox in the left side of the control group and click the “Add trait”-button. To delete the trait from the list, select it in the listbox and click “Delete trait”-button.

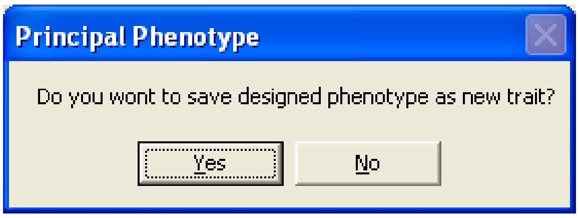
It is supposed, that traits, which you try to unite, are residuals, already adjusted for sex, age and other covariates. Therefore no sex effect and type of adjustment options are presented in the MODEL FORMULATION group. Options *Twins ID, Sample* and *Total variance* have the same meaning as in 5.1.1. The *Maximized component* – combobox defines what variance component you are going to maximize.

In the parameter table you can see the coefficients κ_1, κ_2, …,κ_N; N – is the number of traits in the trait list. The coefficients define how the standardized resulting phenotype can be built from the traits, presented in the list, after standardization. Constraint (fixed value) can be introduced for mean trait value μ and all available components of variance, besides the component, which was selected as *maximized component*. Coefficients κ_i can not be fixed.

The iterations for this type of analysis are more time consuming than for VC analysis, because each iteration is indeed the full VC analysis for the trait formed using current set of scores κ_i. When the maximization is in process, the editbox “Time” below the parameter table shows the time from the beginning of analysis in seconds.

5.4.2. Saving and viewing results of the analysis.

When the maximization is over, you can click the “Save”-button and save the results in a table. For this type of analysis, the maximization is saved as separate table. When you have specified the name of the table in the “Save”-dialog and click “Save”-button, the program suggests saving of the designed principal phenotype as new separate trait column. If you click “Yes”, the trait named “Principal”+<number> appears as the last column in the trait table of the document. You can rename it using “Properties”-item of the column pop-up menu. In the properties- dialog of the new trait you can see what component was maximized, the names of initial traits, their scores κ_i and variance components of resulting trait.



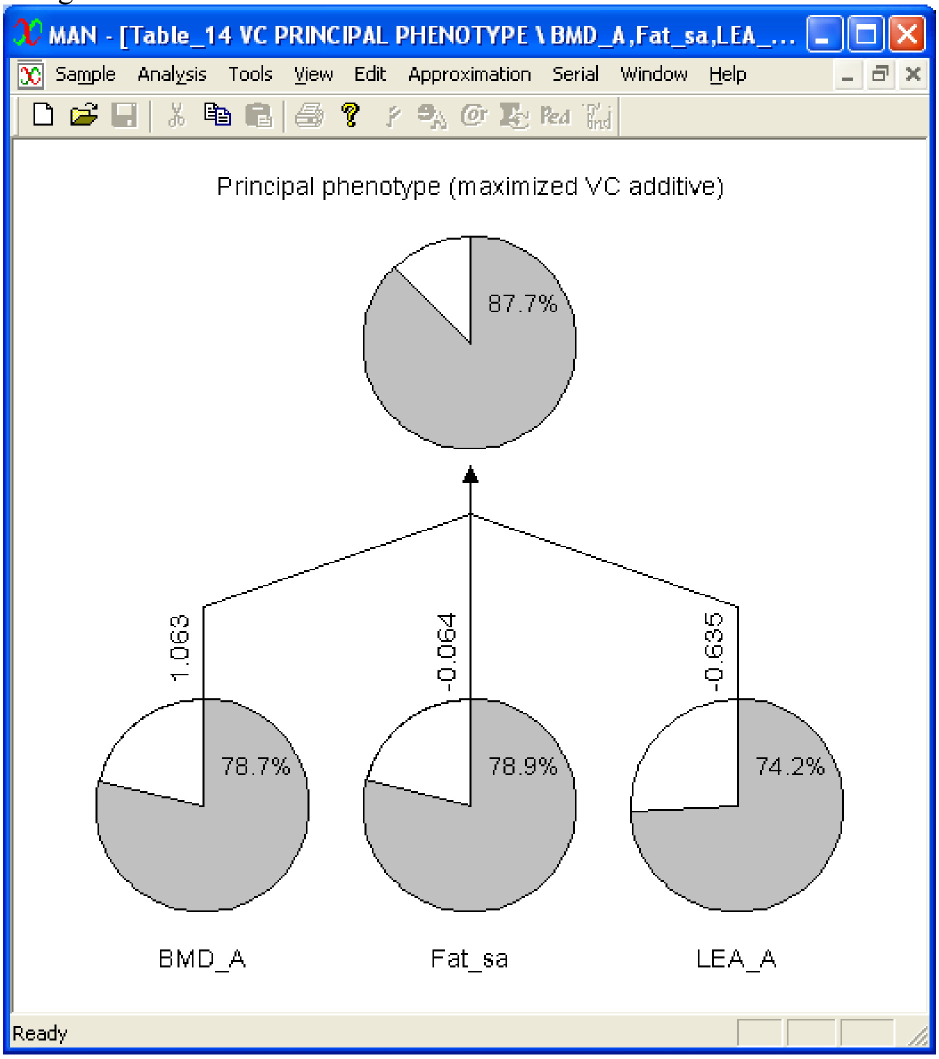
Principal Phenotype

Do you wont to save designed phenotype as new trait?

The table, saving analysis results, appears in the document tree in the RESULTS OF ANALYSIS item after child item, named as the first initial trait in the analysis trait list. In the document tree after the specified table name, you can see the full trait list, delimited by commas. Click on the table icon in the document tree to see its contents. The table includes separate column for each initial trait, where the VC analysis results for the trait after standardization are presented. The last column in the table, named “Principal” contains VC analysis of principal phenotype and trait scores κ_i.

5.4.3. The diagram of principal phenotype construction.

When the results table window of principal phenotype analysis is active, to view the results graphically, click on the “Tools”-“Diagram”-menu item. The diagram window appears. The variance component, which was maximized, is shown in gray for initial traits and for resulting phenotype. Use “Edit”-“Copy”- menu item to put the diagram into the clipboard as enhanced metafile. Then you can use the “Paste” option of Microsoft Office programs to include the diagram in their documents.


MAN - [Table_14 VC PRINCIPAL PHENOTYPE \ BMD_A,Fat_sa,LEA_...
Sample Analysis Tools View Edit Approximation Serial Window Help
Principal phenotype (maximized VC additive)
87.7%
1.063
-0.064
-0.635
78.7%
78.9%
74.2%
BMD_A
Fat_sa
LEA_A
Ready

5.5. Quasi Variance Component Analysis.

VCA was originally proposed for quantitative continuous traits. The approach can be extended to the analysis of categorical traits, namely proposed by Falconer liability X, an unobserved, continuously distributed, quantitative variable with a threshold τ, dividing the population to affected (X>τ) and non-affected (X<=τ) individuals.

If the quantitative trait $\{X_n\}$ is the hypothetic liability with an affection threshold τ relating to a dichotomous affection status $\{B_n\}$, measured on N pedigree members, the estimation of model parameters demand computation of N-dimensional integral of the probability density LH of the liability, on each variable X_n from $-\infty$ to τ, if $B_n=0$, and from τ to $+\infty$, if $B_n=1$. This is a complex task and is the major reason why VC estimations of disease liabilities have been performed only on samples with simple familial structures including MZ and DZ twin pairs and without the inclusion of covariates. Without covariates the two-dimensional integral could be computed numerically for each parameter set only for three co-affection variants of co-twins ([0, 0], [0, 1], [1, 1]). If covariates are included, the LH for each twin pair in the sample should be integrated separately and the computation time grows unlimitedly with sample size.

We have reformulated the VCA LH for quantitative traits in order to to simplify the integration of probability density LH on variables X_n and in this way to enable the inclusion of several covariates in the analysis of a liability.

Consider a pedigree with two measured individuals $\{X_{1k}, X_{2k}\}$, both with covariates $\{Y_{1k}, Y_{2k}\}$ (where k is the number of the pedigree). The variance of X in the sample is σ_{Tot}^2. To fit the data, we will use a VC model that includes only one variance component σ_C^2 shared by both relatives (which may reflect any type of effect, genetic, environmental, or a combination). Trait value X can be considered as the sum of independent (orthogonal) traits T_j (factors): $X_{ik}=T_{1\ ik}+T_{2\ ik}+T_{3\ ik}= a_k + F(Y_{ik}) + s_{ik}$; $T_{1\ ik} = a_k$ is a factor specific for the pedigree. Both individuals belonging to the same pedigree k have the same value $T_{1\ 1k} = T_{1\ 2k} = a_k$. The variance of the factor T_1 in the sample is σ_C^2. $T_{2\ ik} = F(Y_{ik})$ - is the factor presenting the individual trait value, predicted by covariates. The variance of factor T_2 in the sample can be estimated for any function $F(Y_{ik})$ and is σ_{Covar}^2. $T_{3\ ik} = s_{ik}$ is also individual specific $s_{ik} = X_{ik} - T_{1\ ik} - T_{2\ ik}$. Because all factors are supposed to be orthogonal the variance of s_{ik} is $\sigma_R^2=\sigma_{Tot}^2-\sigma_{Covar}^2-\sigma_C^2$.

Let suppose that a_k and s_{ik} are normally distributed in the sample with mean 0 and variance σ_C^2 and σ_R^2 correspondingly. So the probability density LH for a pedigree can be formulated as follows:

$$LH=1/(\sigma_C \sigma_R^2 (2\pi)^{\frac{3}{2}}) \int_{-\infty}^{+\infty} \exp\left(-\frac{a^2}{2\sigma_C^2}\right) \prod_i \exp\left\{-\frac{(X_i-F(Y_i)-a)^2}{2\sigma_R^2}\right\} da \qquad (1)$$

Now let us substitute the continuous normal distribution of a, with a stepwise one, having constant probabilities $p_{-J}, p_{-(J-1)}, \ldots, p_{(J-1)}, p_J$, defined on the intervals having mean values $a_j=j\cdot \beta$, $-J <= j <= J$. The sum of all p_j is equal to 1 and $p_j=p_{-j}$. The parameter β is selected to produce the same variance σ_C^2: $\beta = \frac{\sigma_C}{\sqrt{\sum_{-J}^{J}(j^2 p_j)}}$. The integer number J and set of p_j should approximate the distribution of a. The LH for this model is now:

$$LH=1/(2\pi\sigma_R^2) \sum_{-J}^{J} [p_j \prod_i \exp\left\{-\frac{(X_i-F(Y_i)-j\beta)^2}{2\sigma_R^2}\right\}] \qquad (2)$$

This form of LH can be extended on pedigrees including N measured individuals and having k factors with variances $\sigma_{C\ k}^2$. Each factor corresponds to independent variability source, shared for definite k-th group of individuals in the pedigree. The shifts $j\beta_k$ are included in the exponent only for those individuals who share the appropriate factor. For example the pedigree, consisting of two parents and one offspring, may be described with four shared factors. Two additive genetic factors each with variance equal to $1/2\sigma_{Add}^2$: the first is shared by father and offspring, the second is shared by mother and offspring. Two environment factors, attributed to spouses and household environment, are shared for parents and for the whole nuclear family correspondingly. Here we have 4 overlapped groups of individuals: 1-father and offspring, 2-mother and offspring, 3-parents, 4 – offspring and both parents. In general for each pedigree member n and each shared factor corresponding to k-th group of individuals we can define δ_{n_k}, which equals 1 or 0 depending on relation type of the individual n, which share or not the factor k. For a pedigree having N measured individuals and described by K factors with variances $\sigma_{C\ k}^2$, shared by appropriate K overlapping groups, we have

$$LH = \sum_{J_0}(p_{J_0} \sum_{J_1}(p_{J_1} \ldots \sum_{J_K}(p_{J_K} [\prod_N \{\exp\left(-\frac{\left(X_n-F(Y_n)-\sum_K \delta_{n_k} J_k \beta_k\right)^2}{2\sigma_{R_n}^2}\right)/(\sigma_n\sqrt{2\pi})\}]) \ldots)), \qquad (3)$$

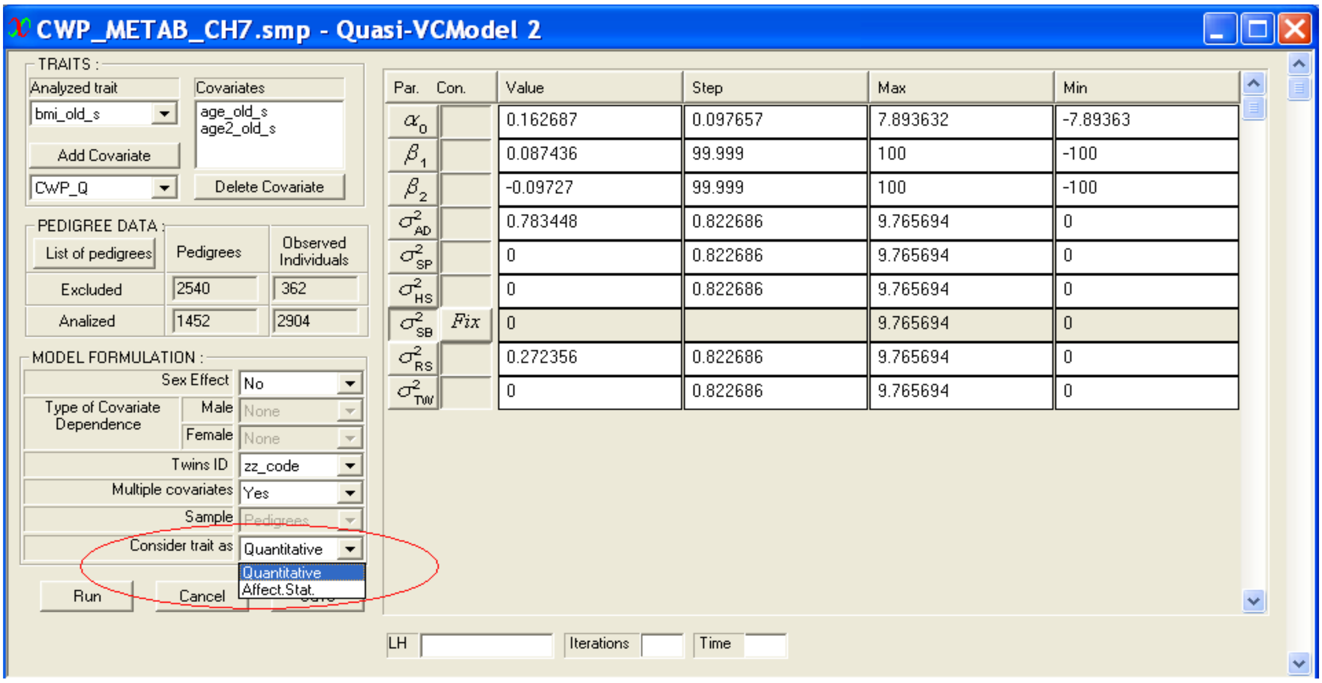
where indexes J_k change from –J to J; p_{J_k} is the probability of definite shift $J_k\beta_k$, corresponding to factor number k; δ_{n_k} equals 1 or 0 depending on belonging to group k; X_n is individual trait value; $F(Y_n)$ is predicted by covariates individual value; because all factors are supposed to be orthogonal $\sigma_{R_n}^2 = \sigma_{Tot}^2 - \sigma_{Covar}^2 - \sum_K \delta_{n_k} \sigma_{C\ k}^2$.

If the sample consists of MZ and DZ twin pairs, we may estimate additive genetic σ_{Add}^2 and common sibs (twin) environment σ_{Sib}^2 components, using LH (2), where for MZ families $\sigma_{CMZ}^2 = \sigma_{Add}^2 + \sigma_{Sib}^2$, $\sigma_{RMZ}^2 = \sigma_{Tot}^2 - \sigma_{Covar}^2 - \sigma_{Add}^2 - \sigma_{Sib}^2$, for DZ families $\sigma_{CDZ}^2 = 0.5\sigma_{Add}^2 + \sigma_{Sib}^2$, $\sigma_{RDZ}^2 = \sigma_{Tot}^2 - \sigma_{Covar}^2 - 0.5\sigma_{Add}^2 - \sigma_{Sib}^2$.

If the trait X is a liability of affection status B with affection threshold τ, the probability density LH (2) or (3) can be integrated independently on each X_i above or below the liability threshold, depending on affection status B_i. The one dimensional integral is a tabulated function. The MLE parameters have in this model version two constraints. First is the liability variance in the sample σ_{Tot}^2=1; second the sample mean of liability equals to zero. So, in this case the individual specific variance component V_{RS} and the constant, included in the covariate adjustment $F(Y_n)$, are not independent parameters. The threshold parameter τ is an estimated parameter.

5.5.1. User interface of Quasi VCA.

To use the analysis click "Analysis"-"Quasi-VCA"-" Univariate" menu item. The analysis window is similar to VCA window. The only difference in MODEL FORMULATION group is "Consider trait as" combobox. Its value define if trait *X* is considered as quantitative trait or as affection status. If affection status is selected L_{AF} parameter appears in the parameter table. Its value is the threshold dividing the sample to affected ($X_i >= L_{AF}$) and not affected ($X_i < L_{AF}$) individuals. The value L_{AF} by default is the mean value of trait X, but it can be changed, to move the threshold. This parameter is not maximized, it is used only on the beginning of the analysis to convert X to binary trait *B*. Other parameters are similar to those of VCA, but applied to liability of binary trait *B*. The liability, as mentioned upper, is normalized to have zero mean and variance equal to1. The liability affection threshold τ (L_{LAB}) is a maximized parameter corresponding to liability scale.



CWP_METAB_CH7.smp - Quasi-VCModel 2

TRAITS :

Analyzed trait: bmi_old_s

Covariates: age_old_s, age2_old_s

CWP_Q

PEDIGREE DATA :

List of pedigrees	Pedigrees	Observed Individuals
Excluded	2540	362
Analized	1452	2904

MODEL FORMULATION :

Sex Effect: No

Type of Covariate Dependence: Male: None; Female: None

Twins ID: zz_code

Multiple covariates: Yes

Sample: Pedigrees

Consider trait as: Quantitative

Par.	Con.	Value	Step	Max	Min
α_0		0.162687	0.097657	7.893632	-7.89363
β_1		0.087436	99.999	100	-100
β_2		-0.09727	99.999	100	-100
σ^2_{AD}		0.783448	0.822686	9.765694	0
σ^2_{SP}		0	0.822686	9.765694	0
σ^2_{HS}		0	0.822686	9.765694	0
σ^2_{SB}	Fix	0		9.765694	0
σ^2_{RS}		0.272356	0.822686	9.765694	0
σ^2_{TW}		0	0.822686	9.765694	0

5.6. Bivariate QVCA.

For bivariate analysis for two quantitative traits we used additional assumptions. For each type of shared factors, included in the univariate analysis we supposed the existence of decomposition to three independent factors, two trait-specific and one common for two traits; each is normally distributed. For each trait the univariate factor is a linear combination of its specific part and the common factor. Hence for each type of factors included in the model, we have 3 summation indexes J for step wise distributions: two for independent specific for each trait sources of variability and one for the common source. For individual specific variability we also suppose the common source of variability for two traits, but here only this common factor is indexed independently for each individual. In bivariate analysis we use estimates of VC for both traits, which we obtained separately for each trait in univariate analyses. The parameters, which are to be found in MLE, are proportions of common variance for each trait $\nu_{1c,}$ ν_{2c} in each type of VC. The correlation coefficient in each type of VC is $R_C^2=\nu_{1C*}\nu_{2C}$. The sign of the correlation coefficient is included in R and can be different for different types of VC (for example genetic additive and sibling environment components). The probability density LH for two quantitative traits is expressed (similar to LH (3) in section 5.5) as a sum through all shift indexes of the products of 2N one-variable normal distributions (each trait for each measured individual). One or both traits can be a liability X with threshold τ of dichotomous affection status B. In this case the appropriate integration of probability density LH on individual liability traits X_n, corresponding to affection status B_n, should be included and τ (or two different $\tau_{1,}$ τ_2) is additional estimated parameter.

For bivariate analysis it is preferable to use the most parsimonious model for each trait. It can be loaded to upper or lower part of the bivariate window as described for bivariate VCA. The menu item “Analysis”-“ Quasi-VCA”-“Bivariate” should be used to load the model, when it is selected (black column) in the result table window (see 2.3.2 section).

The component correlations R for different variance components are presented in the parameter table of the lover window. Parameters ν correspond to shared by two traits proportion of variance, attributable to trait model in the lower window. Parameters ν for the upper trait can be computed as $\nu[1]=R^2/\nu$.

6. Linkage and disequilibrium analysis.

In linkage analysis, each pedigree member is characterized by two functionally different traits, the trait describing the biological function being studied and the specific marker locus, which usually does not have its own phenotypic manifestation of interest, but for which the chromosomal position is known. Suppose that, after analyzing the pedigree sample, we find a statistically significant association between the transmission across generations of the phenotypes of the trait under study and the marker genotypes. Then linkage between a putative locus controlling the trait and the marker locus will have been established and we can find the chromosomal position of this putative locus by point and interval estimation of the recombination fraction between the two loci.

The disequilibrium (or association) analysis establishes connection between genotypes on marker locus and trait phenotypes for individuals selected from the sample by special design. Some types of this analysis can be performed not only on pedigree sample, but also on the sample of unrelated individuals. In the case of pedigree sample, the additional phenotypic correlation between relatives should be taken into account. The basic idea of this analysis is that, if the putative locus regulating the phenotype is sufficiently close to the marker locus and they are in linkage-disequilibrium (LD) with each other. Nevertheless, the association established in the analysis may be not only result of LD, but also probably of some other reasons, for example population stratification. The distance between marker and putative gene locus cannot be estimated in this analysis. As a rule, the analysis is performed on the set of sufficiently closely placed markers, covering the chromosomal region of interest.

6.1 Model based linkage.

Let the segregation analysis has established the major gene (MG) effect for a quantitative phenotype X in the pedigree sample (see section 3) and the most parsimonious Mendelian model for inheritance of this trait was built. Let some marker was genotyped on the same sample and this data on the sample individuals are Y. For each pedigree, the likelihood, including joint inheritance of quantitative phenotype X and marker data Y, can be built. The model of joint inheritance contains parameters described in 3.1 and additional parameters, characterizing marker genotype distribution and joint distribution and inheritance of marker and putative MG.

6.1.1. Joint trait-marker model of inheritance.

Let A_1 and A_2 be the two alleles at the trait locus (major gene), characterized with allele frequency p or genotype frequencies p_1, p_2 (see 3.1). Let the marker locus has M alleles M_m, having population frequencies $\Pr(M_m)=q_m$, $q_1+ q_2+\ldots+ q_M= 1$. Besides these parameters (p and q_m), the joint inheritance of the trait and marker is described by additional parameters.

The first is the recombination fraction, θ, determining linkage between the two loci. It is possible to use θ_m and θ_f – different recombination fractions for each sex. θ or θ_m,θ_f – parameters are used to formulate probability $P(g_i|g_m,g_f)$, where g_m and g_f are genotypes of the parents of the i-th pedigree member. Here genotype g_i is a combination of haplotypes on two loci: putative trait locus and marker locus.

The second group of parameters are elements of allele disequilibrium matrix, $D_{am} = \Pr(A_aM_m) - p_aq_m$ (the difference between the population frequency of the haplotype A_aM_m and the frequency expected when the alleles at the two loci are distributed independently; $a = 1,2$). There are only $M - 1$ independent values of D_{am} because of the constraints $D_{1m} = -D_{2m}$ for any m, and $\sum_m D_{am} = 0$ for any a. We actually use as model parameters Levontin's D' disequilibrium coefficients $\delta_m = D_{1m}/max\{D_{1m}\}$, where $max\{D_{1m}\}$ is the maximal disequilibrium matrix element possible with given gene and marker allele frequencies. The third group are the haplotypes disequilibrium parameters w_{kl} , normalized analogously to δ_m. Parameters p, q_m δ_m, w_{kl} are used to formulate probability of founder (member that has no parents in the pedigree) to have genotype g_i.

The linkage pedigree likelihood includes phenotype-major genotype correspondence parameters the same as described in 3.1.3 and 3.1.4, because the marker genotype does not influence explicitly the trait phenotype. However $\Pr(g_i|g_m,g_f)$-probabilities and founders $\Pr(g_i)$ components include the whole marker-gene genotype. In model-based linkage, the parameters, which were estimated for MG Mendelian most parsimonious model, are not estimated once more for linkage likelihood. Only the specific linkage parameters describing marker allele frequencies q_m (M-1 independent parameters), describing joint population distribution of (marker+MG) haplotypes δ_m(M-1 independent parameters), describing haplotype pairs distribution w_{kl} ((M-1)*(2M-1) independent parameters) and describing joint gene marker inheritance parameter - recombination fraction θ (or θ_m, θ_f), are estimated during the likelihood maximization.

6.1.2. Linkage test

The LRT statistics used to test the null hypothesis of no linkage, H_0: $\theta = 0.5$ (or $\theta_m = \theta_f = 0.5$), can be formulated as follows:

$$\lambda = 2\ln[P(X,Y\mid\hat{\theta})/P(X,Y\mid\theta=0.5)],$$

where $\hat{\theta}$ is the recombination fraction estimated together with other parameters, of the joint distribution in pedigree members of the two phenotypes, the trait and marker. Asymptotically, the test statistic is distributed as a central χ^2 with df = 1 (provided $\hat{\theta}$ is not restricted to being ≤0.5).

Additionally the significance of disequilibrium parameters can be tested in appropriate LRT.

6.1.3. User interface of model-based linkage analysis.

To start the analysis, we need to choose one of MG models from the saved segregation analysis results. To do this, click on the appropriate table icon in the document tree. When the table window opens, select the column with desirable Mendelian model (the column should appear in black). Click on “Analysis”-“Linkage”-“Model based” menu item. If the trait or covariate, used in the analysis, was deleted from the sample, or the selected column does not contain Mendelian model, you will get the error message. Otherwise, the analysis window opens.

In the analysis window you can see the trait, covariate, model formulation options, parameter constraints and estimates of the selected Mendelian model, but cannot change them.

In the TRAITS control group the only enabled control is the “Marker”-combobox. You can select the marker from the list of all marker traits, available in the sample and having number of alleles not greater than 5. If you are going to use a marker with greater number of alleles, you should factorize it (see 1.6.5). The number of alleles in the selected marker locus determines the number of appropriate frequency and disequilibrium parameters, which appear in the parameter table (see 6.1.1).

In the MODEL FORMULATION group, three specific linkage options are available.

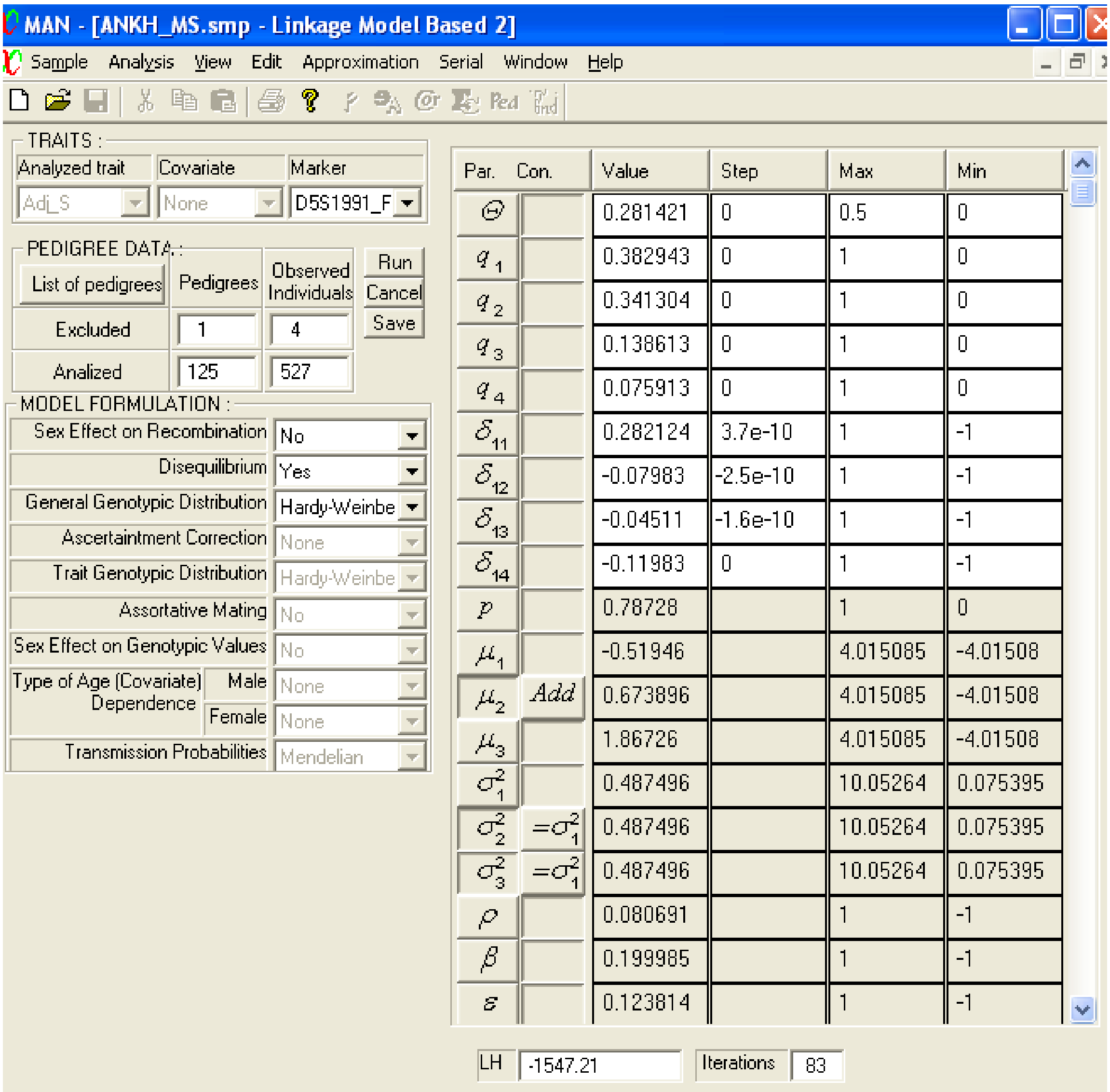
Sex effect on recombination. By default “No”. If “Yes” was selected, two sex specific recombination fractions θ_m,θ_f appear in the parameter table.

Disequilibrium. By default “No”. All allelic disequilibrium parameters are supposed to be zero. If “Yes” was selected M-1 δ_m parameters appear in the parameter table.

General Genotypic Distribution. By default “Hardy-Weinberg” (the Hardy-Weinberg equilibrium for haplotypes is supposed. If “Arbitrary” was selected the appropriate number of w_{kl} disequilibrium parameters appear in the parameter table.

In the parameter table the constraints for segregation model parameters can not be changed. For additional linkage parameters, only “Fixed” constraint is possible to introduce in the parameter table. To do this type the desired value in the appropriate string of “Value”-column of the parameter table and then click on the parameter button.

When the model and parameter constraints are specified, click “Run” to start maximization.



TRAITS :

Analyzed trait	Covariate	Marker
Adj_S	None	D5S1991_F

PEDIGREE DATA :

List of pedigrees	Pedigrees	Observed Individuals
Excluded	1	4
Analized	125	527

MODEL FORMULATION :

Sex Effect on Recombination		No
Disequilibrium		Yes
General Genotypic Distribution		Hardy-Weinbe
Ascertainment Correction		None
Trait Genotypic Distribution		Hardy-Weinbe
Assortative Mating		No
Sex Effect on Genotypic Values		No
Type of Age (Covariate) Dependence	Male	None
	Female	None
Transmission Probabilities		Mendelian

Par.	Con.	Value	Step	Max	Min
Θ		0.281421	0	0.5	0
q_1		0.382943	0	1	0
q_2		0.341304	0	1	0
q_3		0.138613	0	1	0
q_4		0.075913	0	1	0
δ_{11}		0.282124	3.7e-10	1	-1
δ_{12}		-0.07983	-2.5e-10	1	-1
δ_{13}		-0.04511	-1.6e-10	1	-1
δ_{14}		-0.11983	0	1	-1
p		0.78728		1	0
μ_1		-0.51946		4.015085	-4.01508
μ_2	Add	0.673896		4.015085	-4.01508
μ_3		1.86726		4.015085	-4.01508
σ_1^2		0.487496		10.05264	0.075395
σ_2^2	$=\sigma_1^2$	0.487496		10.05264	0.075395
σ_3^2	$=\sigma_1^2$	0.487496		10.05264	0.075395
ρ		0.080691		1	-1
β		0.199985		1	-1
ε		0.123814		1	-1

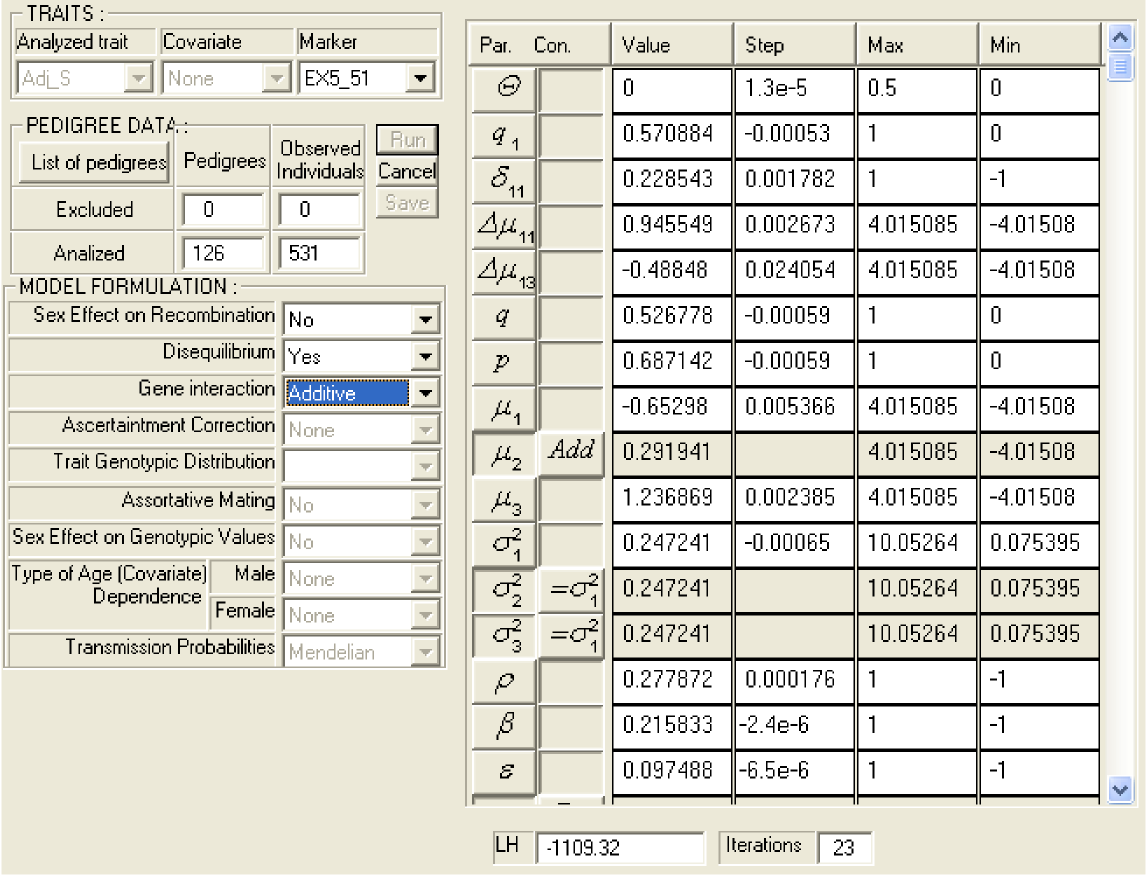
LH -1547.21 Iterations 83

6.2 Joint segregation-linkage (JSL) analysis.

In JSL analysis all parameters including parameters of segregation model are estimated together in the maximization procedure. To use this analysis select the column with desirable Mendelian model in the segregation analysis results table and click on “Analysis”-“Linkage”-“JSL” menu item. The parameters of segregation model appear as initial values in the parameter table, you can edit them here, if you need. The constraints of segregation model cannot be changed in the JSL window, if you need some other constraints use another segregation model as base for JSL analysis. The linkage test is the same as in 6.1.2.

6.3 Joint segregation-linkage digene (JSLD) analysis.

This analysis supposes that a MG model was established for a trait. The putative major gene G_{maj} is a suitable approximation to describe the general inheritance pattern, which can be done more precise by introducing an additional di-allelic trait gene G_{sm} (having smaller effect on the trait), which is linked to the investigated marker M. Formally we suppose that M and G_{sm} are on the same chromosome and G_{maj} on some other. The genotypic values $\mu_{1,}$ $\mu_{2,}$ μ_3 of G_{maj} are supposed to be a mean value for three genotypes possible as combinations of selected genotype of G_{maj} and three genotypes of $G_{sm.}$ To describe genotypic values for complex genotype $\{G_{maj}$, $G_{sm\}}$, we need in general 6 additional parameters. We introduce parameters Δ_{11}, Δ_{13}, Δ_{21}, Δ_{23}, Δ_{31}, Δ_{33}, as follows: $\mu_{11}= \mu_1+\Delta_{11}$ ($g_{maj}=1$, $g_{sm}=1$); $\mu_{13}= \mu_1+\Delta_{13}$ ($g_{maj}=1$, $g_{sm}=3$); $\mu_{21}= \mu_2+\Delta_{21}$ ($g_{maj}=2$, $g_{sm}=1$); $\mu_{23}= \mu_2+\Delta_{23}$ ($g_{maj}=2$, $g_{sm}=3$); $\mu_{31}= \mu_3+\Delta_{31}$ ($g_{maj}=3$, $g_{sm}=1$); $\mu_{33}= \mu_3+\Delta_{33}$ ($g_{maj}=3$, $g_{sm}=3$). $\mu_{12,}$ $\mu_{22,}$ μ_{32} can be computed from the condition that $\mu_{1,}$ $\mu_{2,}$ μ_3 are mean values of $G_{maj}=1,2,3$. If gene interaction is supposed to be additive or multiplicative, we need only two Δ parameters, because $\Delta_{11}=\Delta_{21}=\Delta_{31}$ and $\Delta_{13}=\Delta_{23}=\Delta_{33}$. For additive gene interaction genotypic values are $\mu_{11}= \mu_1+\Delta_{11}$; $\mu_{13}= \mu_1+\Delta_{13}$; $\mu_{21}= \mu_2+\Delta_{11}$; $\mu_{23}= \mu_2+\Delta_{13}$; $\mu_{31}= \mu_3+\Delta_{11}$; $\mu_{33}= \mu_3+\Delta_{13}$.



TRAITS :

Analyzed trait	Covariate	Marker
Adj_S	None	EX5_51

PEDIGREE DATA :

List of pedigrees	Pedigrees	Observed Individuals
Excluded	0	0
Analized	126	531

MODEL FORMULATION :

Sex Effect on Recombination	No
Disequilibrium	Yes
Gene interaction	Additive
Ascertaintment Correction	None
Trait Genotypic Distribution	
Assortative Mating	No
Sex Effect on Genotypic Values	No
Type of Age (Covariate) Dependence – Male	None
Type of Age (Covariate) Dependence – Female	None
Transmission Probabilities	Mendelian

Par.	Con.	Value	Step	Max	Min
θ		0	1.3e-5	0.5	0
q_1		0.570884	-0.00053	1	0
δ_{11}		0.228543	0.001782	1	-1
$\Delta\mu_{11}$		0.945549	0.002673	4.015085	-4.01508
$\Delta\mu_{13}$		-0.48848	0.024054	4.015085	-4.01508
q		0.526778	-0.00059	1	0
p		0.687142	-0.00059	1	0
μ_1		-0.65298	0.005366	4.015085	-4.01508
μ_2	Add	0.291941		4.015085	-4.01508
μ_3		1.236869	0.002385	4.015085	-4.01508
σ_1^2		0.247241	-0.00065	10.05264	0.075395
σ_2^2	$=\sigma_1^2$	0.247241		10.05264	0.075395
σ_3^2	$=\sigma_1^2$	0.247241		10.05264	0.075395
ρ		0.277872	0.000176	1	-1
β		0.215833	-2.4e-6	1	-1
ε		0.097488	-6.5e-6	1	-1

LH -1109.32 Iterations 23

For multiplicative they are $\mu_{11}= \mu_1\cdot\Delta_{11}$; $\mu_{13}= \mu_1\cdot\Delta_{13}$; $\mu_{21}= \mu_2\cdot\Delta_{11}$; $\mu_{23}= \mu_2\cdot\Delta_{13}$; $\mu_{31}= \mu_3\cdot\Delta_{11}$; $\mu_{33}= \mu_3\cdot\Delta_{13}$. In comparison with JSL model we have additional parameter *q*- the allele frequency of G_{sm} gene and Δ parameters (6 or 2, as was discussed above). Disequilibrium parameters δ and recombination fraction θ relate to loci M and G_{sm}.

To specify type of phenotypic interaction of genes G_{maj}, G_{sm} in the analysis window use the option *Gene Interaction* in the MODEL FORMULATION control group. By default the value is "General" and six Δ parameters appear in the parameter window. For "Additive" and "Multiplicative" options only two parameters Δ_{11}, Δ_{13} are estimated (the meaning see above).

The linkage test is the same as in 6.1.2. The most parsimonious *Gene Interaction* option can be chosen by comparison of "General" and "Additive" option by central χ^2 with df = 4.

This analysis is much more time-consuming than JSL.

6.4 Linkage and disequilibrium between genetic markers.

For two genetic markers, having M and K alleles, this analysis estimates allele frequencies p_j (M-1 independent parameters), q_i (K-1 independent parameters), disequilibrium coefficients δ_{ij} ((M-1)·(K-1) independent parameters)and recombination fraction θ (or θ_m, θ_f). If it is known that positions of two markers are close, the linkage test (6.1.2) for them can be used as a check for genotyping quality. The estimation of δ coefficients we need to characterize the LD pattern in the region.

6.4.1. User interface of marker linkage analysis.

Click "Analysis"-"Linkage"-"Markers"-menu item to open the analysis window.

In the TRAITS control group select "First marker" and "Second marker". The appropriate number of parameters appears in the parameter table.

In the MODEL FORMULATION group only two options are available:

Sex effect on recombination. By default "No", the general recombination fraction θ appears in the parameter table. If "Yes" was selected, θ_m and θ_f parameters are estimated for males and females.

Disequilibrium. By default "No", allelic equilibrium is supposed. If "Yes", δ_{ij} ((M-1)·(K-1) independent parameters) appear in the parameter table and are estimated during LH maximization.

6.5. Transmission disequilibrium test.

The transmission disequilibrium test (TDT) was formulated by Spielman et al. (Am J Hum Genet 1993; 52: 120-132) to test linkage between a marker locus and the trait being studied, provided that there is linkage disequilibrium between the marker locus and the gene taking part in the control of the trait. Initially, the test was used for binary traits (affected-unaffected).

6.5.1. Alisson's quantitative tests for parent-offspring trios.

Allison proposed the TDT for use with quantitative traits (Am J Hum Genet 1997;60: 676– 90). The family-based sample data includes parent-offspring trios with at least one heterozygous at the marker locus parent. The data analyzed contain two sub-samples, one having the marker allele transmitted to the offspring from the heterozygous parent, and the other having the untransmitted marker allele. The analysis of difference in offspring phenotypes between these sub-samples is the base of TDT. Using different statistics, Allison proposed five variants of quantitative TDT, named Q_1, Q_2, Q_3, Q_4, Q_5. Q_1 and Q_3 use Student's *t*-distribution, Q_2 uses central χ^2 with df = 1, Q_4 – Z-statistics, Q_5 – F-statistics.

Allison also proposed the TDT for an extreme-threshold (ET) sampling design defined as follows. Let Z_U and Z_L be upper and lower thresholds such that trios having offspring trait values between these thresholds are excluded from the analysis. In the particular case $Z_U = Z_L$, all the trios are used. Allison showed that this ET design increases the power of the test, but did not dwell in detail on how to choose the optimal thresholds.

6.5.2. Extreme-offspring design.

If the sample contains nuclear families with more than one offspring, some extensions of TDT were proposed, which take advantage from the greater offspring number. Keeping intact the rationale behind the construction of the TDT, we proposed an extreme-offspring (EO) sampling design that removes the problem of establishing the optimal proportion of excluded trios, and that substantially increases the power of the TDT (Malkin et al., Genet Epidemiol 2002; 23: 234– 44). For each parent pair, to form the trio one selects the offspring having the most extreme trait value among those siblings whose trait values belong to the predefined tails of the distribution. The median of the offspring trait distribution divide the distribution into left and right tail. Of course, there is no offspring selection for parent pairs, who have only one offspring with a trait value outside the intermediate trait range. Thus, for each given proportion of excluded offspring, we have the same number of trios for both EO and ET testing. The design can be applied for $Q_1 \div Q_5$ tests (for Q_1 test only if there is no phenotype selection condition).

6.5.3. Offspring-mean test (OMT).

This test represents an extension of Allison's Q3 test applied to familial offspring mean trait value. The OMT uses all offspring in the informative families and is of special advantage in the analysis of quantitative traits with substantial correlation between sibs. It also obviates measuring the trait values of both parents. Let the trait in offspring have two orthogonal random components with zero mean: between family *b* and within family *w*. Then the trait value of *j*-th offspring in the *i*-th family is modeled as $\xi_{ij} = \mu + b_i + w_{ij}$, where μ is the sample mean. Determine for each family *i* with at least one heterozygous parent the mean value q_{i1} for all offspring, which have inherited the selected allele A1 from the heterozygous parent, and the mean q_{i0} for all offspring, which have not, i.e., for offspring having transmission status (TS) 1 and 0, respectively. Then $q_{i1} = \mu + b_i + \frac{1}{n_{i1}}\sum_{j=1}^{n_{i1}} w_{ij1}$, $q_{i0} = \mu + b_i + \frac{1}{n_{i1}}\sum_{j=1}^{n_{i0}} w_{ij0}$, where n_{i1} and n_{i0} are for each family the numbers of offspring having TS=1 and TS=0. Let N_1 families in the sample have offspring with TS=1, N_0 families have offspring with TS=0 and N_{01} families have offspring with both TS=0 and TS=1. Let $M_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} q_{i1}$ and $M_0 = \frac{1}{N_0}\sum_{i=1}^{N_0} q_{i0}$ be the sample means for all informative families. Then we can test the hypothesis of no linkage disequilibrium between the trait and marker by computing the test statistics

$$T = \frac{M_1 - M_0}{\sqrt{Var(M_1) + Var(M_0) - 2Cov(M_1, M_0)}}.$$

The statistical significance of this test can be assessed by referring to the Student t-test with N-2 df, where N is the number of informative families in the sample. The estimate for covariance of means with different TS is

$$Cov(M_1, M_0) = \frac{N_{01}Var(b)}{N_1 N_0} = \frac{N_{01}}{N_1 N_0}[Var(\xi) - Var(w)],$$

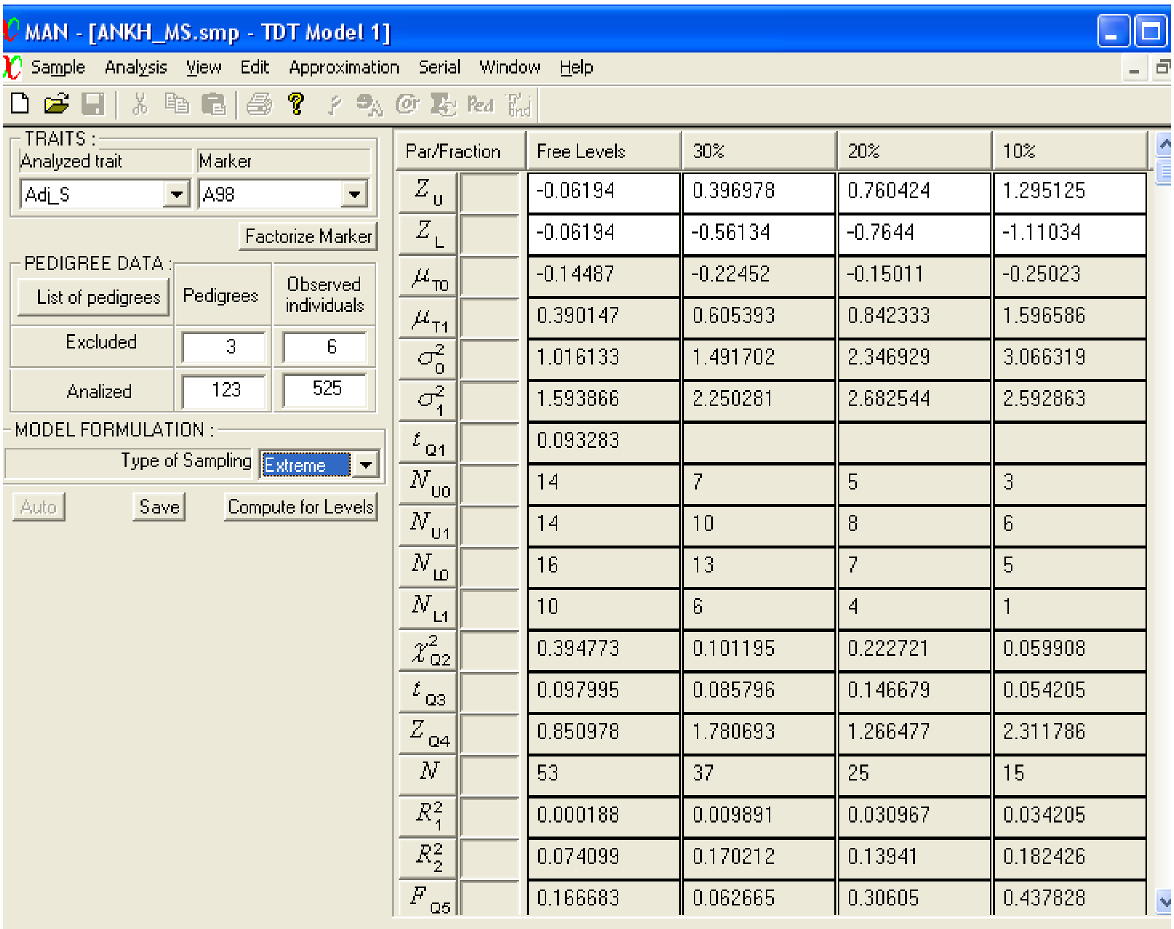
where $Var(w)$ can be estimated as a half of the mean of squared differences between sibling trait values. Both $Var(\xi)$ and $Var(w)$ estimates are computed for offspring of informative families. For both heterozygous parents having offspring of the same genotype, only families with homozygous offspring are informative. Nuclear families having different offspring genotypes are always informative, both in the case when two parents are heterozygous and when parents are not genotyped. Then the offspring having minimal number of A1 alleles have transmission status 0 and all the rest transmission status 1. An alternative definition, to assign the transmission status 1 to offspring with maximal number of A1 alleles may also be used, but it makes difference only for families having offspring of three different genotypes.

6.5.4. User interface for TDT analysis window.

Click “Analysis”-“Linkage”-“TDT”-menu item to open the analysis window.

In the TRAITS control group select “Analyzed Trait” and “Marker”. If marker has more than 2 alleles, it should be factorized before the analysis to dichotomous scheme. By default one allele is taken as “selected” allele *A* and all others are united in the second allele *a*.

Click the “Factorize Marker”-button to view or change the dichotomous scheme.



TRAITS :

Analyzed trait: Adj_S

Marker: A98

PEDIGREE DATA :

	Pedigrees	Observed individuals
Excluded	3	6
Analized	123	525

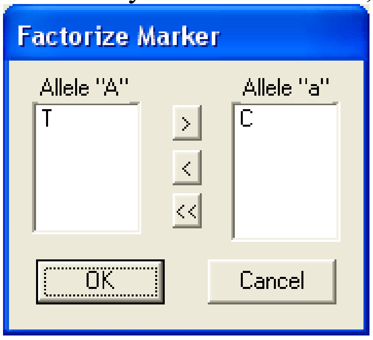
MODEL FORMULATION :

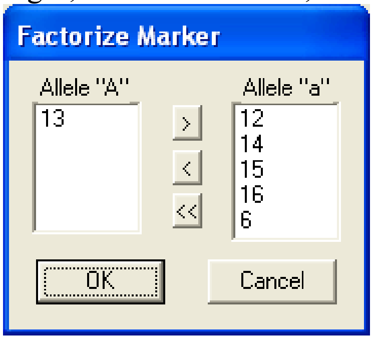
Type of Sampling: Extreme

Par/Fraction	Free Levels	30%	20%	10%
Z_U	-0.06194	0.396978	0.760424	1.295125
Z_L	-0.06194	-0.56134	-0.7644	-1.11034
μ_{T0}	-0.14487	-0.22452	-0.15011	-0.25023
μ_{T1}	0.390147	0.605393	0.842333	1.596586
σ^2_0	1.016133	1.491702	2.346929	3.066319
σ^2_1	1.593866	2.250281	2.682544	2.592863
t_{Q1}	0.093283			
N_{U0}	14	7	5	3
N_{U1}	14	10	8	6
N_{L0}	16	13	7	5
N_{L1}	10	6	4	1
χ^2_{Q2}	0.394773	0.101195	0.222721	0.059908
t_{Q3}	0.097995	0.085796	0.146679	0.054205
Z_{Q4}	0.850978	1.780693	1.266477	2.311786
N	53	37	25	15
R^2_1	0.000188	0.009891	0.030967	0.034205
R^2_2	0.074099	0.170212	0.13941	0.182426
F_{Q5}	0.166683	0.062665	0.30605	0.437828

The “Factorize Marker”- dialog contains two lists of marker alleles, the alleles in the left listbox *Allele “A”* correspond to “selected” allele. To move the marker allele from the left list to the right, select the desired allele in the left listbox (the selected string will be blue) and click “>”-button. To move from right to left, select the allele in the right listbox and click “<”-

button. The “<<”-button moves all alleles from the right to left. Click on the “OK” button to confirm you selection. If you click “Cancel”, all changes, which were made, are canceled.





In the MODEL FORMULATION control group, you should select the *Type of sampling*. By default “Random”. For each informative parent pair (at least one heterozygous parent) the first offspring, matching to the phenotype selection condition, is used for the trio.

“Extreme” – for informative parent pair the most extreme offspring in the nuclear family matching the phenotype selection condition, is selected for the trio, see 6.5.2. The whole offspring sample trait distribution is used to determine the left or right tail percentile, to which the offspring belong.

“Mean” – for informative parent pair all offspring, matching the phenotype selection condition, are used for the test as described in 6.5.3.

In the parameter table, you can see and change the phenotype selection condition, which is defined by upper and lower thresholds Z_U, Z_L. The offspring phenotype matching to the phenotype threshold should be less than Z_L or not less than Z_U. You can edit Z_U, Z_L only in the first column of parameter table named “Free levels”, by default in this column $Z_U = Z_L$ and both are equal to offspring phenotype distribution median (all genotyped offspring of the informative nuclear families can be used). After you have edited Z_U, Z_L or both, click on the “Compute for Levels” button to view in the column the recomputed values for all tests. Three other columns “30%”, “20%”, “10%”, contains levels corresponding to appropriate parts of upper and lower distribution tails and cannot be changed. The columns are given to illustrate the tendency of extreme-threshold sampling design (see 6.5.1) for a definite trait-marker pair. Only the contents of the “Free levels”-column you can save as a column in the results table (“Save”-button).

Each time as you change the quantitative trait in the *Analyzed Trait*-combobox the *Type of sampling* changes to its default value “Random”, the marker factorization also changes to its default (initial state), the default thresholds and all tests are recomputed for all four columns in the parameter table.

Each time as you change the marker, or marker factorization, or type of sampling all four columns are recomputed for the default thresholds.

If the marker has more than two alleles, the “Auto”-button in the left-lower corner of the analysis window is enabled. Click on this button compute the tests for the selected type of sampling and for all possible factorizations of type: one allele versus all others. The results are added as separate columns to the TDT table, corresponding to the same trait-marker pair. If such a table does not exist, a new table with default name “Table.”+<number> is created.

6.5.5. Values displayed in the TDT parameter table.

The values displayed in parameter table are not only p-values of the tests, but also some values, which are included in the computed statistics and are useful for understanding of trait-marker interaction.

Z_U, Z_L define the phenotype selection condition: the offspring phenotype should be less than Z_L or not less than Z_U.

μ_{T0} – the mean trait value in the sub-sample of offspring, having selected allele (A) transmitted from the heterozygous parent (for OMT this is mean of family means);

μ_{T1} – the mean trait value in the sub-sample of offspring, having selected allele (A) non-transmitted from the heterozygous parent (for OMT this is mean of family means);

σ^2_0 – the trait variance in the sub-sample of offspring having selected allele (A) transmitted from the heterozygous parent (for OMT this is variance of family means);

σ^2_1 – the trait variance in the sub-sample of offspring having selected allele (A) non-transmitted from the heterozygous parent (for OMT this is variance of family means);.

t_{Q1} – p-value of Q_1 test;

N_{U0} – number of offspring used in the test having selected allele (A) transmitted and trait value not less than Z_U;

N_{U1} – number of offspring used in the test having selected allele (A) non-transmitted and trait value not less than Z_U;

N_{L0} – number of offspring used in the test having selected allele (A) transmitted and trait value less than Z_L;

N_{L1} – number of offspring used in the test having selected allele (A) non-transmitted and trait value less than Z_L;

χ^2_{Q2} - p-value of Q_2 test;

t_{Q3} – p-value of Q_3 test, accounting for not equal sizes of transmitted and non-transmitted sub-samples;

Z_{Q4} – Z-statistics for Q_4 test.

N – number of parent-offspring trios taking part in Q_5 test.

R^2_1- multiple regression coefficient of two parent genotypes on the offspring phenotype;

R^2_2- multiple regression coefficient of three genotypes (parents and offspring) on the offspring phenotype;

F_{Q5} - p-value of Q_5 test.

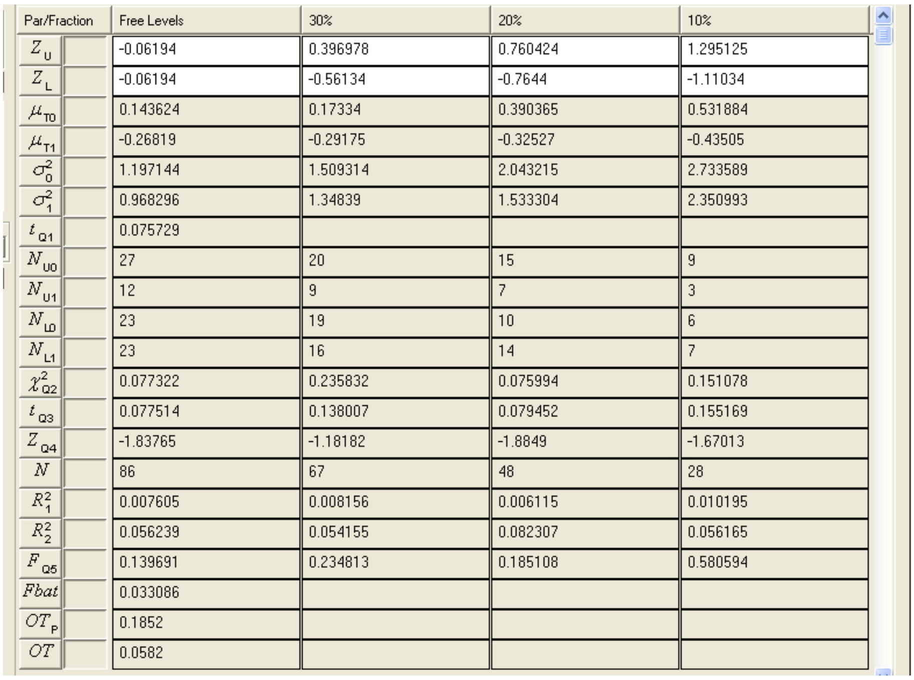
6.5.6. The external tests.

It is possible to compute for the defined trait-marker pair three additional TDT like tests, implemented in the external executables.

The first is the family-based association test proposed by Horvath et al.(Eur J Hum Genet 2001; 9: 301–6.) and implemented in FBAT program (executable file fbat1.4_win32.exe available from http://www.biostat.harvard.edu/~fbat).

The second and third are two different versions of the orthogonal test proposed by Abecasis et al. (Am J Hum Genet 2000; 66: 279– 92.) with adjustment for the parent phenotypes (OT_P) and without it (OT). Both tests are included in the QTDT program (executable file qtdt.exe available from http://www.sph.umich.edu/csg/abecasis/QTDT).

If you have downloaded these executable files, copy them to the directory where the MAN executable file maxlh1.exe is. Now you can add the p-values of these tests to the parameter table of TDT analysis. To do this, click on "Analysis"-"External"-"FBAT" menu item. The menu item will be checked and, when the TDT analysis window is active, in the parameter table appears an additional line named *Fbat* , which contains the test p-value. To include orthogonal tests click "Analysis"-"External"-"QTDT" menu item. The menu item will be checked and, when the TDT analysis window is active, in the parameter table appear two additional lines named OT_P and *OT.* If the menu items are checked, each time as you change trait, marker or marker factorization in the TDT analysis window the new data input files will be made for external executable and they will start in the background regime using appropriate batch files and produce their output. The p-values from the output are included in the first column of the parameter table and can be saved together with all other tests and parameters. All the produced files (input and output of the external programs) are placed in the directory corresponding to environment variable "TMP" for the user (for Windows XP it is "%USERPROFILE%\Local Settings\Temp").



Par/Fraction		Free Levels	30%	20%	10%
Z_U		-0.06194	0.396978	0.760424	1.295125
Z_L		-0.06194	-0.56134	-0.7644	-1.11034
μ_{T0}		0.143624	0.17334	0.390365	0.531884
μ_{T1}		-0.26819	-0.29175	-0.32527	-0.43505
σ^2_0		1.197144	1.509314	2.043215	2.733589
σ^2_1		0.968296	1.34839	1.533304	2.350993
t_{Q1}		0.075729			
N_{U0}		27	20	15	9
N_{U1}		12	9	7	3
N_{L0}		23	19	10	6
N_{L1}		23	16	14	7
χ^2_{Q2}		0.077322	0.235832	0.075994	0.151078
t_{Q3}		0.077514	0.138007	0.079452	0.155169
Z_{Q4}		-1.83765	-1.18182	-1.8849	-1.67013
N		86	67	48	28
R^2_1		0.007605	0.008156	0.006115	0.010195
R^2_2		0.056239	0.054155	0.082307	0.056165
F_{Q5}		0.139691	0.234813	0.185108	0.580594
$Fbat$		0.033086			
OT_P		0.1852			
OT		0.0582			

6.6. Pedigree disequilibrium test (PDT).

Consider the monogenic model of trait inheritance, in which it is assumed that the trait genotypes on pedigree members are exactly their marker genotypes. Let $P(X_n, C_n \mid \mu)$ be the pedigree sample likelihood of this model when it is assumed that all the marker genotypes have the same genotypic value μ for the trait in each pedigree member. Further, let $P(X_n, C_n \mid \mu_g)$ be the same pedigree likelihood defined by the MG model for which the three genotypic values μ_g that determine the trait control (g = 1, 2 and 3 for genotypes A_1A_1, A_1A_2 and A_2A_2, respectively) are distinguished. If the marker has more then two alleles, it should be factorized using some dichotomous scheme.

Introduce the test in the form:

$$\text{LRT} = 2\ln[P(X_n, C_n \mid \hat{\mu}_g) / P(X_n, C_n \mid \overline{\mu})],$$

where $\hat{\mu}_g$ are the maximum likelihood estimates of the μ_g model parameters, and $\overline{\mu}$ is the parameter maximizing the null hypothesis model $P(X_n, C_n \mid \mu)$. We expect that, if this null hypothesis is true, the LRT is distributed asymptotically as χ^2 with 2 df (the difference between the numbers of estimated parameters in the two likelihoods).

To test the significance of trait regression on the number of A_1 alleles, we can compare $P(X_n, C_n \mid \mu)$ with the additive model $P(X_n, C_n \mid \mu_g, \mu_2 = (\mu_1 + \mu_3)/2)$, the appropriate LRT is distributed asymptotically as χ^2 with 1 df.

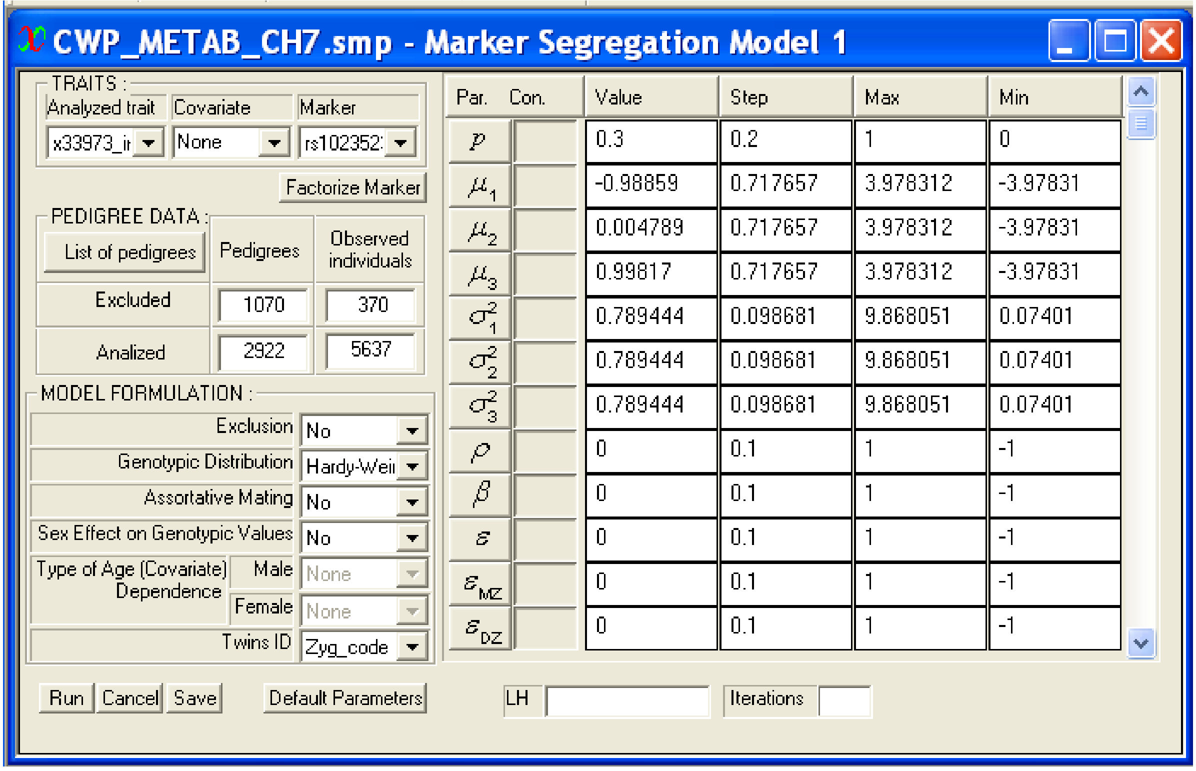
The LRT is used to reject the null hypothesis that all marker genotypes exhibit the same mean trait value. Analogously, if the segregation model includes some kind of covariate dependence of genotypic values (see 3.1.4), more complex LRT can be formulated with appropriate number of df.

In PDT the complete pedigree data are analyzed instead of only members of informative nuclear families as in TDT. The non-genotyped pedigree members are also included in the test, for such pedigrees summation through all compatible marker genotypes is done. The test is very sensitive to disequilibrium. Note, however, that in a stratified population an association can also be found in the absence of linkage.

6.6.1. User interface for PDT.

Click “Analysis”-“Linkage”-“Marker segregation”-menu item to open the analysis window.

In the TRAITS control group select “Analyzed Trait”, “Covariate” and “Marker”. If marker has more than 2 alleles, it should be factorized before the analysis to dichotomous scheme. Click the “Factorize Marker”-button to view or change the dichotomous scheme (about factorization see 6.5.4 section).



CWP_METAB_CH7.smp - Marker Segregation Model 1

TRAITS :

Analyzed trait	Covariate	Marker
x33973_ir	None	rs102352

PEDIGREE DATA :

List of pedigrees	Pedigrees	Observed individuals
Excluded	1070	370
Analized	2922	5637

MODEL FORMULATION :

Exclusion: No
Genotypic Distribution: Hardy-Wei
Assortative Mating: No
Sex Effect on Genotypic Values: No
Type of Age (Covariate) Dependence: Male None; Female None
Twins ID: Zyg_code

Par.	Con.	Value	Step	Max	Min
p		0.3	0.2	1	0
μ_1		-0.98859	0.717657	3.978312	-3.97831
μ_2		0.004789	0.717657	3.978312	-3.97831
μ_3		0.99817	0.717657	3.978312	-3.97831
σ_1^2		0.789444	0.098681	9.868051	0.07401
σ_2^2		0.789444	0.098681	9.868051	0.07401
σ_3^2		0.789444	0.098681	9.868051	0.07401
ρ		0	0.1	1	-1
β		0	0.1	1	-1
ε		0	0.1	1	-1
ε_{MZ}		0	0.1	1	-1
ε_{DZ}		0	0.1	1	-1

The MODEL FORMULATION control group is similar to that of segregation analysis (see 3.4.1) The only difference is "Twins ID" combobox, which enables to select integer trait defining twin sets and type as described in 5.1.1, if the sample contains twins.

The parameter table is also similar to that of segregation analysis (see 3.4.2), but if Twins ID trait is selected two additional partial correlation parameters ε_{MZ}, ε_{DZ} appear in the parameter table for MZ and DZ twins correspondingly.

7. Serial analysis.

The linkage and disequilibrium analyses in MAN work with single trait-marker pair. As a rule for genetic research genotyping is executed for a set of markers covering a definite chromosomal region. It is also often the case that a set of different traits is supposed to be linked or associated to the same marker set. For these cases we designed the serial analysis. The user should select from the sample the set of traits and markers, which are to be analyzed, and should formulate the collection of tests to perform for each trait-marker pair. Then for each possible trait-marker pair the desired collection of the analyses is performed automatically and results are saved in the tables. For adjacent markers haplotypes can be reconstructed and used in the analysis.

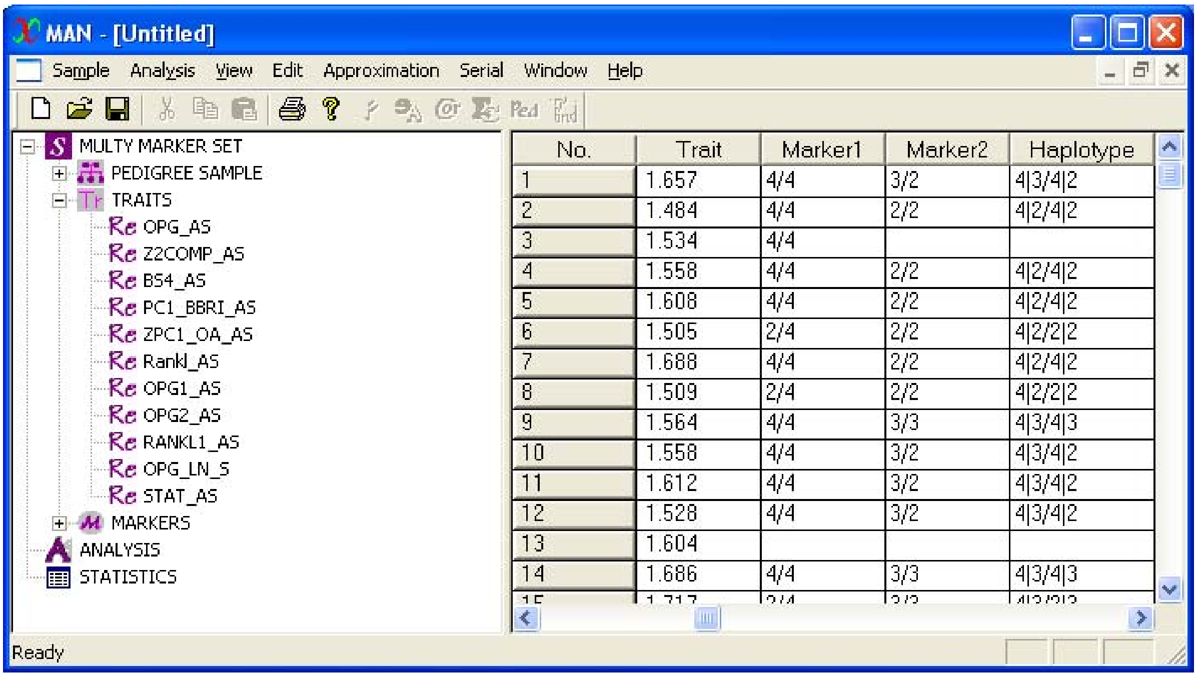
To select the results of interest you can apply filters, which can include any combination of parameters for the formulated set of analyses. The results can be also presented in graphical form. The LD pattern for a marker set can be computed and displayed.

7.1 Creating new serial analysis document.

To use serial analysis, you should have a sample file (extension ".smp"), which include as separate columns all traits and marker data (300 columns can be specified).

Open this file, then to create the new serial analysis click "Analysis"-"Serial Analysis"-menu item. The new window of serial analysis document will be opened. The left side of the window contains the document tree, the right a data table. In the right side you can see only four traits: real trait named "Trait" and three markers "Marker1", "Marker2" and "Haplotype".

This reflects the way how the serial analysis will run. The data of each pair of markers adjacent in the marker list are loaded into the di-allelic markers "Marker1", "Marker2", then haplotypes are reconstructed for these two markers and loaded into the 4-allelic marker "Haplotype", then sequentially data of each trait from the trait list are loaded into "Trait" column, if some new derived traits were specified, their contents will be recalculated with new "Trait" data. Then all analyses specified are performed and results are saved.

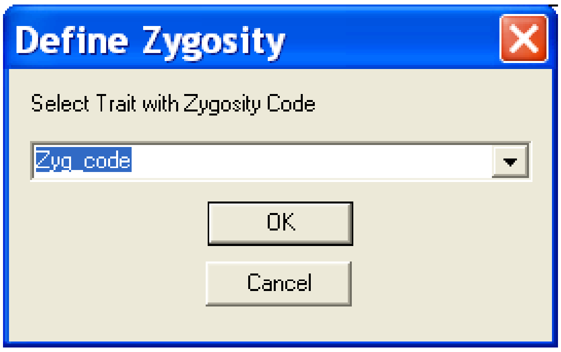


No.	Trait	Marker1	Marker2	Haplotype
1	1.657	4/4	3/2	4|3/4|2
2	1.484	4/4	2/2	4|2/4|2
3	1.534	4/4		
4	1.558	4/4	2/2	4|2/4|2
5	1.608	4/4	2/2	4|2/4|2
6	1.505	2/4	2/2	4|2/2|2
7	1.688	4/4	2/2	4|2/4|2
8	1.509	2/4	2/2	4|2/2|2
9	1.564	4/4	3/3	4|3/4|3
10	1.558	4/4	3/2	4|3/4|2
11	1.612	4/4	3/2	4|3/4|2
12	1.528	4/4	3/2	4|3/4|2
13	1.604			
14	1.686	4/4	3/3	4|3/4|3
15	1.717	2/4	[illegible]	[illegible]

If TDT or PDT are specified for haplotype, the analysis is performed for all possible factorizations of type: one allele versus all others.

7.1.1. Specifying zygosity for samples with twins.

If the sample contains twins, they should be identified with some integer variable, which for each twin set should have the same value and should be unique for each separate set within nuclear family. The monozygotic twins should have positive ID, the dizygotic twins - negative. If you need identify twins in serial analysis (for VCA or PDT analysis) click **Serial-Define Zygosity** menu item. This option is available only before specifying analyses in Serial Analysis Window.


Define Zygosity

Select Trait with Zygosity Code

Zyg_code

When the *Define Zygosity Dialog* openes select in the combobox the appropriate integer trait containing Twin ID. After clicking OK the trait will be available in the *Twin ID* combobox in VCA and PDT analysis windows.

7.2. Defining quantitative traits for serial analysis

In the document tree open the item TRAITS by click on "+" or double click on the item name. You can see the list of all real traits, which were included in the smp-file. If you don't need for serial analysis some of the traits, you can delete it from the list. To do this, select the trait by mouse click on the appropriate trait name in the tree and press "Delete" key on the keyboard (note that there is no undo option for this operation). When the trait is deleted the cursor moves automatically to the next trait item. So you can easily delete a group of adjacent traits by pressing on "Delete" key more times.

As it was mentioned above (5.2.1), if you are going to use some analysis based on likelihood maximization (all analyses in MAN besides TDT), it is recommended to include in the serial list of traits only standardized traits. Otherwise the maximization procedure for some of the traits can have difficulties with very small or very great values, which cannot pass the intermediate maximization limitation conditions, it likely ends with error message and the serial analysis stops. To rerun it, you should clear all already computed results (click on menu item "Serial"-"Clear Iterations"), then it is possible to delete from the serial list of traits the trait, which have caused the error, and to start analysis with new trait list.

It is possible to include standardization or some other transformation explicitly in the serial analysis. For this purpose you can use menu items "Edit"-"Add Trait"-"Formula" and "Edit"-"Add Trait"-"Standardize" to create new traits. For example you can specify calculation of square or logarithm of "Trait" as "Trait1" and then "Trait2"as standardized "Trait1". You can also specify deletion of extremely long distribution tails for standardized traits. But you should take into account that the trait cycle is the internal cycle of serial analysis; each time as the next trait from the list is loaded into “Trait” column, all specified new traits are recalculated and this takes appropriate time.

7.3. Defining markers for serial analysis.

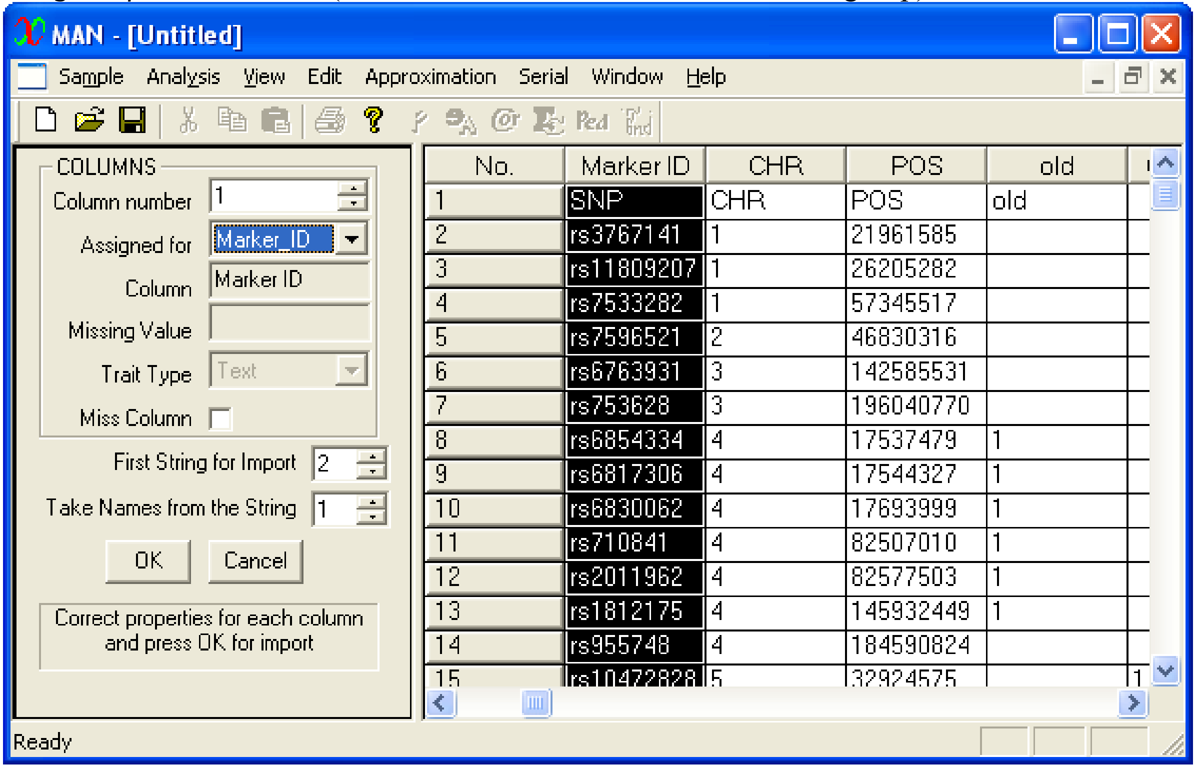
In the document tree open the item MARKERS by click on “+” or double click on the item name. Only diallelic markers, which were available in the smp-file are presented in the marker list. You can delete the markers from the list in the same way as for real traits (7.1.1).

The order of markers is significant, if you are going to use haplotypes of adjacent markers or to compute the disequilibrium pattern. There are to ways to establish the order of the markers.

Fist way, you can move the markers in the tree manually as follows. Select the marker, which you want to move by mouse click. Then press the “↑” key on the keyboard to move the selected marker up, or “↓”key to move the selected marker down.

The other way is to specify directly the chromosomal position of all markers. To do this you should have an information file containing a table including columns: marker name, chromosome name, marker position. The order of columns is not significant. The file can be an Excel worksheet or tab-delimited text. Each name of markers in your document tree should fully coincide with some marker name in you information file. To import the information, click on menu item “Serial”-“Markers”-“Import positions”. In the open file dialog select the name of your information file.

In the right side of the opened window you will see the contents of your file. In the left part of the window specify *First string for import* – the number of string (corresponding to No.-column in the right sub-window), where the first marker name is. Then, first, click on the header of the column, where all marker names are (the column should appear in black). In the *Assigned for*- combobox (left sub-window , COLUMNS control group) select “Marker ID”.



COLUMNS

Column number: 1

Assigned for: Marker_ID

Column: Marker ID

Missing Value:

Trait Type: Text

Miss Column: □

First String for Import: 2

Take Names from the String: 1

Correct properties for each column and press OK for import

No.	Marker ID	CHR	POS	old	
1	SNP	CHR	POS	old	
2	rs3767141	1	21961585		
3	rs11809207	1	26205282		
4	rs7533282	1	57345517		
5	rs7596521	2	46830316		
6	rs6763931	3	142585531		
7	rs753628	3	196040770		
8	rs6854334	4	17537479	1	
9	rs6817306	4	17544327	1	
10	rs6830062	4	17693999	1	
11	rs710841	4	82507010	1	
12	rs2011962	4	82577503	1	
13	rs1812175	4	145932449	1	
14	rs955748	4	184590824		
15	rs10472828	5	32924575		1

Second, click on the column with chromosome names and select in the *Assigned for*- combobox option “Chromosome”. Third, click on the column with chromosomal positions and select in the *Assigned for*- combobox option “Distance”. After that, click OK. If some markers, defined for serial analysis, have not got position or some markers in your information file were not found in the serial marker list, you will see an error message. Read the message attentively, probably some marker names in your information file are no the same as in the marker list of serial analysis.

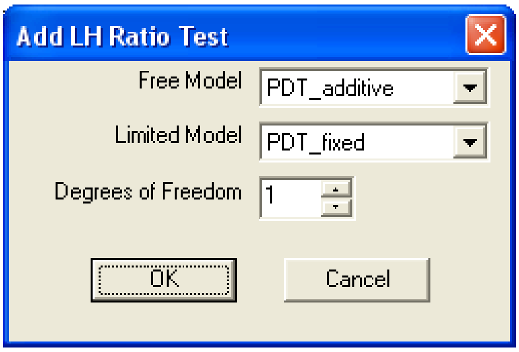
If you have imported the position information successfully, the order of markers changes according to the position information. You can no longer change the order of markers manually by arrow keys.

7.4. Specifying the set of analyses.

TDT, PDT, Segregation analysis and model based linkage (also JSL and JSLD) can be included in the set of analyses. To use likelihood ratio test for PDT or linkage, you should specify maximization for both models, included in the LRT.

7.4.1. Specifying LRT.

For all linkage or disequilibrium analyses, using maximum likelihood estimation (MLE), you should specify at leas two models of the same type: more general and constrained. To construct LRT click on menu item “Serial”-“New”-“LH Ratio”. The “Add LH Ratio Test”-dialog opens. The upper combobox, named “Free model” contains the name list of all MLE models (besides segregation), which were already defined. Select the model, which you will use as more general. Then in the lower combobox, named “Limited model” you will get a name list of all models of the same type as that, selected in the upper combobox, besides itself (if there is not any model of the same type, the list will be empty). Select from the list the model that you are going to use as constrained model. Then specify in “Degrees of Freedom”- editbox the difference in df between upper and lower models and click “OK”. The new LRT named <free model>/ <limited model> appears in the document tree item ANALYSIS. Now it is possible to use p-value of this LRT for filtering results.



Add LH Ratio Test

Free Model: PDT_additive

Limited Model: PDT_fixed

Degrees of Freedom: 1

7.4.2. TDT.

. Click “Analysis”-“Linkage”-“TDT”. The TDT window opens. If you haven’t specified derived traits, only one option “Trait” is available in “Analyzed trait”-comboboox. In “Marker”-combobox you can select “Marker1” or “Haplotype”. In the MODEL FORMULATION control group select “Type of sampling”, for example “Extreme”. If you want to calculate QTDT and/or FBAT, click on “Analysis”-“External”-“QTDT” and/or “Analysis”-“External”-“FBAT” correspondingly. The external tests should appear in the parameter table. In the “MODEL NAME”-editbox (below the MODEL FORMULATION control group) specify the name. It is worse to specify meaningful names for example “TDT extreem” or “TDT haplotypes”, if you have selected “Haplotype” as marker. Then click on menu item “Serial”-“Add analysis”. The TDT-window closes and in the document tree under the item ANALYSIS appears the name of new analysis that you have specified in “MODEL NAME”-editbox.

Double click on the model name in the document tree opens the analysis window with options that you have specified. You can change some options. To save changes click menu item “Serial”-“Add analysis”. If you have not changed the MODEL NAME, you will be asked whether you want to create new model or to correct the old. Otherwise the new model will be created.

7.4.3. PDT.

If the sample contains twins, specify zygosyty in "Twins ID" combobox. If trait defining zygosyty code is not available, delete all already determined analysis and use **Serial-Define Zygosity** menu item (see section 7.1.1).

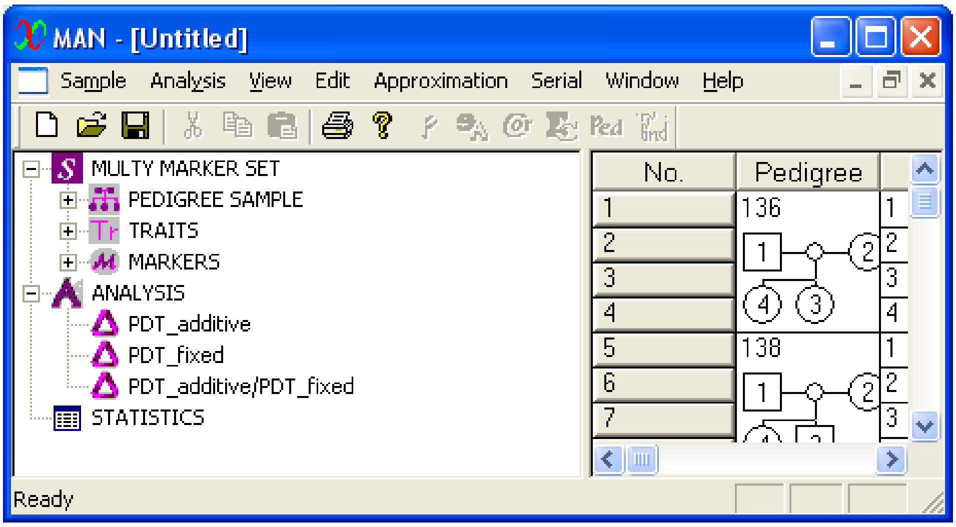
For PDT analysis we should specify at least two models so that one can be treated as general and the other as constrained model. Then the LRT of these two models should be distributed asymptotically as χ^2 with number of df equal to number of additional constraints in the second model.

For example, if we don't include sex effect and compare model with three free genotypic values μ_1, μ_2, μ_3 versus model where μ_1=μ_2=μ_3 , χ^2 distribution with number of df=2 should be used for LRT.

Often we are interested in the presence or absence of significant regression of phenotype on the number of selected marker alleles (significance of additive component) accounting for familial correlation. In this case we use as more general model the additive model, where μ_2=(μ_1+ μ_3)/2, and as constrained - the model where μ_1=μ_2=μ_3 (χ^2 distribution with df=1).

As one example, let specify the last analysis. Click “Analysis”-“Linkage”-“Marker segregation” menu item. The analysis window opens. In the parameter table of the window click on the “μ_2”-button. Select from the popup menu the item “Add” (additive model), the constraint button appears right to the “μ_2”-button. Then specify in MODEL NAME edit box the name “PDT_additive” and click menu item “Serial”-“Add analysis”. The new model appears in the document tree. Click once more on “Analysis”-“Linkage”-“Marker segregation” menu item. In the analysis window apply two constraints. Click on the “μ_1”-button and select “=μ_2” from the popup menu. Click on the “μ_3”-button and select “=μ_2” from the popup menu. Specify in MODEL NAME edit box the name “PDT_fixed” and click menu item “Serial”-“Add analysis”.

To construct LRT click on menu item “Serial”-“New”-“LH Ratio”. In “Add LH Ratio Test”-dialog select in the upper combobox the more general model “PDT_additive” and in lower the restricted model “PDT_fixed”, then specify in “Degrees of Freedom”- editbox 1 (the difference in df between upper and lower model) and click “OK”. The new LRT “PDT_additive/ PDT_fixed” appears in the document tree and now it is possible to use p-value of this LRT for filtering results.



7.4.4 Segregation models.

The segregation models without covariates can also be specified, to use it as base for model based linkage, JSL and JSLD. As a rule the building of most parsimonious Mendelian segregation model is specific for each trait. So for model based linkage and JSLD analyses it is worse to use this option, if you have to analyze only one trait and a set of markers. You need to know, what constraints should be applied for most parsimonious model. It means that preliminary segregation analysis should be done for the trait to test the MG effect and the most parsimonious model should be built.

For JSL you can use any Mendelian segregation model as a base for analysis. The model will be estimated once for each quantitative trait and the estimated parameters will be used as initial values for JSL analysis with each marker.

Click “Analysis”-“Segregation”-“Conventional”. The segregation analysis window opens. If you have specified some derived traits, select the desired trait in the “Analyzed trait”-combobox. Specify options of most parsimonious model in MODEL FORMULATION control group. Specify in the parameter table all constraints. Define the model name and click menu item “Serial”-“Add analysis”.

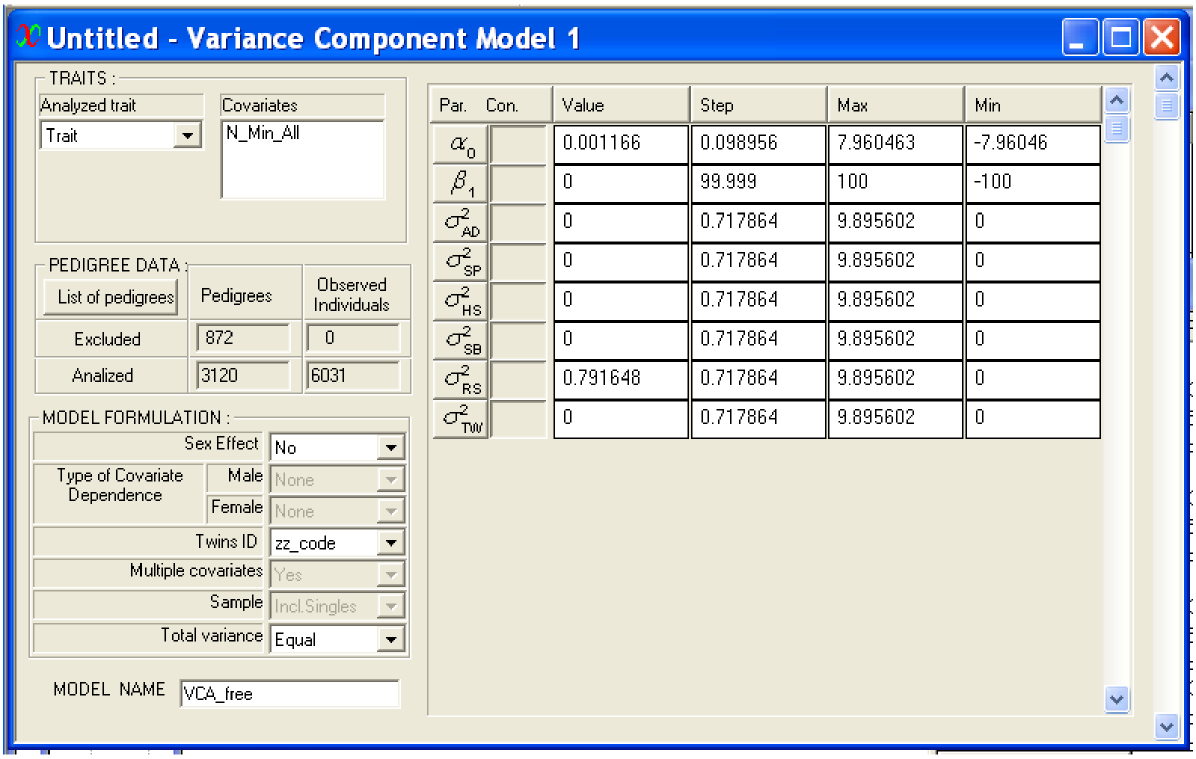
7.4.5. Model based linkage.

The segregation models should be specified before model based linkage, JSL and JSLD anlyses. If there is no at least one segregation model in the ANALYSES item of document tree, the corresponding menu items in ”Analysis”-“Linkage”-menu are disabled. So, specify all desirable segregation models (7.4.4) and then click on menu item-”Analysis”-“Linkage”-“Model based”. The analysis window opens. If you have specified a number of segregation models, the list of models is available in “Take initial values from”-conbobox (below the parameter table). Select the desirable model. The analysis window will change in accordance to properties of the model that you have selected. Now define in “Marker”-combobox “Marker1” or “Haplotype”. Define the linkage analysis in the MODEL FORMULATION group (see 6.1.3). Define constraints, if you need, in the parameter table. Type the model name in the MODEL NAME edit box and click menu item “Serial”-“Add analysis”. To use the linkage test, you should specify at least two models: free and restricted. Then define the new LRT using “Serial”-“New”-“LH Ratio”- menu item as described in 7.4.1.

Analogously you can specify JSL and JSLD analysis. See 6.2 and 6.3 for model properties and LRT construction.

7.4.6. Variance component analysis.

In section 5.2.2 we discuss the using of VC-analysis for testing disequilibrium between trait and genetic markers. This option can be used in serial analysis. Click in the Serial Analysis window "Analysis"-"Variance Component"-"Univariate" menu item. Specify "Twin ID", if twins are included in the sample. In the analysis window only one covariate "N Min All" is presented in the covariate list box. This covariate denote the number of minor alleles of diallelic marker. For each marker, included in serial analysis this integer trait will be produced automatically and its regression on each quantitative trait, included in serial analysis will be tested. The parameter β_1 is the regression coefficient of the trait on the number of alleles. To test the significance of association between trait and marker two analyses should be specified: the first with free β_1 and the second with constrain β_1=0. The LRT for this two analyses should be defined as described in 7.4.1 with 1 DF.



TRAITS :

Analyzed trait	Covariates
Trait	N_Min_All

PEDIGREE DATA :

List of pedigrees	Pedigrees	Observed Individuals
Excluded	872	0
Analized	3120	6031

MODEL FORMULATION :

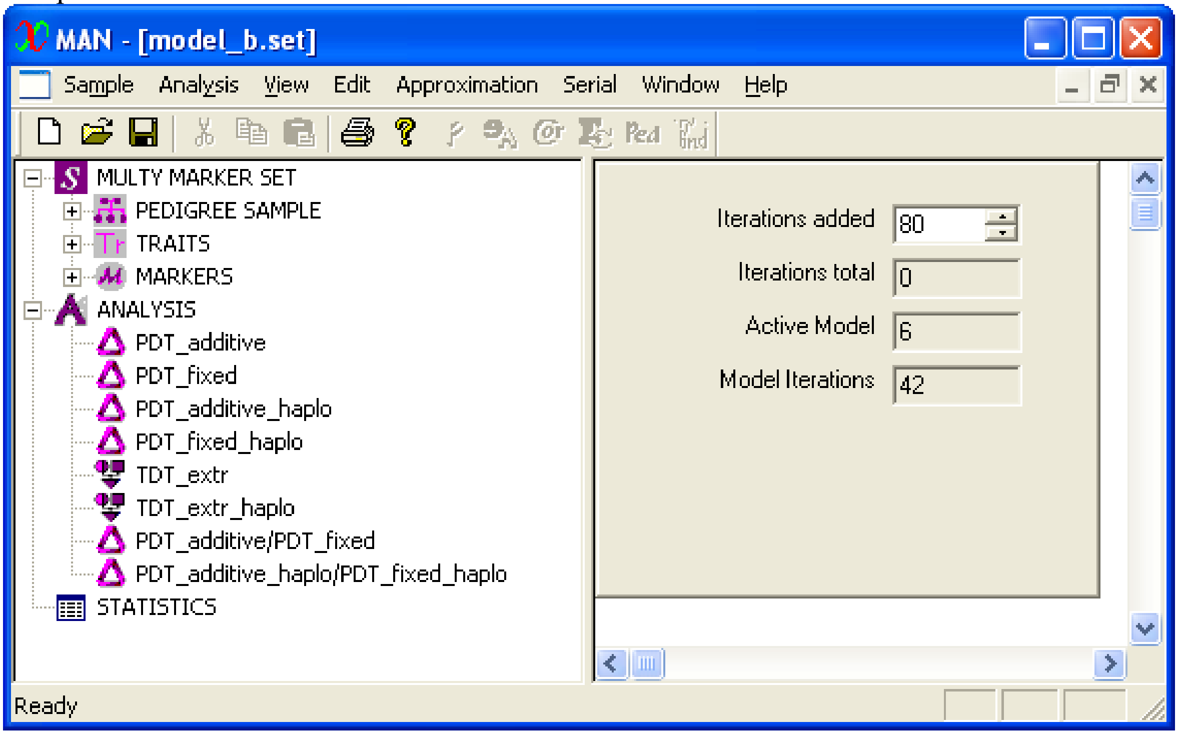
Sex Effect		No
Type of Covariate Dependence	Male	None
	Female	None
Twins ID		zz_code
Multiple covariates		Yes
Sample		Incl.Singles
Total variance		Equal

MODEL NAME VCA_free

Par.	Con.	Value	Step	Max	Min
α_0		0.001166	0.098956	7.960463	-7.96046
β_1		0	99.999	100	-100
σ^2_{AD}		0	0.717864	9.895602	0
σ^2_{SP}		0	0.717864	9.895602	0
σ^2_{HS}		0	0.717864	9.895602	0
σ^2_{SB}		0	0.717864	9.895602	0
σ^2_{RS}		0.791648	0.717864	9.895602	0
σ^2_{TW}		0	0.717864	9.895602	0

7.5. Running serial analysis.

When you have specifyed all models, which you need, click on menu item “Serial”-“Run”. The computation starts. In the right side of document window you will see the current computation statistics.



Iterations added	80
Iterations total	0
Active Model	6
Model Iterations	42

Ready

The total number of iteration is computed as number of traits, N_T, multiplied on number of markers, N_M. If some segregation models were specified the total number of iterations is

$N_T(N_M+1)$. In the *Iterations added* field you will see the number of residual iterations to be computed. In the *Iterations total* field you will see the number of iterations already fully computed (computed for all models).

Active model is the number of model which is estimated now for current iteration. The models defined for haplotype correspond to four sequential numbers, because four marker factorizations (dichotomous schemes) are estimated. *Model Iterations* shows the number of iterations of single maximization. When the number of fully computed iterations is nonzero, it is possible to see parameters of already estimated model. To do this, click on "View"-"Model Parameters" menu item. In the right side of the document window you will see the table of parameters corresponding to the upper model in the ANALYSIS item of the document tree. To view any other model, click on its name in the tree. The contents of the table in the right sub-window will change correspondingly. The number of strings in the table corresponds to number of MLE estimations already done for the model. Models estimated for haplotype have four times more strings. To return to computation statistics click once more "Serial"-"Run".

If the number of traits and markers is sufficiently high, the computation can take a lot of time. You can stop the computation at any time. To do this, click on menu item "Serial"-"Cancel". After you see the message, that the computation was stopped, you can save the serial document and close the program. When you are ready to continue the calculation, open the serial analysis document and click "Serial"-"Run". The serial analysis use for computation the data from the sample (smp file), for which it was created. For this purpose it saves the full path to this file. If this file was moved or cannot be find using this path, you will be asked to find the corresponding smp file manually. If the data of used trait were not changed, the computation continues.

If for some trait –marker pair the maximization ends with error, the serial analysis stops. In this case you should delete the appropriate trait from the trait list and start the analysis once more (see the paragraph below). The trait that caused error can be analyzed individually in the smp-file, because here you can edit initial parameter values and parameter limits.

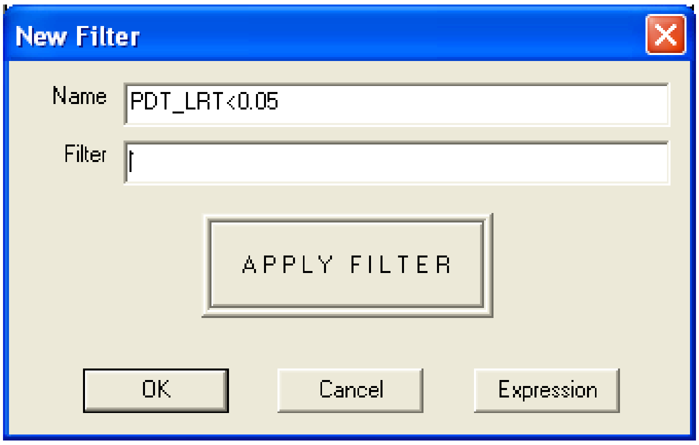
When the number of iterations is nonzero you cannot do any changes in the trait or marker list and in the list of specified models. If you need to do this, stop the computation, then click "Serial"-"Clear Iterations". Because all results already computed will be deleted, you should confirm the performance of the command. Then you can change the definition of serial analysis.

7.6. Viewing results.

7.6.1. Tables of estimated model parameters.

When the serial computation is over, you can see parameters estimated for each model as tables. To view them, click on "View"-"Model Parameters" menu item or on "P" icon on the toolbar. In the right side of the document window you will see the table of parameters corresponding to the upper model in the ANALYSIS item of the document tree. To switch between models click on the appropriate model name in the ANALYSIS item. The structure of all tables is similar. The four first columns contain the name of trait, the name of marker, selected allele and its frequency. For haplotype analysis the "Marker" column contains the names of adjacent markers, separated with "&"-symbol. The "Allele" column in this case contain the denotation of selected haplotype: alleles of the appropriate markers separated with "|"-symbol. For MLE-models the fives column, named "LH" contains the maximum log-likelihood of the model. Other columns (up to the end of the table) contain the estimated model parameters. The name of parameter is shown in the header of the column. By default the tables are sorted through markers in the order, coinciding with that of marker list in the document tree. To sort the tables through trait, click on the header of the “Trait” column.

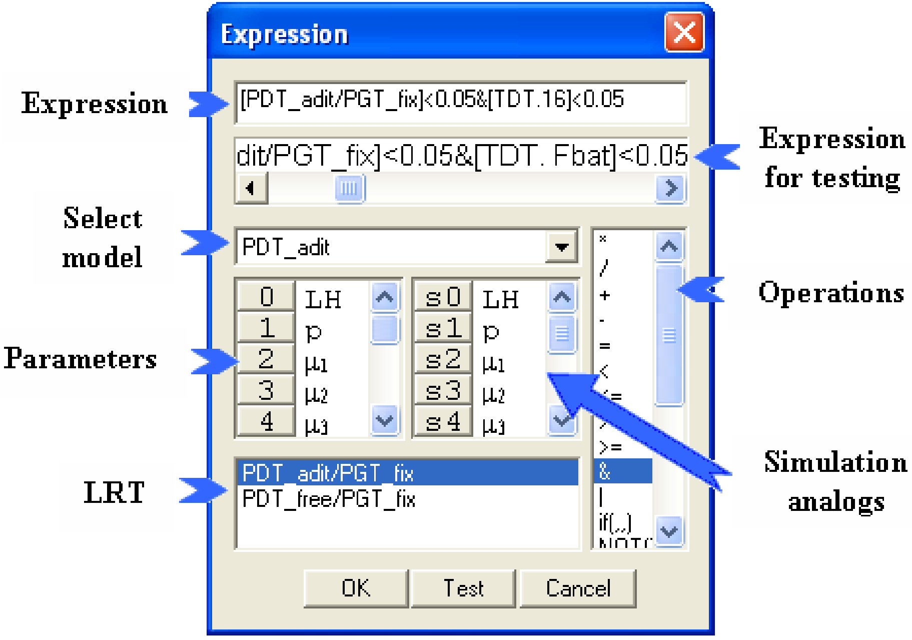
7.6.2. Constructing and applying filters.



Name: PDT_LRT<0.05

Filter:

The number of strings in serial analysis results may be very large. To view only results of interest you can build a logical expression, including parameters of all models and/or p-values of all defined LRTs, to use it as filter for results. To construct a filter click on the menu item “Serial”-“New”-“Filter”. The “New Filter”-dialog opens. Type the name of the filter in the “Name”-edit box. It is worse to give meaningful names, because the name will be displayed in the document tree. To construct the filter expression click on “Expression” button. The “Expression”-dialog opens. In 1.6.1 we described the creation of traits by formula. There the expression dialog is used to construct the formula for computing the new trait. Using this formula the trait is computed for each individual string (the names of traits in the expression are substituted for individual trait values and the expression is computed). Now we should define an expression, which should be computed for each iteration. This form of the expression dialog is used both for serial analysis and for simulations. So instead individual trait values as in 1.6.1 we include in the expression the definite parameters of the models, that are available in the document tree and were estimated for each iteration (corresponding to appropriate trait-marker pair in serial analysis or to single iteration in simulations).



[PDT_adit/PGT_fix]<0.05&[TDT.16]<0.05

dit/PGT_fix]<0.05&[TDT. Fbat]<0.05

PDT_adit

PDT_adit/PGT_fix
PDT_free/PGT_fix

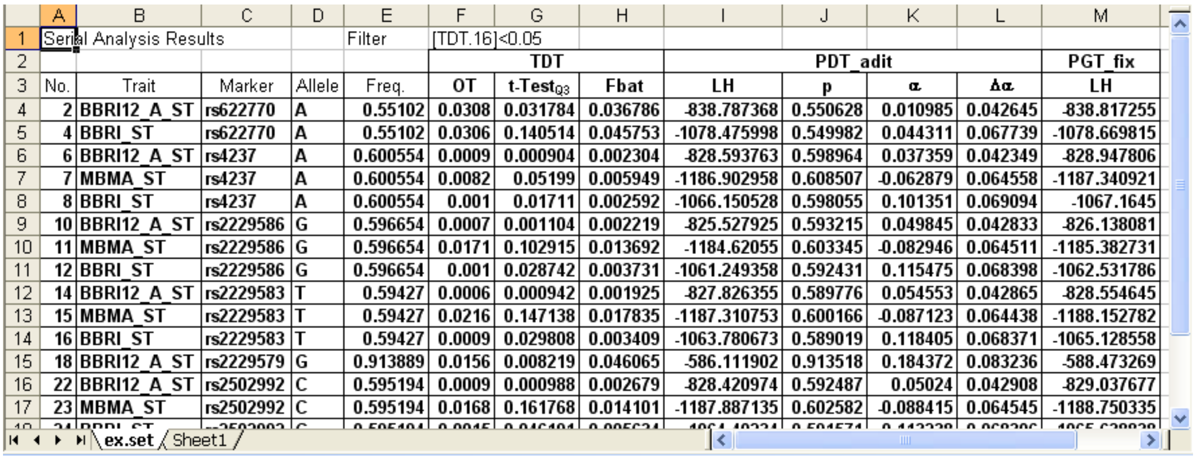
The syntax of the simple expression including only one parameter is as follows:

[<model_name>.<parameter_Number>].

The parameter number 0 corresponds to the model log-likelihood. The p-value of LRT is included in the expression as [<LRT_name>]. The expression using the defined syntax should be formed in the upper editbox. The combobox below is used to define the model, from which you want to choose parameter for the expression. The list of all defined models is available in the combobox. Two listboxes below the combobox contain: the left - all parameters of the appropriate model (as a rule estimated in MLE procedure) and the right - so called “simulation analogs”(SA),if they are defined for the appropriate parameter. For example in PDT the appropriate genotypic values μ_1, μ_2, μ_3 are estimated in MLE procedure. For each trait-marker pair the SA for each of genotypic values is the mean value of trait for the individuals having the appropriate marker genotype. For TDT no MLE is performed and no SA for the parameters are defined (the second combobox is empty). The SA parameters can be used for simulations, it is not recommended to include them in expression for serial analysis. When the selected in the combobox model changes, the contents of both listboxes below changes appropriately. In the parameter listbox you see all parameter names as defined in the parameter table of the model. In the left side of the listbox you can see buttons with parameter numbers. Double click on the parameter item in the listbox copies [<model_name>.<parameter_Number>] (the parameter denotation) in the upper edit box, where expression is formed. The lowest listbox contains the list of all defined LRTs. And each LRT p-value also can be copied in the expression as [<LRT_name>] by double click on it. You can combine in one expression LRT p-values and parameters of different models using logical (&-and,|-or), comparison (<,<=,>,>=) and other operators and functions. The list of mathematical functions is the same as in 1.6.1 and is available in the most right listbox. Double click on the appropriate item copies the item in the expression (into the upper editbox). To test the expression, which you have built, click on “Test”-button. If there are some syntax errors in the expression, you will see the error message; otherwise in the control below the expression-editbox you will see the expression in more testable form. All parameters will be given in form [<model_name>.<parameter_Name>].

If the expression is formed right, click OK. The expression dialog closes and the expression appears in the “Filter”-editbox. The type of expression for the filter should be logical. If you click on OK button, the filter will be not applied. It only will be saved and appears in the document tree as sub-item in the STATISTICS item. If you want to see only those results, for which the filter value is true, click the “APPLY FILTER”-button and you will see in the parameter tables for all models only appropriate strings. If you have defined a number of filters only one of them can be applied at once, its icon will be red. To change the current applied filter double click on the other filter in the document tree and when the filter dialog opens click on the “APPLY FILTER”-button. If you want to remove action of any filter, click on “View”-“Remove Filter” menu item.

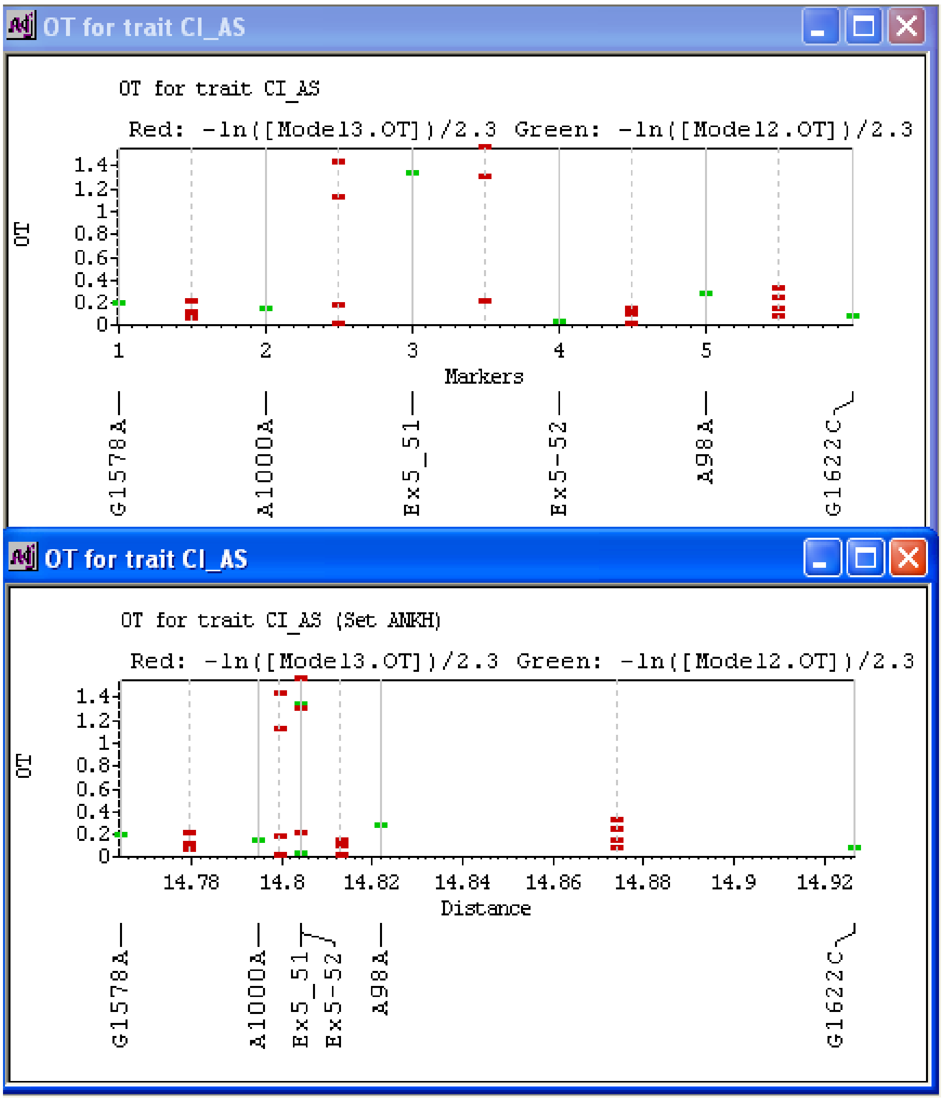
7.6.3. Exporting results to Excel.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Serial Analysis Results				Filter	[TDT.16]<0.05							
2						TDT			PDT_adit				PGT_fix
3	No.	Trait	Marker	Allele	Freq.	OT	t-Test$_{Q3}$	Fbat	LH	p	α	Δα	LH
4	2	BBRI12_A_ST	rs622770	A	0.55102	0.0308	0.031784	0.036786	-838.787368	0.550628	0.010985	0.042645	-838.817255
5	4	BBRI_ST	rs622770	A	0.55102	0.0306	0.140514	0.045753	-1078.475998	0.549982	0.044311	0.067739	-1078.669815
6	6	BBRI12_A_ST	rs4237	A	0.600554	0.0009	0.000904	0.002304	-828.593763	0.598964	0.037359	0.042349	-828.947806
7	7	MBMA_ST	rs4237	A	0.600554	0.0082	0.05199	0.005949	-1186.902958	0.608507	-0.062879	0.064558	-1187.340921
8	8	BBRI_ST	rs4237	A	0.600554	0.001	0.01711	0.002592	-1066.150528	0.598055	0.101351	0.069094	-1067.1645
9	10	BBRI12_A_ST	rs2229586	G	0.596654	0.0007	0.001104	0.002219	-825.527925	0.593215	0.049845	0.042833	-826.138081
10	11	MBMA_ST	rs2229586	G	0.596654	0.0171	0.102915	0.013692	-1184.62055	0.603345	-0.082946	0.064511	-1185.382731
11	12	BBRI_ST	rs2229586	G	0.596654	0.001	0.028742	0.003731	-1061.249358	0.592431	0.115475	0.068398	-1062.531786
12	14	BBRI12_A_ST	rs2229583	T	0.59427	0.0006	0.000942	0.001925	-827.826355	0.589776	0.054553	0.042865	-828.554645
13	15	MBMA_ST	rs2229583	T	0.59427	0.0216	0.147138	0.017835	-1187.310753	0.600166	-0.087123	0.064438	-1188.152782
14	16	BBRI_ST	rs2229583	T	0.59427	0.0009	0.029808	0.003409	-1063.780673	0.589019	0.118405	0.068371	-1065.128558
15	18	BBRI12_A_ST	rs2229579	G	0.913889	0.0156	0.008219	0.046065	-586.111902	0.913518	0.184372	0.083236	-588.473269
16	22	BBRI12_A_ST	rs2502992	C	0.595194	0.0009	0.000988	0.002679	-828.420974	0.592487	0.05024	0.042908	-829.037677
17	23	MBMA_ST	rs2502992	C	0.595194	0.0168	0.161768	0.014101	-1187.887135	0.602582	-0.088415	0.064545	-1188.750335

If MSOffice Excel is installed on your computer, you can export results into Excel worksheet. To do this you should select for all available models parameter columns, which you want to export. To do this, click on the “P”–icon in the toolbar or on menu item “View”-“Model Parameters”. Click on the desired model name in the document tree; the appropriate parameter table will be displayed in the right side of the document window. To select any parameter, click on its column header. The name in the header appears in red. This means that the parameter is selected. You can select any number of parameters.
To remove selection click on the red parameter header once more, the parameter name will appear in black (the selection is removed). Then click on the next model and select its parameters and so on. When all desirable parameters are selected click menu item “Serial”-“Export to Excel”. The new Excel worksheet will be opened with the table in it. If any filter was applied, only strings, for which the filter value is true, will be exported. The filter expression will be also included into the worksheet. The columns Trait, Marker, Allele, Freq.(the frequency of allele) are always included in the results table.

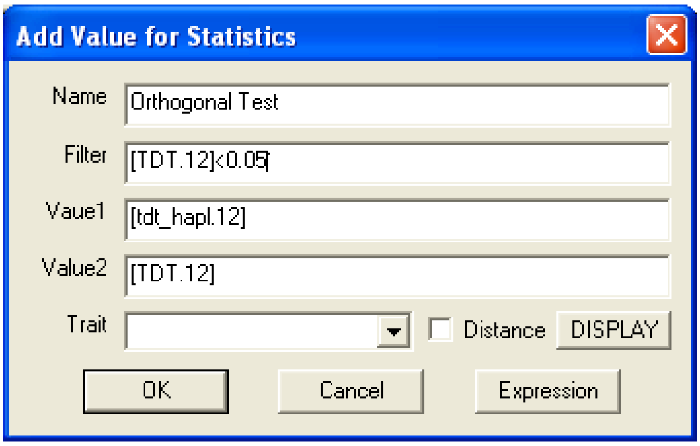
7.7. Options for graphical display.



You may want to see some numerical value (constructed from model parameters and LRT p-values) displayed for all markers together on the same graph. It is suitable, for example, to compare test values for all markers and selected trait. If the logarithm of the test p-value is displayed, the most significant tests are peaks of the graph. Two different values can be displayed on the graph at once. The graph can be displayed in two forms: as dependence on marker number or, if marker positions were defined, as dependence on chromosomal distance. The values for tests/models, defined for adjacent haplotypes, are displayed in the middle between positions of the markers, which form the haplotype. For haplotypees results are displayed for all dichotomous factorizations. The graph can be copied to the clipboard as enhanced metafile and then can be inserted for example into Office documents.

7.7.1. Defining values for graphical display.

To define new graph click "Serial"-"New"-"Stat Value". The *Add Value for Statistics*- dialog opens. In the *Name*-editbox you should specify the name of the graph (it is preferable to give meaningful names). In the *Filter*, *Value1* and *Value2* editboxes the expressions in the same form as described in 7.6.2 should be specified. To do this, click on the appropriate editbox. When the cursor appears in the editbox, click on the *Expression*-button. After the *Expression*-dialog opens, form expression as described in 7.6.2 and click *OK*-button. The expression appears in the appropriate editbox (*Filter*, *Value1* or *Value2*). The value of filter should be logical or empty. If the filter was defined the values will be shown only for those markers, for which the filter value is true. The values defined for *Value1* and *Value2* can be of type real or integer. You may define expression for only one value or for both. In the last case the scale of the graph is the same for both values, this should be taken into account, when you define two values. After you have defined filter and values you can click "OK"-button and save the definition for later using with any quantitative trait of the document trait list. The graph definition appears as sub-item with *S*-icon within the document tree item STATISTICS.



Add Value for Statistics

Name: Orthogonal Test

Filter: [TDT.12]<0.05

Vaue1: [tdt_hapl.12]

Value2: [TDT.12]

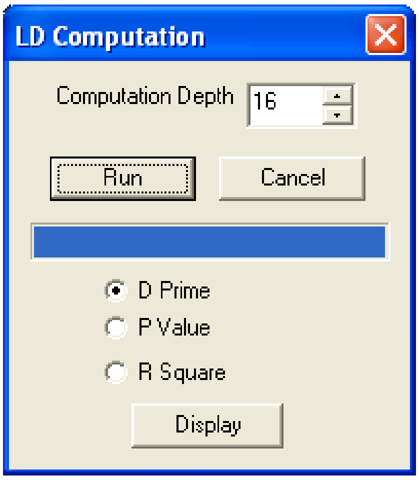
Trait:

7.7.2. Displaying the graph for different quantitative traits.

If you want to see the graph immediately, select the quantitative trait in the *Trait*-combobox. If you want to see horizontal scale as the chromosomal position, mark the *Distance*-checkbox. Then click *DISPLAY*-button. The graph window opens. If you have checked the *Distance*-checkbox and the marker list contains markers positioned on different chromosomes, you will see a number of graph windows (separate graph for each chromosome having more then two tests to represent, accounting for filter, if defined). Simultaneously the graph definition appears as sub-item with *S*-icon within the document tree. If you want to see the analogous graph for another quantitative trait, double click on this sub-item in the document tree. In the opened dialog you can specify the desirable trait and click *DISPLAY*-button, to open the graph window (or windows) for this trait. When you double click any sub-item of STATISTICS-item in the document tree you can also edit its name and defined expression, if you need. If you have a number of graph windows opened, you can see them simultaneously. To do this, minimize all document windows and then click the menu item "Window"-"Tile". If you want to equalize scales, you can edit each graph by clicking "Edit"-"Edit Curve" as described in 1.9.3. You can copy the graph into clipboard. To do this click "Edit"-"Copy" menu item, when the graph window is active.

7. 8. Disequilibrium between markers.

To compute disequilibrium between markers click "Serial"-"Markers"-"LD Computation". In the "LD Computation"-dialog you should specify only one parameter *Computation depth*. This is the size of marker neighborhood, for which LD computation will be done. If the depth is equal to 1, LD is computed only for adjacent markers. If depth is *n*, the computation is done so, that for each marker the LD results for *n* adjacent to the left and *n* adjacent to the right markers in the marker list will be computed. LD coefficients are computed for pairs of markers. P-value for significant LD is computed for LRT as described in 6.4. To start the computation click the "Run"-button. You can see the progress of computation. If the number of markers is great and the computation depth is significant, the computation can take a relatively long time. You can stop the computation by clicking "Cancel"-button.



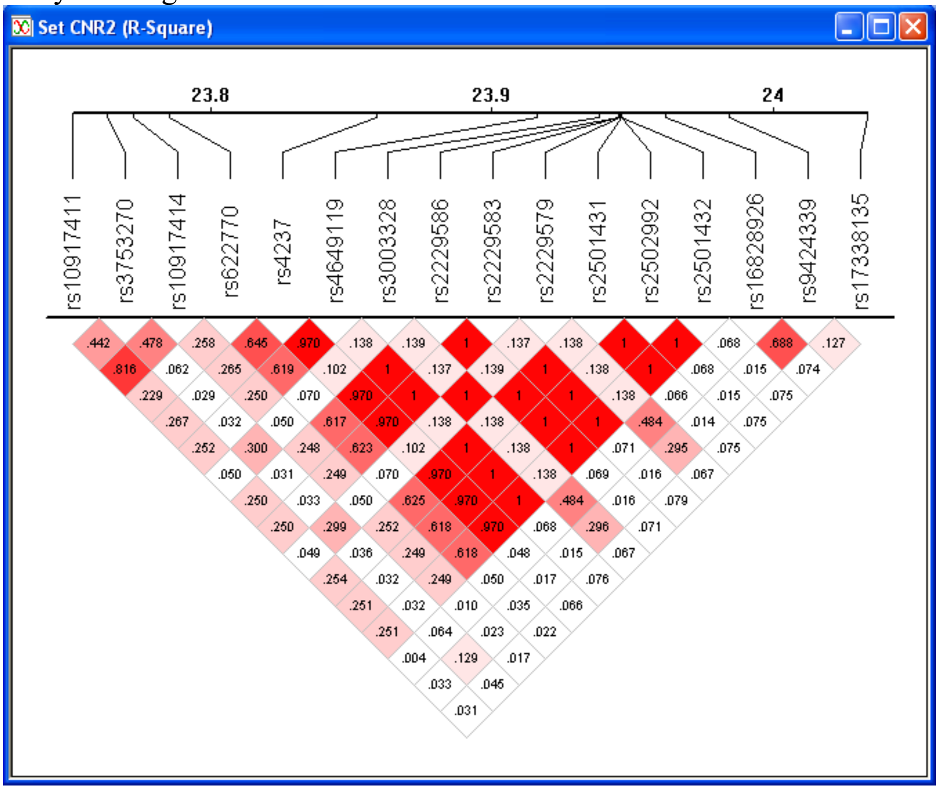
LD Computation

Computation Depth 16

● D Prime

○ P Value

○ R Square



Set CNR2 (R-Square)

After that the serial analysis document can be saved and closed. When you open the file once more, you can continue with LD computation from the point, where you have stopped. To do this click “Serial”-“Markers”-“LD Computation” menu item and then click “Run”-button in the “LD Computation”-dialog. When the computation is over the results can be displayed as graph. There are three options available: *D prime*, *R square* and *P value*. Select the desirable option and click “Display”-button. The graph window opens. If the chromosomal positions were specified for the markers, they are shown in the upper part of the graph.

If your marker list contains markers, positioned on different chromosomes, a number of graph windows will be opened (separate graph for each chromosome, having at least two markers to represent). The graph window is resizable and you can change its width and height to get the most suitable graph according to number of markers and computation depth. You can copy the graph into clipboard. To do this, click the menu item “Edit”-“Copy”, when the graph window is active.

8. Simulation.

MAN includes the possibility to perform simulations for pedigree based studies. The most often aim of simulation is to investigate power of tests as a function of sample size and structure, genetic and environmental parameters of the trait model and marker properties. We also can test the type I error level for samples of definite structure and get critical values for the cases, when we suppose possible deviation from the asymptotic null-distribution. It is also possible to get any statistics including parameters of estimated models and statistics of simulated sample realizations.

The structure and the number of the simulated pedigrees in each sample can be defined either by sequential visual drawing of a particular pedigree structure with its further multiplication and adding to the sample, or by copying of a pedigree structure from an already existing sample.

The simulated model of the trait inheritance can be formulated in accordance with parameterization options defined in section 3.1 (permitting extension for controlled by two loci quantitative traits). Up to 6 genetic loci and one or two quantitative traits can be simulated.

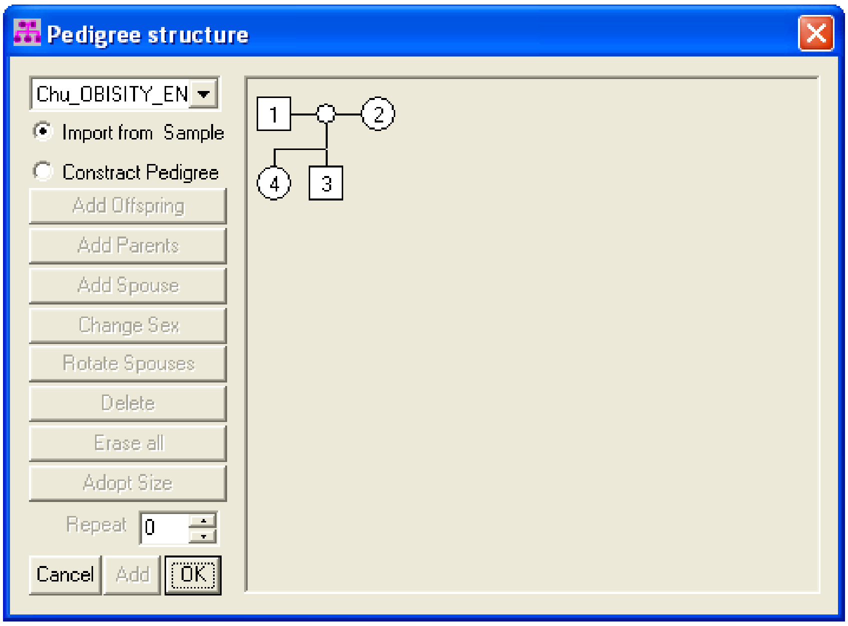
The design of simulation study includes the set of analyzing models, methods of their testing and LRT comparisons.

8.1 New simulation document.

8.1.1 The sample pedigree structure.

If you need to perform simulation study for a sample, for which you already have a smp-document file, open it. You will copy from this file not only the pedigree structure but also array of missing individuals for quantitative and genetic marker traits. This information is taken from two last real traits in the trait table. The last trait will be used for individuals with missing quantitative trait values, the trait before it - for missing genetic markers. So if you want to do simulation with all individuals measured, create to new real traits, which have no missing values. If you want to use some existent marker and existent quantitative trait (for example, named “Marker1” and “Trait1”) as a standards for missing individuals in simulated markers and traits, create to new traits as follows. Use “Edit”-“Add Trait”-“Formula” menu item (see 1.6.1) and create the first trait by formula *if(IsMis(Marker1),[MissVal],1)* and the second trait by formula *if(IsMis(Trait1),[MissVal],1).*

Click menu item “Serial”-“New”-“Simulation”. The “Pedigree structure”-dialog opens.



The radio buttons in the left upper corner of the dialog enable two options *Import from Sample* and *Construct Pedigree*. If some samples (smp-files) are already opened, *Import from Sample* is selected by default and you can select the desired sample from the combobox above the radio buttons. In the right part of the dialog you can see the pedigree, which index (zero based) appears in the edit box in the left lower corner of the dialog. By changing this index, you can look through all pedigrees in the sample. If you want to copy pedigree structure from the selected sample, click the “OK”-button.

If you want to build the pedigree structure for simulation manually select *Construct Pedigree* and perform the procedure described in 1.1.2.

After clicking “OK”-button the dialog closes and the new simulation document opens. The new pedigrees appear in the right side of the document window.

If you click in “Pedigree structure”-dialog the “Cancel”-button, an empty simulation document opens. You can close it or click “Serial”-“Pedigrees”-“Construct” menu item to try once more to create the pedigree structure.

8.2 Model of traits for simulation.

When the pedigree structure is defined, the trait model for simulation can be formulated. As mentioned above, six genetic loci (Gen1-Gen6) are simulated for each member of the pedigree sample. It is supposed that loci are positioned sequentially on the chromosome and for each pair of adjacent loci recombination fraction should be defined. Certainly, if you define for some interval recombination fraction equal to 0.5, it means that loci before this interval and after are on different chromosomes.

Any two loci from the list can be selected as controlling the quantitative trait in additive manner. For each of two trait controlling loci heritability (H_i^2) and dominance level should be defined. The loci, which control the trait, can be only di-allelic. The frequency of allele, corresponding to minor quantitative phenotype, can be defined with discreetness 0.1 (0.1, 0.2, …, 0.9). Other four loci can have up to five alleles. Each locus has two options of allele frequency selection: equal (all alleles have equal frequencies) and proportional (the frequency of each allele is proportional to its number).

Allelic disequilibrium can be simulated for one pair of loci. At least one of loci in the pair should control the quantitative trait. The trait is simulated to have total variance 1. Genetic variance for two loci is $V_{gen}= H_1^2+ H_2^2- H_1^2\cdot H_2^2$; trait residual in the pedigree is simulated to be N-variable normal (N is the number of individuals in the pedigree) with three parameters being partial correlations between parents-offspring, spouses and siblings.

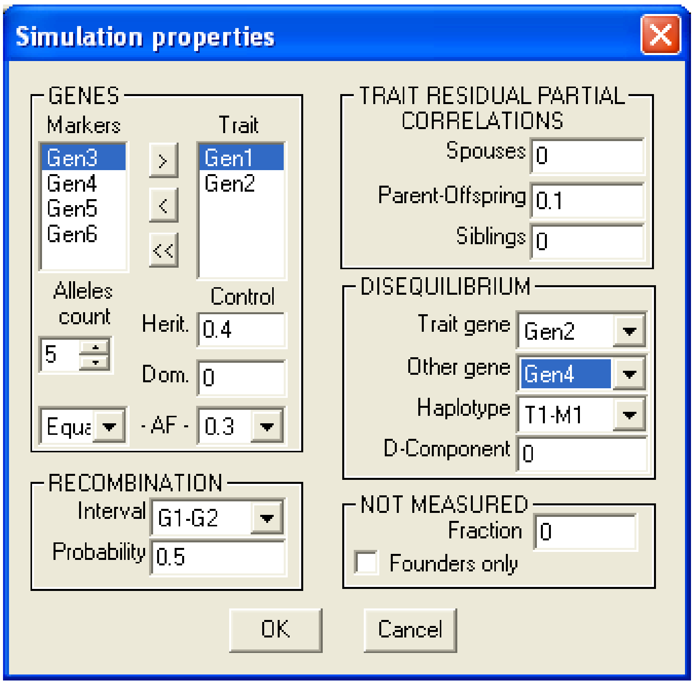
To define the trait model click menu item “Serial”-“Trait Model”. The “Simulation properties” dialog opens.

GENES - control group defines the properties of genetic loci and their relation to the trait control. When you open the dialog at first all loci are in the list named “Markers”. You should move the loci or locus controlling the quantitative trait to the list “Trait”, using buttons “>”,”<”,”<<” positioned between two lists.

Now, to specify loci properties for each member of “Markers” list, select the appropriate locus in the list (it appears in blue) and specify for it the count of alleles in the editbox below the list. Then in the combobox below (“AF”) specify the type of allele frequency simulation: equal or proportional. Then select in the “Markers” list the next locus, its parameters appear in the controls below the list and you can edit them.

To specify the properties of loci in the “Trait” list select the upper locus. Specify the part of variance controlling by the locus (non-negative value less then 1) in the “Herit.”-editbox. Specify the extent of dominance in the “Dom.”-edit box (value from -1 to 1). 1

correspond to heterozygous genotypic value equal to lower phenotype homozygote, -1 - heterozygous genotypic value equals to upper phenotype homozygote, 0 –additive model (heterozygous genotypic value equals to mean of two homozygous). In the “AF”- combobox below the “Trait” list specify the frequency of allele corresponding to lower genotypic value homozygote. If you have two loci in the “Trait” list, select the lower one and specify properties for it. You should taken into account that the value, that you specify in the “Herit.”- editbox for the second locus (H_2^2) denote the part of the residual variance ($1-H_1^2$), which it controls, so the part of trait variance, which controls the second locus is $H_2^2 \cdot (1-H_1^2)$.



Simulation properties

GENES

Markers: Gen3, Gen4, Gen5, Gen6

Trait: Gen1, Gen2

Alleles count: 5

Equa

Control

Herit. 0.4

Dom. 0

- AF - 0.3

RECOMBINATION

Interval: G1-G2

Probability: 0.5

TRAIT RESIDUAL PARTIAL CORRELATIONS

Spouses: 0

Parent-Offspring: 0.1

Siblings: 0

DISEQUILIBRIUM

Trait gene: Gen2

Other gene: Gen4

Haplotype: T1-M1

D-Component: 0

NOT MEASURED

Fraction: 0

□ Founders only

RECOMBINATION- control group defines recombination probability between loci with adjacent numbers. By default the recombination fraction for all five intervals is 0.5. Select the appropriate interval in the “Interval”-combobox. Then in the “Probability”-editbox specify the recombination probability for the selected interval.

TRAIT RESIDUAL PARTIAL CORRELATIONS - control group defines parameters of trait residual normal distribution. Partial correlations for spouses, parent-offspring and siblings should be specified in the appropriate editboxes. This can simulate the additional environmental or polygenic phenotype correlation. If you have sufficiently great families this values should be restricted (their order should be 1/N, where N is the maximal size of the pedigree), because the inverse matrix (the correlation matrix) should be positively defined. If for some pedigree the correlation matrix is not so, the trait values for that pedigree cannot be calculated (missing values will apear for the whole pedigree).

DISEQUILIBRIUM - control group defines for what pair of loci allelic disequilibrium should be simulated. As mentioned above at least one of loci should appear in the “Trait” list of the GENES - control group. In the “Trait gene”- combobox select the desirable locus. Then in the “Other gene”- combobox you can select any of other five loci. If the number of alleles of the “Other gene” is K, you should specify K-1 independent parameters of D prime matrix (by default they are 0). To do this select the appropriate pair of alleles in the “Haplotype”- combobox and specify the appropriate D’$_{ij}$ parameter (value between 1 and -1).

NOT MEASURED - control group can be used only if you have constructed the sample pedigree structure manually (pedigree structure was not imported from the existing smp-file). For this case you can specify the proportion of not measured individuals in "Fraction"-edit box.

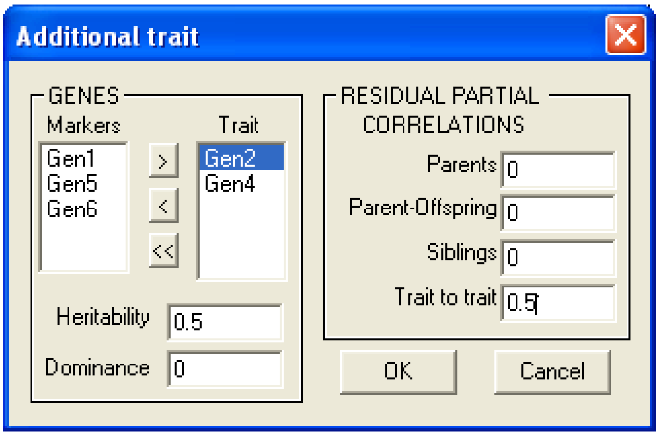
Then after the simulation of the trait the defined proportion of randomly selected individuals will get missing trait values. If you check "Founders only", the defined proportion will be applied only to founders. So you can for example simulate the sample of siblings without measured parents, which is often the feature of sampling design. Check "Founders only" and specify "Fraction" of not measured founders as 1).

After specifying all features of the trait-marker model click "OK"-button. The simulated realization of quantitative trait "SimTrait" six markers "Gen1"-"Gen6" appear in the trait table (right part of the document window).

8.2.1 Second trait for bivariate analysis.

If you are going to do simulation for bivariate analysis (section 4), you need additional trait having different genes controlling it and probably environmental correlation with the first trait. If you have already defined properties of genetic loci and quantitative trait "SimTrait", you can define an additional trait as follows.

Click menu item "Edit" - "Add Trait" - "Bivariate (Simulation)". The "Additional trait" – dialog opens.



Markers: Gen1, Gen5, Gen6

Trait: Gen2, Gen4

Heritability: 0.5

Dominance: 0

Parents: 0

Parent-Offspring: 0

Siblings: 0

Trait to trait: 0.5

GENES - control group is analogous to that described above for "Simulation properties"-dialog. In the "Marker" listbox only those loci are presented, which was already defined as di-allelic. Move the desirable loci (one or two) into the "Trait" listbox. Define for each locus the *Heritability* and *Dominance* parameters as described above.

RESIDUAL PARTIAL CORRELATIONS - control group define the additional features for 2N-normal distribution for two trait residuals. Relatives partial correlations for the first trait are taken as defined in the Simulation properties"- dialog. Now you should define analogous parameters for the second trait and additional parameter individual partial self-correlation for two traits. The appropriate editbox is named "Trait to trait". When all parameters are defined, click "OK" button. The dialog closes and additional trait "SimTrB" appears in the trait table.

8.2.2 Derived traits.

You can define a number of additional derived traits using menu items "Edit"-"Add Trait"-"Formula" and "Edit"-"Add Trait"-"Standardized". During the simulation process for each realization after simulation of main quantitative traits "SimTrait" and "SimTrB"(if defined), all derived traits will be recalculated using new data.

8.3 Specifying the set of analyses.

The set of analyses (models and LRTs) is specified analogously to those described for serial analysis in 7.4. Additionally to analyses listed in 7.4, if more quantitative traits than one were defined, you can also specify bivariate analysis models (section 4) for simulation.

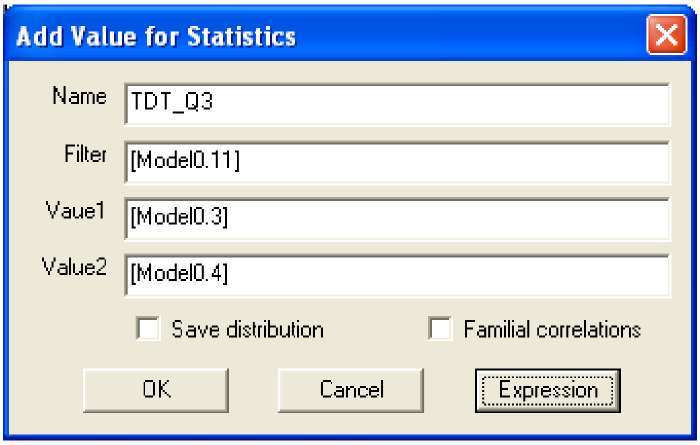
8.4 Specifying statistics.

If you have specified any LRT, you will see the number and proportion of significant tests for it in “Iteration Statistics” window (see below 8.6) automatically. Additionally, for any TDT analysis the number and proportion of significant tests for Q2 (χ^2test), Q3 (*t*-test) and Q5 (F-test) will be shown by default. The desired P-value is specified before the simulation starts, if the “Save Statistics Only” mode was selected. Otherwise any P-value can be specified in the “P-value”- editbox in “Iteration Statistics” window and the number and proportion of significant tests will be recalculated after click on the compute button.

If you want to see some other statistics you should specify their properties as described below.

8.4.1. Defining values for display in statistics window.

Click menu item “Serial”-“New”-“Stat Value”. The “Add Value for Statistics”-dialog opens. In the dialog you should specify the *Name* of statistics. There are three editboxes, which define the way of statistics computation: *Filter, Value1, Value2*. These fields should contain expressions formed from simulation model parameters and LRT p-values. To construct expression in the desirable editbox click on it and when the text cursor appears in the field click on the expression button. The “Expression” dialog opens. The syntax of expressions and working with “Expression” dialog is described in 7.6.2 section. The expression in the *Filter* field should be logical. The types of *Value1, Value2* may be integer or real.



Add Value for Statistics

Name TDT_Q3

Filter [Model0.11]

Vaue1 [Model0.3]

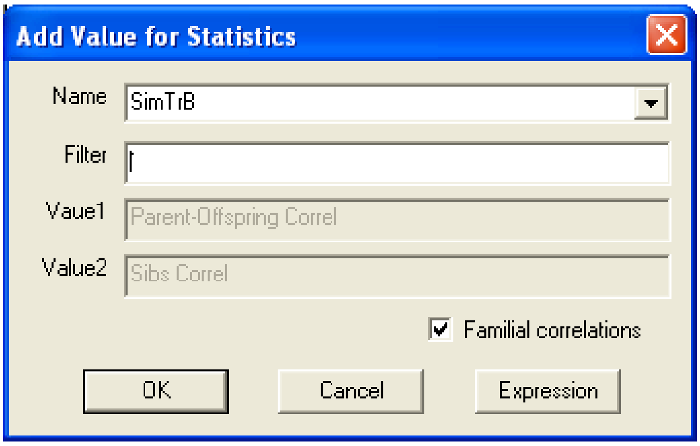
Value2 [Model0.4]

□ Save distribution □ Familial correlations

If the *Filter* field is empty the statistics will be computed using all simulation replicates. If you define only one expression *Value1* or *Value2*, you will see in the “Iteration Statistics” window the mean value of defined expression and its variance. If you specify both *Value1* and *Value2*, you will see means and variances for both expressions and additionally correlation coefficient between them. If you specify *Filter*, the statistics will be computed using only those replicates, for which the value of the filter is true and additionally you will see in the “Iteration Statistics” window the proportion of simulation replicates, for which the filter expression was true. If you specify only the *Filter* expression and both fields *Value1* and *Value2* are empty, only proportion of true will be displayed.

If you check the “Save distribution” combobox, the distribution of the expression *Value1* (or *Value2*, if *Value1* is empty) is computed and you will see critical values (percentiles) for 1, 2.5, 5, 25, 50, 75, 95, 97.5, 99 percents.

There is additional specific type of statistics that also can be specified. This is statistics for parent-offspring and sibling correlations, computed for each realization of any defined quantitative trait. To specify this statistics check the “Familial correlations” checkbox. The appearance of the dialog varies. In the upper part of the window appears the “Name”-combobox, in which you can select the desirable quantitative trait from the list of traits already defined. If you specify the *Filter* expression, the statistics will be stored only for simulation replicates, for which the filter value is true. If you need this type of statistics, you should specify it before the simulation starts, because current trait realizations are not saved.


Add Value for Statistics

Name: SimTrB

Filter:

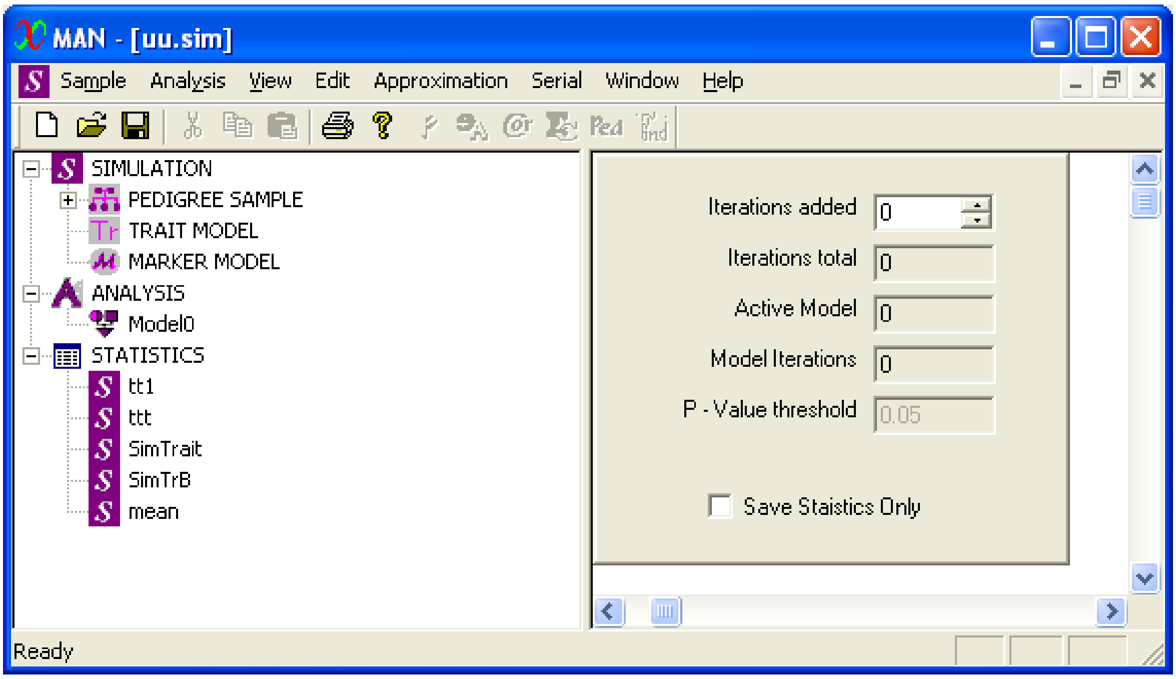
Vaue1: Parent-Offspring Correl

Value2: Sibs Correl

☑ Familial correlations

8.5 Running simulation.

When you have specified the desirable set of analyses, click on the menu item “Serial”-”Run”.


MAN - [uu.sim]

Iterations added: 0

Iterations total: 0

Active Model: 0

Model Iterations: 0

P - Value threshold: 0.05

□ Save Staistics Only

Ready

In the right side of the document window appears the dialog, where you should specify the number of simulation replicates (in the “Iterations added”-editbox) and the type of results saving. If you don’t check “Save Statistics Only”-checkbox, all parameters of specified models and likelihoods will be saved in the simulation document. In this case you can specify new statistics (excluding familial correlations) at any moment of simulation, because all parameters to compute any expression statistics are available. The negative feature in this case is that, if the number of simulation replicates is sufficiently great, the size of the document file on the disk grows rapidly. If you have checked “Save Statistics Only”, the size of the document will not grow, but you will be able to see only those statistics, that you have specified before the start of simulations. If you have checked “Save Statistics Only”, the “P-Value threshold” editbox will be available and you should specify P-value, for which you want to see the proportion of significant tests by default (see 8.4).

If the number in the “Iterations added”-editbox is specified (non-zero), you should click once more “Serial”-”Run”. The simulation starts. In the dialog you will see the number of simulation replicates already computed an added to the document. If this number is not zero, you can not change the pedigree structure, trait and marker simulation properties and the set of analyses.

At any moment you can see the current statistics. To do this, click menu item “View”-“Statistics”.

If the model parameters are saved for all simulation replicates (“Save Statistics Only” was not checked), you can see the parameters of all already computed simulation replicates as tables. To view them, click on “View”-“Model Parameters” menu item or on “P” icon on the toolbar. In the right side of the document window you will see the table of parameters corresponding to the upper model in the ANALYSIS item of the document tree. To switch between models click on the appropriate model name in the ANALYSIS item.

To view the current realization of traits and markers click on PEDIGREE SAMPLE icon in the document tree.

To view the number of iterations, which are to be computed, click menu item “Serial”-“Run”.

You can stop the simulation by clicking menu item “Serial”-“Cancel”. After that you can save and close the simulation document. When you are ready to continue the simulations, open the simulation document (file extension “.sim”). Click menu item “Serial”-“Run”, in the right side of the document window you will see the dialog. The “Iterations total”-edit box shows the number of simulation replicates, already saved in the document. Specify in the “Iterations added”-editbox the number of replicates, which you are going to add and click “Serial”-”Run”. The simulation process will be continued.

8.5.1. Running the set of simulations.

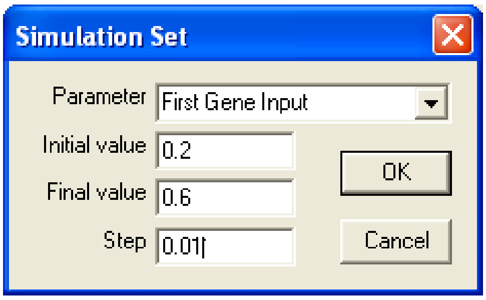
Often we want to see the test powers or other statistics as a function of some properties of the sample. This means that as a rule we should repeat the simulation with similar set of sample parameters and the same set of analyses. If we have already computed a desirable number of simulation replicates, we can use the simulation document to specify easy the new simulation with similar parameters. To do this, open the simulation document, then click menu item “Serial”-“Clear Iterations”. This action delete all saved simulation replicates and need to be confirmed.

When the number of replicates is zero, you can change the set of specified models (delete some of them or add other). Then save the document with another name and start the simulation. The simulation inherits all properties (sample structure and trait model) of the previous simulation, which was not changed. The simulation also inherits the list of nondefault statistics specified. So, if you have deleted some model or LRT, which take part in statistics, you will get an error. To avoid this before running the simulation, you can test what parameters are used in specified statistics. To do this, double click on the appropriate sub-item of the STATISTICS item in the document tree. You will see the dialog, where the specified expressions can be seen and edited.

If for example you need to test the power as a function of the number of the pedigrees, or number of offspring in the nuclear families, or number of generations in the pedigrees, you can change the sample structure. To do this clear computed replicates and then click “Serial”-“Pedigrees”-“Edit”. In the “Pedigree structure” dialog you can specify the new sample structure. Then save the document with a new name and start the simulation. All specifications, which were not changed (simulated trait model, set of analyses, specified statistics) will be inherited from the previous simulation.

The most often case is to test power as a function on the size of supposed effect of some gene on the phenotype or on joint properties of gene and tested marker. To do this you also can, after deleting of the computed replicates, edit the trait model parameters. Click menu item “Serial”-“Trait Model”. The dialog shows the properties, which were specified previously and enable to edit them.

If you need to compute statistics as a function of only one trait model parameter with defined step. You can specify the set of simulations automatically. To do this specify the simulation, then click “Serial”-“Run”. Specify the number of replications, which you want to compute for each simulation of the set. Then click “Serial”-“Run Set”. The “Simulation Set”-dialog opens.



Simulation Set

Parameter: First Gene Input

Initial value: 0.2

Final value: 0.6

Step: 0.01

In this dialog you should select the parameter, for which you want to do simulations with step. The options in the “Parameter”-combobox are as follows:

Spouses Corr means partial residual correlation for spouses (p0);

Parent-Offspring Cor means partial residual correlation for parent-offspring (p1);

Sibling Cor - partial residual correlation for siblings (p2);

First Gene Input means proportion of variance controlled by the upper gene in the “Trait” list of the“Simulation properties” dialog (p3);

Second Gene Input means proportion of residual variance controlled by the lower gene in the “Trait” list of the “Simulation properties” dialog (p4);

First Gene Frequency is allele frequency of the upper gene in the “Trait” list of the “Simulation properties” dialog (p5);

Second Gene Frequency is allele frequency of the lower gene in the “Trait” list of the “Simulation properties” dialog (p6);

Disequilibrium is disequilibrium parameter specified for the first “Haplotype” item in the DISEQUILIBRIUM control group of the “Simulation properties” dialog (p7).

After selecting parameter you should specify its initial and final value and step. So the simulation will be done at first for initial parameter value, for each next simulation the parameter value will increase on step while the value is less than the specified final value. Each simulation will be saved in the separate file. If the name of file, where you specified all other properties is <file_name>, separate files of the set will be named “<file_name>_p<parameter_number>_<parameter_value>”, where <parameter_number> is the number of parameter type in the list (from 0 for *Spouses Corr* to 7 for *Disequilibrium*); <parameter_value> is the value of parameter for this simulation multiplied on 100.

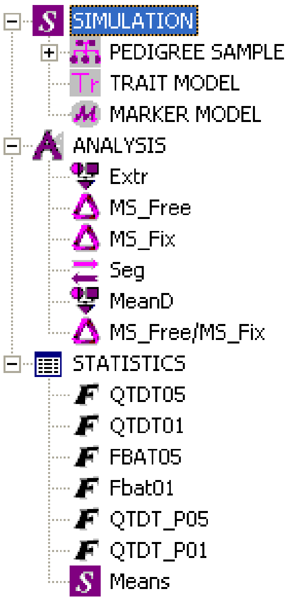
8.6. Viewing results of simulation.

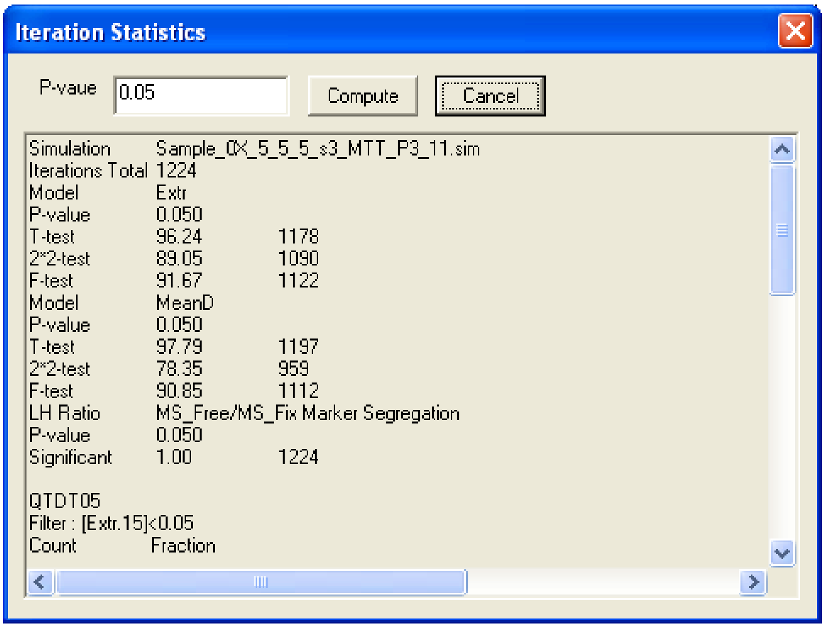
The results of simulation can be seen at any time after the simulation was started. If the model parameters are saved for all simulation replicates (“Save Statistics Only” was not checked), you can see the parameters of all already computed simulation replicates as tables. To view them, click on “View”-“Model Parameters” menu item or on “P” icon on the toolbar. In the right side of the document window you will see the table of parameters corresponding to the upper model in the ANALYSIS item of the document tree. When the new replicate is fully computed additional string is added to the table of model parameters. To switch between models click on the appropriate model name in the ANALYSIS item.

To view the current realization of traits and markers click on PEDIGREE SAMPLE icon in the document tree.

To view the number of iterations, which are to be computed, click menu item “Serial”-“Run”.

You also can view the current results for specified statistics (8.4). Click “View”-“Statistics”. The “Iteration Statistics” window opens. P-value in the upper edit box is used for default statistics: number and proportion of significant tests through already computed simulation replicates for all defined LRTs and TDTs (8.4). If P-value editbox enables to change its value, type the desirable threshold and click “Compute”- button. The contents of the lower editbox will be recalculated for current p-value and current state of simulation (number of already computed replicates).





P-vaue 0.05

```
Simulation      Sample_0X_5_5_5_s3_MTT_P3_11.sim
Iterations Total 1224
Model           Extr
P-value         0.050
T-test          96.24           1178
2*2-test        89.05           1090
F-test          91.67           1122
Model           MeanD
P-value         0.050
T-test          97.79           1197
2*2-test        78.35           959
F-test          90.85           1112
LH Ratio        MS_Free/MS_Fix Marker Segregation
P-value         0.050
Significant     1.00            1224

QTDT05
Filter : [Extr.15]<0.05
Count          Fraction
```

In the upper string you see the name of the simulation document file; the next shows the number of replicates already added (Iterations Total). After that, you will see all default features corresponding to models and LRTs, which you have specified.

Below is the list of all sub-items of the STATISTICS item. For each item you see its name and expressions, which were constructed for all fields. If you have specified statistics having only the *Filter* field defined, the output will be like the following:

```
QTDT05
Filter : [Extr.15]<0.05
Count   Fraction
 1202    0.9820
```

If the *Filter* and both possible values (8.4.1) were specified and *Save distribution* was checked, the output will be like the following:

```
Means
Filter : [Extr.11]<0.001   Value : [Extr.3]   Value : [Extr.4]
Count   Fraction           Mean V1  Mean V2  Varian1     Varian2   Correl.
   842     0.6879           -0.503   0.5121   0.0164      0.0194    0.2201
Distribution size 500
Percent  0.01    0.025    0.05     0.25     0.5         0.75      0.95      0.975     0.99
CritV   -0.245  -0.266   -0.294   -0.417   -0.498      -0.590    -0.724    -0.763    -0.809
```

8.6.1. Exporting results to Excel.

The results in the “Iteration Statistics” are formatted as tab-delimited text. You can select the contents of the results editbox using the mouse and copy them into clipboard by <Ctrl-C> hotkey. Then you can use *Paste*- options to put them into Excel-worksheet.

Tables of model parameters can be also exported into Excel-worksheet. To do this select for all models the parameter columns, which you want to export (as described in 7.6.3) and click menu item “Serial”-“Export to Excel”.

8.7. Joint distribution of various tests and combined p-value.

To combine the results of the a number of separate association tests (different variants of TDT and PDT) applied to the same trait-marker data, we formulated a multiple comparison procedure (MCP), which enables to compute combined P-values, the probability of erroneous rejection of the general null hypothesis of no linkage disequilibrium, which unites all certain null hypotheses of separate tests. MCP was constructed based on the joint simulated null-distribution for all separate tests. To account for the extent of correlation between results of different tests applied to the same data, we propose to use for simulation the same sample pedigree structure as in the actual sample. The trait pattern of inheritance should be simulated using the inheritance model exhibiting the observed familial correlations, but assuming that the tested marker had no effect on the trait variation (null-distribution). We recommend to compute about 50000 simulation replicates.

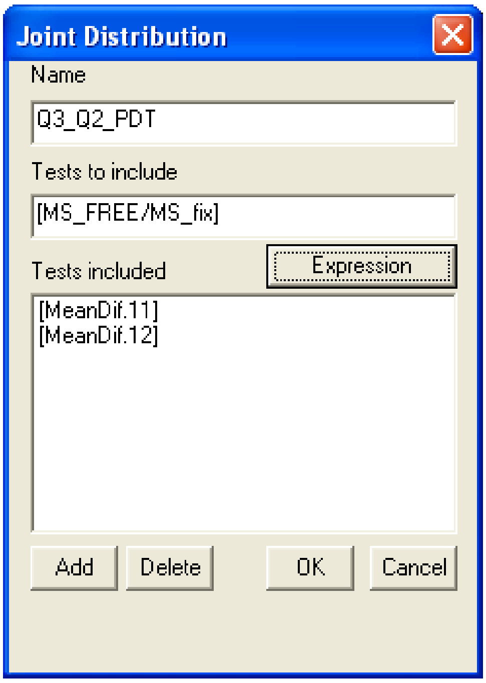
8.7.1. Multiple comparison procedure (MCP).

Each point of the simulated distribution is characterized by a vector of one-dimensional P-values (P_1, P_2, P_3,…,P_k) with components corresponding to separate association tests. The computation of the MCP P-value included two steps. First, to compare any pair of distribution points as more or less significant, we determined a function S (P_1, P_2, P_3,…, P_k) as a point significance level and computed this level for each distribution point. Second, to estimate the appropriate MCP P-value, P_{MCP}, for each vector (P_{1_0}, P_{2_0}, P_{3_0},…, P_{k_0}), we computed the fraction *f* of distribution points, which were more or equally significant in comparison with the level $S_0=S$ (P_{1_0}, P_{2_0}, P_{3_0},…, P_{k_0}). Thus, *f* is used as an estimator for P_{MCP} (the probability to reject erroneously the null hypothesis using the S_0 level as a criterion for rejection).

There are a number of ways to define the significance level function *S*. For example, we can define it as the minimum of one-dimensional P-values, *S* (P_1, P_2, P_3,…, P_k) = min (P_1, P_2, P_3,…, P_k) (simulative analog of Bonferroni correction), or as their maximum. To account for the relationship between all P-values, we use the population level for asymmetrical (A) and symmetrical (B) rejection regions as follows. Define that the distribution point I, having coordinates P_{1_i}, P_{2_i}, P_{3_i},…, P_{k_i}, belong to the rejection region of type A, if it satisfies the condition ($P_{1\_i}<P_1$) & ($P_{2\_i}<P_2$) & ($P_{3\_i}<P_3$) & …&($P_{k\_i}<P_k$), where & denotes the logical AND. Let $P_{(1)}<P_{(2)}<P_{(3)}<\ldots<P_{(k)}$ be the set of arranged P-values regardless of the serial numbers of tests for which each P-value was computed. Define that the distribution point *i* belongs to the rejection region of type B, if it satisfies the condition ($P_{i(1)}<P_{(1)}$) & ($P_{i(2)}<P_{(2)}$) &($P_{i(3)}<P_{(3)}$) &…& ($P_{i(k)}<P_{(k)}$). Define S_A (P_1, P_2, P_3,…, P_k) as the fraction of distribution points belonging to rejection asymmetrical region A and S_B (P_1, P_2, P_3,…, P_k) as the fraction of distribution points belonging to symmetrical rejection region B. Thus, we have two scores for arranging the distribution. Despite different arrangements, the resulting estimated MCP P-values of both types are generally of the same order. We compute both of them in order to use the more conservative estimator as the general adjusted P-value.

8.7.2. Generating of the joint null distribution.

If you have already get the simulation document having the set of tests, including all tests, you are going to combine, the trait model, assuming that the tested marker had no effect on the trait variation (null-distribution), and sufficient number of replicates, open it. Click menu item “Serial”-“New”-“Joint Distribution”. The *Joint Distribution*-dialog opens.



Joint Distribution

Name: Q3_Q2_PDT

Tests to include: [MS_FREE/MS_fix]

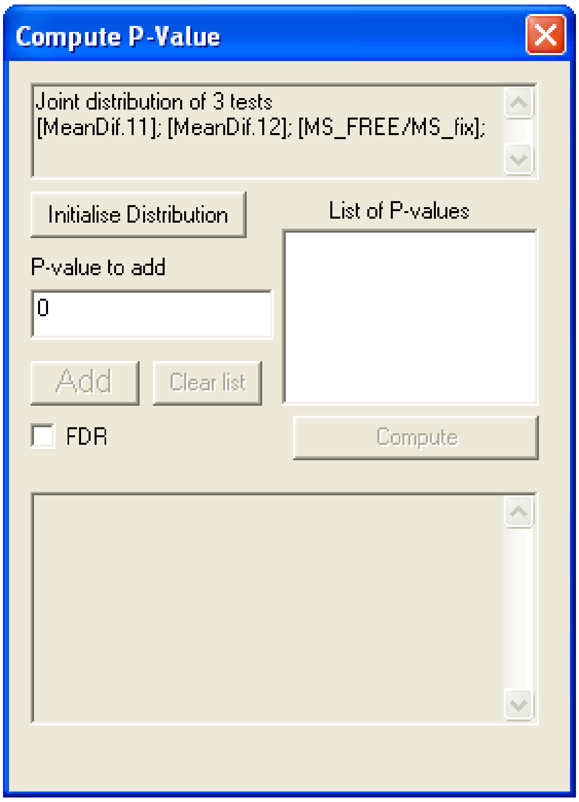
Tests included:
[MeanDif.11]
[MeanDif.12]

In the “Name”-editbox you should specify the name of distribution, which will further appear as sub-item of the simulation document tree. You should also include into the listbox expressions for all p-values, which you want to include in the joint distribution. To do this, click on the “Test to include”-editbox and then click “Expression”-button. In the “Expression”-dialog formulate the expression, which value should be equal to the desired p-value (as

described in 7.6.2) and click OK. The expression appears in the "Test to include"-editbox. Now click "Add"-button the expression moves to the listbox below. In the same way specify in the "Test to include"-editbox the next expression and add it to the list and so on. If you want to delete some expression from the listbox, select it with the mouse cursor and then click the "Delete"-button. When all desirable p-value expressions are in the listbox, click OK. The program begins to order the joint distribution in two ways described above (8.7.1) and to compute the appropriate combined p-values. For a big number of replicates the process can take some minutes. You will see the progress of computation in the lower part of the dialog.
You can break the process by clicking "Cancel"-button. If the distribution generation was performed successfully, the dialog closes and the new sub-item with appropriate name appears in the STATISTICS-item of the document tree.

8.7.3. Computing combined P-value for MCP.

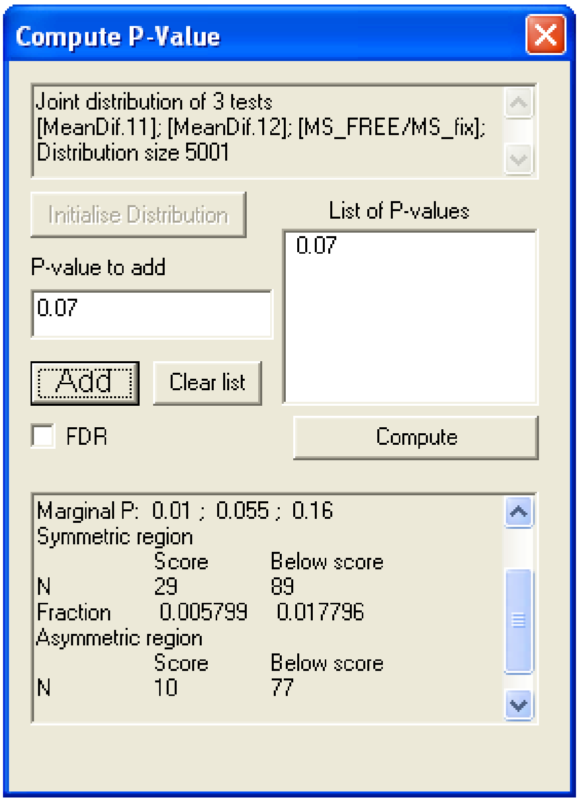
To use the joint distribution for computing combined p-value double click on its icon in the document tree. The "Compute P-Value"-dialog opens. If after the generation of the distribution you have closed the document file, the definition of the joint distribution items is saved, but the distribution itself should be initialized once more each time as the document is opened. If this is the case, click the "Initialize Distribution"-button. The progress of the initialization is shown in the lower part of the dialog.



Compute P-Value

Joint distribution of 3 tests
[MeanDif.11]; [MeanDif.12]; [MS_FREE/MS_fix];

If the buttons "Add", "Clear list" and "Compute" are enabled, this means that the distribution is already initialized. To compute combined p-value you should place the P-values vector into the listbox named "List of P-values" in the same order as the component tests are listed in the upper editbox. To do this type (or copy using <Ctrl-V>) the first p-value into the "P-value to add"-editbox and click the "Add" button, the value appears in the list. Then type the next p-value and click "Add" once more. Repeat these actions up to all values appear in the list. Now click the “Compute”-button. In the results editbox in the lower part of the window you will see computation for two type of significance scoring: population level of asymmetric (8.7.1 S_A) and symmetric (8.7.1 S_B) regions. For both types you will see the score value (column”Score”) and count of replication having less or equal score (column”Below score”). The most conservative (greater) proportion of replications below score can be taken as an estimator of combined P-Value.



```
Joint distribution of 3 tests
[MeanDif.11]; [MeanDif.12]; [MS_FREE/MS_fix];
Distribution size 5001
```

List of P-values: 0.07

P-value to add: 0.07

☐ FDR

```
Marginal P:  0.01 ;  0.055 ;  0.16
Symmetric region
                Score         Below score
N               29            89
Fraction        0.005799      0.017796
Asymmetric region
                Score         Below score
N               10            77
```

You can clear the list of p-values using “Clear list”-button, specify the new P-values vector and click once more the “Compute”-button. New result will be added to the previous contents of the results editbox. You can select the contents of the result editbox by the mouse and copy it into clipboard by <Ctrl-C> hotkey.

You should recognize that if you analyze the very significant vector of P-values, you distribution should have the appropriate number of replication. The combined P-value can be treated as reliable if the number of replication below score is about 100.

Using the joint distribution you can see the relation between theoretical and experimental false discovery rate (FDR) on you sample for selected set of tests. To do this, check the “FDR”-checkbox. Specify the value in the “P-value to add”-editbox and click the “Compute”-button. In the results window you will see the proportion of replicates having at leas one test from the set significant on the specified FDR level, accounting for the number of tests in the defined list of tests.