Version 17

Genetics

"The real voyage of discovery consists not in seeking new landscapes, but in having new eyes."

Marcel Proust

JMP Statistical Discovery LLC
SAS Campus Drive
Cary, North Carolina 27513-2414

17.1

The correct bibliographic citation for this manual is as follows: JMP Statistical Discovery LLC 2022–2023. *JMP® 17 Genetics*. Cary, NC: JMP Statistical Discovery LLC

JMP® 17 Genetics

JMP Statistical Discovery LLC, SAS Campus Drive, Cary, North Carolina 27513-2414.

October 2022

March 2023

Get the Most from JMP

Whether you are a first-time or a long-time user, there is always something to learn about JMP.

Visit JMP.com to find the following:

- live and recorded webcasts about how to get started with JMP
- video demos and webcasts of new features and advanced techniques
- details on registering for JMP training
- schedules for seminars being held in your area
- success stories showing how others use JMP
- the JMP user community, resources for users including examples of add-ins and scripts, a forum, blogs, conference information, and so on

jmp.com/getstarted

Contents

Genetics

Chapter 1

Learn about JMP

Documentation and Additional Resources

Learn about JMP documentation, such as book conventions, descriptions of each JMP document, the Help system, and where to find additional support.

Contents

Formatting Conventions in JMP Documentation

These conventions help you relate written material to information that you see on your screen:

- Sample data table names, column names, pathnames, filenames, file extensions, and folders appear in Helvetica (or sans-serif online) font.
- Code appears in `Lucida Sans Typewriter` (or monospace online) font.
- Code output appears in *`Lucida Sans Typewriter`* italic (or monospace italic online) font and is indented farther than the preceding code.
- **Helvetica bold** formatting (or bold sans-serif online) indicates items that you select to complete a task:
  - buttons
  - check boxes
  - commands
  - list names that are selectable
  - menus
  - options
  - tab names
  - text boxes
- The following items appear in italics:
  - words or phrases that are important or have definitions specific to JMP
  - book titles
  - variables
- Features that are for JMP Pro only are noted with the JMP Pro icon JMP PRO. For an overview of JMP Pro features, visit jmp.com/software/pro.

Note: Special information and limitations appear within a Note.

Tip: Helpful information appears within a Tip.

JMP Help

JMP Help in the Help menu enables you to search for information about JMP features, statistical methods, and the JMP Scripting Language (or *JSL*). You can open JMP Help in several ways:

- Search and view JMP Help on Windows by selecting **Help > JMP Help**.
- On Windows, press the F1 key to open the Help system in the default browser.
- Get help on a specific part of a data table or report window. Select the Help tool from the **Tools** menu and then click anywhere in a data table or report window to see the Help for that area.
- Within a JMP window, click the **Help** button.

Note: The JMP Help is available for users with Internet connections. Users without an Internet connection can search all books in a PDF file by selecting **Help > JMP Documentation Library**. See "JMP Documentation Library" for more information.

JMP Documentation Library

The Help system content is also available in one PDF file called *JMP Documentation Library*. Select **Help > JMP Documentation Library** to open the file. You can also download the Documentation PDF Files add-in if you prefer searching individual PDF files of each document in the JMP library. Download the available add-ins from community.jmp.com.

The following table describes the purpose and content of each document in the JMP library.

Document Title	Document Purpose	Document Content
Discovering JMP	If you are not familiar with JMP, start here.	Introduces you to JMP and gets you started creating and analyzing data. Also learn how to share your results.
Using JMP	Learn about JMP data tables and how to perform basic operations.	Covers general JMP concepts and features that span across all of JMP, including importing data, modifying columns properties, sorting data, and connecting to SAS.

Document Title	Document Purpose	Document Content
Basic Analysis	Perform basic analysis using this document.	Describes these Analyze menu platforms: • Distribution • Fit Y by X • Tabulate • Text Explorer Covers how to perform bivariate, one-way ANOVA, and contingency analyses through Analyze > Fit Y by X. How to approximate sampling distributions using bootstrapping and how to perform parametric resampling with the Simulate platform are also included.
Essential Graphing	Find the ideal graph for your data.	Describes these Graph menu platforms: • Graph Builder • Scatterplot 3D • Contour Plot • Bubble Plot • Parallel Plot • Cell Plot • Scatterplot Matrix • Ternary Plot • Treemap • Chart • Overlay Plot The book also covers how to create background and custom maps.
Profilers	Learn how to use interactive profiling tools, which enable you to view cross-sections of any response surface.	Covers all profilers listed in the Graph menu. Analyzing noise factors is included along with running simulations using random inputs.

Document Title	Document Purpose	Document Content
Design of Experiments Guide	Learn how to design experiments and determine appropriate sample sizes.	Covers all topics in the DOE menu.
Fitting Linear Models	Learn about Fit Model platform and many of its personalities.	Describes these personalities, all available within the Analyze menu Fit Model platform: • Standard Least Squares • Stepwise • Generalized Regression • Mixed Model • Generalized Linear Mixed Model • MANOVA • Loglinear Variance • Nominal Logistic • Ordinal Logistic • Generalized Linear Model

Document Title	Document Purpose	Document Content
Predictive and Specialized Modeling	Learn about additional modeling techniques.	Describes these Analyze > Predictive Modeling menu platforms: • Neural • Partition • Bootstrap Forest • Boosted Tree • K Nearest Neighbors • Naive Bayes • Support Vector Machines • Model Comparison • Model Screening • Make Validation Column • Formula Depot Describes these Analyze > Specialized Modeling menu platforms: • Fit Curve • Nonlinear • Functional Data Explorer • Gaussian Process • Time Series • Matched Pairs Describes these Analyze > Screening menu platforms: • Explore Outliers • Explore Missing Values • Explore Patterns • Response Screening • Predictor Screening • Association Analysis • Process History Explorer

Document Title	Document Purpose	Document Content
Multivariate Methods	Read about techniques for analyzing several variables simultaneously.	Describes these Analyze > Multivariate Methods menu platforms: • Multivariate • Principal Components • Discriminant • Partial Least Squares • Multiple Correspondence Analysis • Structural Equation Models • Factor Analysis • Multidimensional Scaling • Multivariate Embedding • Item Analysis Describes these Analyze > Clustering menu platforms: • Hierarchical Cluster • K Means Cluster • Normal Mixtures • Latent Class Analysis • Cluster Variables

Document Title	Document Purpose	Document Content
Quality and Process Methods	Read about tools for evaluating and improving processes.	Describes these Analyze > Quality and Process menu platforms: • Control Chart Builder and individual control charts • Measurement Systems Analysis (EMP and Type 1 Gauge) • Variability / Attribute Gauge Charts • Process Screening • Process Capability • Model Driven Multivariate Control Chart • Legacy Control Charts • Pareto Plot • Diagram • Manage Limits • OC Curves
Reliability and Survival Methods	Learn to evaluate and improve reliability in a product or system and analyze survival data for people and products.	Describes these Analyze > Reliability and Survival menu platforms: • Life Distribution • Fit Life by X • Cumulative Damage • Recurrence Analysis • Degradation • Repeated Measures Degradation • Destructive Degradation • Reliability Forecast • Reliability Growth • Reliability Block Diagram • Repairable Systems Simulation • Survival • Fit Parametric Survival • Fit Proportional Hazards

Document Title	Document Purpose	Document Content
Consumer Research	Learn about methods for studying consumer preferences and using that insight to create better products and services.	Describes these Analyze > Consumer Research menu platforms: • Categorical • Choice • MaxDiff • Uplift • Multiple Factor Analysis
Genetics	Learn about methods that are available in JMP to help you analyze your genetic data and use that data to simulate a breeding program to predict the optimum genetic crosses to make.	Describes these Analyze > Genetics menu platforms: • Marker Statistics • Marker Simulation
Scripting Guide	Learn about taking advantage of the powerful JMP Scripting Language (JSL).	Covers a variety of topics, such as writing and debugging scripts, manipulating data tables, constructing display boxes, and creating JMP applications.
JSL Syntax Reference	Read about many JSL functions on functions and their arguments, and messages that you send to objects and display boxes.	Includes syntax, examples, and notes for JSL commands.

Additional Resources for Learning JMP

In addition to reading JMP help, you can also learn about JMP using the following resources:

Search JMP

If you are not sure where to find a statistical procedure, do a search across JMP. Results are tailored to the window that you launch the search from, such as a data table or report.

1. Click **Help** > **Search JMP**. Or, press Ctrl+comma.
2. Enter your search text.
3. Click the result that contains the procedure that you want.

   On the right, you can see a description and the location of the procedure.
4. Click the corresponding button to open or go to a result.

JMP Tutorials

You can access JMP tutorials by selecting **Help > Tutorials**.

If you are not familiar with JMP, start with the **Beginners Tutorial**. It steps you through the JMP interface and explains the basics of using JMP.

The rest of the tutorials help you with specific aspects of JMP, such as designing an experiment and comparing a sample mean to a constant.

Sample Data Tables

All of the examples in the JMP documentation suite use sample data. Select **Help > Sample Data Folder** to open the sample data directory.

To view an alphabetized list of sample data tables or view sample data within categories, select **Help > Sample Index**.

Sample data tables are installed in the following directory:

On Windows: C:\Program Files\SAS\JMP\17\Samples\Data

On macOS: \Library\Application Support\JMP\17\Samples\Data

In JMP Pro, sample data is installed in the JMPPRO (rather than JMP) directory.

To view examples using sample data, select **Help > Sample Index** and navigate to the Teaching Resources section.

Learn about Statistical and JSL Terms

For help with statistical terms, select Help > Statistics Index. For help with JSL scripting and examples, select **Help > Scripting Index**.

Statistics Index Provides definitions of statistical terms.

Scripting Index Lets you search for information about JSL functions, objects, and display boxes. You can also edit and run sample scripts from the Scripting Index and get help on the commands.

Learn JMP Tips and Tricks

When you first start JMP, you see the Tip of the Day window. This window provides tips for using JMP.

To turn off the Tip of the Day, clear the **Show tips at startup** check box. To view it again, select **Help > Tip of the Day**. Or, you can turn it off using the Preferences window.

JMP Tooltips

JMP provides descriptive tooltips (or *hover labels*) when you hover over items, such as the following:

- Menu or toolbar options
- Labels in graphs
- Text results in the report window (move your cursor in a circle to reveal)

- Files or windows in the Home Window
- Code in the Script Editor

Tip: On Windows, you can hide tooltips in the JMP Preferences. Select **File > Preferences > General** and then deselect **Show menu tips**. This option is not available on macOS.

JMP User Community

The JMP User Community provides a range of options to help you learn more about JMP and connect with other JMP users. The learning library of one-page guides, tutorials, and demos is a good place to start. And you can continue your education by registering for a variety of JMP training courses.

Other resources include a discussion forum, sample data and script file exchange, webcasts, and social networking groups.

To access JMP resources on the website, select **Help > JMP on the Web > JMP User Community** or visit https://community.jmp.com.

Free Online Statistical Thinking Course

Learn practical statistical skills in this free online course on topics such as exploratory data analysis, quality methods, and correlation and regression. The course consists of short videos, demonstrations, exercises, and more. Visit jmp.com/statisticalthinking.

JMP New User Welcome Kit

The JMP New User Welcome Kit is designed to help you quickly get comfortable with the basics of JMP. You'll complete its thirty short demo videos and activities, build your confidence in using the software, and connect with the largest online community of JMP users in the world. Visit jmp.com/welcome.

Statistics Knowledge Portal

The Statistics Knowledge Portal combines concise statistical explanations with illuminating examples and graphics to help visitors establish a firm foundation upon which to build statistical skills. Visit jmp.com/skp.

JMP Training

SAS offers training on a variety of topics led by a seasoned team of JMP experts. Public courses, live web courses, and on-site courses are available. You might also choose the online e-learning subscription to learn at your convenience. Visit jmp.com/training.

JMP Books by Users

Additional books about using JMP that are written by JMP users are available on the JMP website. Visit jmp.com/books.

The JMP Starter Window

The JMP Starter window is a good place to begin if you are not familiar with JMP or data analysis. Options are categorized and described, and you launch them by clicking a button. The JMP Starter window covers many of the options found in the Analyze, Graph, Tables, and File menus. The window also lists JMP Pro features and platforms.

- To open the JMP Starter window, select **View** (**Window** on macOS) **> JMP Starter**.
- To display the JMP Starter automatically when you open JMP on Windows, select **File > Preferences > General**, and then select **JMP Starter** from the Initial JMP Window list. On macOS, select **JMP > Preferences > General > Initial JMP Starter Window**.

JMP Technical Support

JMP technical support is provided by statisticians and engineers educated in SAS and JMP, many of whom have graduate degrees in statistics or other technical disciplines.

Many technical support options are provided at jmp.com/support, including the technical support phone number.

Chapter 2

Introduction to Genetics

Overview of Genetic Analysis

Genetics provides two methods in JMP to help you analyze your genetic data and use that data to simulate a breeding program to predict the optimum genetic crosses to make.

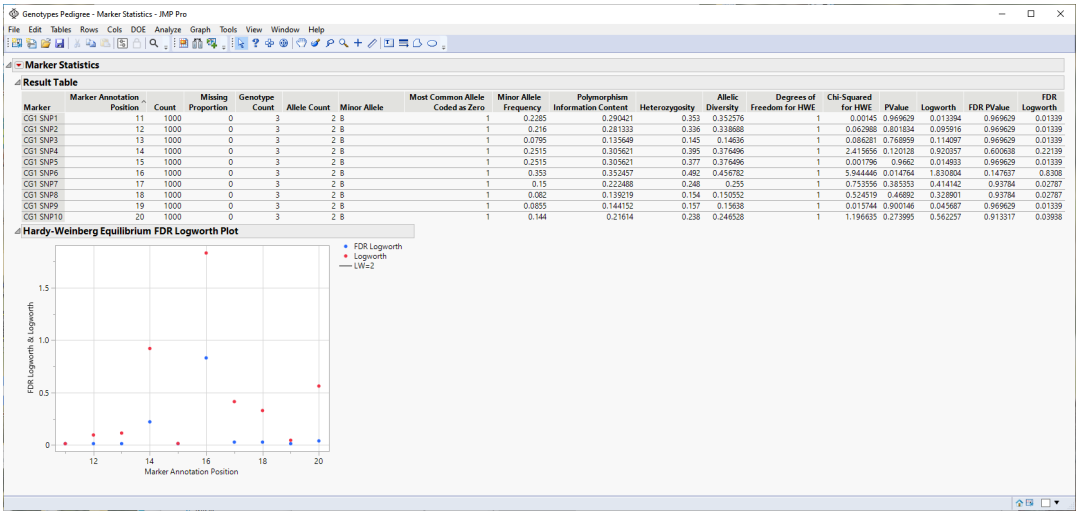
- The Marker Statistics platform provides a convenient method for exploring several properties of all the biallelic markers in a data set, for the purpose of quality control (QC) and possibly selecting markers to be removed from the analysis See "Marker Statistics".
- The Marker Simulation platform simulates the progeny from a specified set of crosses using biallelic markers and predictor formulas generated using the Response Screening platform (See the Predictive and Specialized Modeling book.) saved in your data table. This process enables you to test various crosses to estimate which crosses will generate progeny with the desired combinations of traits. See "Marker Simulation".

Chapter 3

JMP PRO Marker Statistics

The Marker Statistics platform provides a convenient method for exploring several properties of all the biallelic markers in a data set, for the purpose of quality control (QC) and possibly selecting markers to be removed from the analysis.

Figure 3.1 Example Analysis Using the Marker Statistics Platform



The Marker Statistics platform is available only in JMP Pro.

Contents

Overview of the Marker Statistics Platform

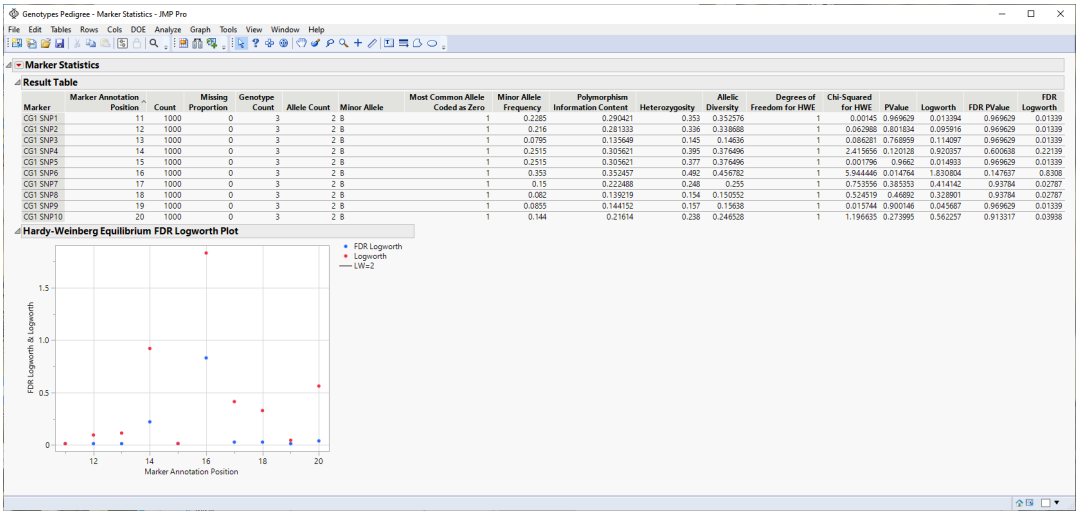
This process calculates various single marker measures, such as polymorphism information content, allele diversity, heterozygosity, minor allele frequency, allele and genotype frequencies, tests for Hardy-Weinberg equilibrium (HWE). It also calculates various measurements of linkage disequilibrium, as well as an overall test for linkage disequilibrium, between pairs of markers.

Example of Marker Statistics

Learn how to calculate marker statistics from your genetic data.

1. Select **Help > Sample Data Folder** > **Life Sciences** and open Genotypes Pedigree.jmp.
2. Select **Analyze > Genetics > Marker Statistics**.
3. Open the Markers column, select the first ten marker columns, and click **Marker**.
4. Click **OK**.

Figure 3.2 The Marker Statistics Report



The Result Table provides statistics for each of the markers analyzed. See “Result Table”.

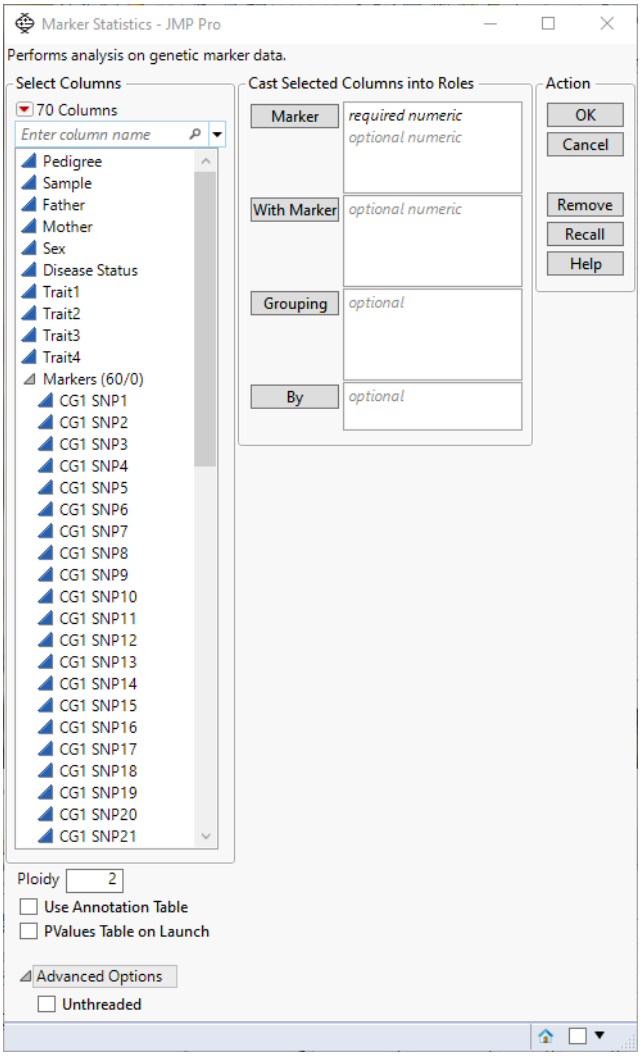
The Hardy-Weinberg Equilibrium plot shows how close each of the markers are to equilibrium. The column number of each marker in the JMP table is plotted on the *x*-axis. The Logworth-adjusted p-value and -log10 (p-value) for HWE for each marker is plotted on the *y*-axis. Markers at or below the reference line are at equilibrium, markers above the line are not. The farther from the reference a marker is, the further the marker is from equilibrium.

Note: To see the Levels and Counts tables, click the Marker Statistics red triangle menu and select the corresponding option.

Launch the Marker Statistics Platform

To calculate the marker statistics for your experimental data, launch the Marker Statistics platform by selecting **Analyze > Genetics > Marker Statistics**.

Figure 3.3 The Marker Statistics Launch Window



For information about the options in the Select Columns red triangle menu, see *Using JMP*.

Marker Select desired marker columns and click **Marker** to specify the markers that you want to analyze.

With Marker Select desired marker columns and click **With Marker** to specify the markers to use in pairwise linkage disequilibrium comparisons with the markers specified in X, Marker above.

Grouping Analyzes the rows assigned to each level of the specified column separately. All results are presented in a single table and report.

By Produces a separate report for each level of the By variable. If more than one By variable is assigned, a separate report is produced for each possible combination of the levels of the By variables.

Ploidy Enables you to specify the ploidy level of the experimental organism under investigation.

Use Annotation Table Enables you to access annotation information contained in a separate data table. After you click **OK**, a window appears, prompting you to specify the name and location of the annotation table.

PValues Table on Launch Creates a data table for the *p*-values of the associated estimated marker statistics. This data table is linked to the Result Table in the Marker Statistics report.

Unthreaded Suppresses multithreading. Deselect this option for improved computational speed.

Required Data Format for the Marker Statistics Platform

Most of the processes in JMP assume that the input table has a particular data structure. JMP distinguishes between tall and wide data sets. A tall data table has samples as columns and molecular entity (for example, marker, gene, clone, protein, or metabolite) as rows, whereas a wide data table is the transpose of the tall, having the samples as rows and molecular entity as columns.

When specifying the input data set for a process, it is important to know the required form. Marker Statistics requires a wide data table. The Transpose platform under the Tables menu enables you to transform your data tables between tall and wide forms.

Marker data must be encoded in the one-column, numeric format. Typically, in this format, diploid individuals homozygous for the least common, or minor allele, are represented in the table by a 2, whereas the heterozygotes are represented by a 1. Homozygotes for the most common allele are represented by a 0.

The Marker Statistics Report and Options

The Marker Statistics report enables you to explore several properties of all the biallelic markers in a data set. This process calculates various single marker measures, as well as various measurements of linkage disequilibrium.

The Marker Statistics red triangle menu contains the following options:

Levels Table Select this option to view the "Additional Example of the Marker Statistics Platform". The Levels table indicates the number of genotypes observed for each marker.

Counts Table Select this option to view the "Counts Table". The Counts table indicates the number of individuals exhibiting each genotype for each marker.

Result Table Select this option to view the "Result Table".

PValue Plot Select this option to view the "Hardy-Weinberg Equilibrium FDR Logworth Plot".

LD Decay Plot Select this option to view the "LD Decay Plot".

Note: This option is available when one or more With Markers are specified.

Save Result Table Use this option to save the Result table as a JMP table.

Save Proportion of Missing Markers Use this option to add a column to the original data table containing the proportion of missing markers in each row.

Select Where Use this option to subset the Result table based on specified criteria.

Select Columns Select row in the Result table and then use this option to select the corresponding columns in the original data table.

Recode Marker Recodes all the selected marker columns in the original data table by switching the major and minor alleles.

Levels Table

The Levels table indicates the number of genotypes observed for each marker. To show the Levels table, click the Marker Statistics red triangle and select **Levels Table**.

Counts Table

The Counts table indicates the number of individuals exhibiting each genotype for each marker. To show the Counts table, click the Marker Statistics red triangle and select **Counts Table**.

Result Table

Marker Annotation Position This column lists the chromosomal location of each marker is listed here provided you specify an annotation table listing the chromosomal locations of the markers. If no position information is supplied, the column number of each marker in the input JMP table is listed instead.

With Marker Annotation Position This column lists the chromosomal location of each comparison marker is listed here provided you specify an annotation table listing the chromosomal locations of the markers. If no position information is supplied, the column number of each comparison marker in the input JMP table is listed instead.

Distance between Marker Pairs This column lists the distance between pairs of markers. If no position information is supplied, the number of columns between each comparison marker in the input JMP table is listed instead.

Count The total number of individuals observed for each marker.

Missing Proportion The proportion of the individuals missing data for each marker or pair of markers.

Genotype Count The number of genotypes observed for each marker.

Allele Count The number of alleles observed for each marker.

Minor Allele The allele at each marker that occurs less frequently.

Most Common Allele Coded as Zero The allele at each marker that occurs most frequently. Individuals homozygous for this allele are coded as 0 for that marker.

Minor Allele Frequency (MAF) The MAF represents the proportion of the minor allele for each marker in the observed population. Assuming a biallelic marker locus M with alleles M_1 and M_2. A sample of N individuals of even ploidy k can therefore have $k + 1$ different genotypes at the locus. The number of individuals with i ($i = 0, 1, 2, \ldots, k$) copies of allele M_1 and j ($j = 0, 1, 2, \ldots, k$) copies of allele M_2 is denoted by N_{ij}. The number n_1 of copies of allele M_1 can be found directly by summation: $n_1 = 0 \times N_{00} + 1 \times N_{10} + 2 \times N_{20} + \ldots + k \times N_{k0}$. The sample frequency of allele M_1 is written as $p_1 = n_1 / (k \times n)$, frequency of allele M_2 is written as $p_2 = 1 - p_1$, and the sample frequency for each genotype carrying u copies of allele M_1 and v copies of allele M_2 is written as $P_{ij} = N_{ij} / n$.

Polymorphism Information Content (PIC) The PIC (Botstein *et al.*, 1980; Hilderbrand, Torney, and Wagner, 1992) measures the probability of differentiating the allele transmitted by a given parent to its child given the marker genotype of father, mother, and child.

$$PIC = 1 - \sum_{u=1}^{k} p_u^2 - \sum_{u=1}^{k-1} \sum_{v=u+1}^{k} 2p_u^2 p_v^2$$

Heterozygosity (HET) The heterozygosity, sometimes called the observed heterozygosity, is simply the proportion of heterozygous individuals in the observed population.

$$Het = 1 - P_{ii} - P_{jj} (i < j)$$

Allelic Diversity The allelic diversity, sometimes called the expected heterozygosity, is the expected proportion of heterozygous individuals in the data set when HWE holds.

$$Div = 1 - p_1^2 - p_2^2$$

Degrees of Freedom for HWE The HWE degrees of freedom are calculated using the formula $k(k-1)/2$, where k is the number of alleles found for a given marker that is being tested.

Chi-Squared for HWE Under ideal population conditions, the two alleles an individual receives, one from each parent, are independent, so that $P_{ii} = p_1^2$, $P_{jj} = p_2^2$ and $P_{ij} = 2p_1 p_2$ (i and $j = 0, 1, 2, \ldots, k$). The factor of 2 for heterozygotes recognizes the fact that M_1 / M_2 and M_2 / M_1 genotypes are generally indistinguishable. This statement about allelic independence within loci is called Hardy-Weinberg equilibrium (HWE). Forces such as selection, mutation, and migration in a population or nonrandom mating can cause departures from HWE. The chi-square goodness-of-fit test can be used to test markers for HWE (null hypothesis of $P_{ii} = p_1^2$, $P_{jj} = p_2^2$, and $P_{ij} = 2p_1 p_2$) is:

$$X_T^2 = \sum_u \frac{(n_{uu} - n\tilde{p}_u^2)^2}{n\tilde{p}_u^2} + \sum_u \sum_{v \neq u} \frac{(n_{uv} - 2n\tilde{p}_u\tilde{p}_v)^2}{2n\tilde{p}_u\tilde{p}_v}$$

Chi-Squared for LD (ChiSQLD) The *chi*-square statistic provides an estimate of the strength of the measures of linkage disequilibrium using the composite linkage disequilibrium (CLD) coefficient (Weir 1979) that does not require the assumption of HWE and uses only allele and two-locus genotype frequencies.

$$X_T^2 = \sum_{u=1}^{k} \sum_{v=1}^{l} \frac{nD_{uv}^2}{\tilde{p}_u\tilde{p}_v}$$

For biallelic markers, k and $l = 2$, and this test has 1 degree of freedom.

Composite LD (D) The measure of the composite linkage disequilibrium (CLD) coefficient (Weir 1979), which does not assume HWE, and is written as follows: $D_{12} = p_{12} + p_{1/2} - 2p_1 p_2$, where:

- p_{12} is the frequency of gamete $M_1 M_2$
- $p_{1/2}$ is the joint frequency of alleles M_1 and M_2 at two different gametes
- p_1 and p_2 are the frequencies of alleles M_1 and M_2 at two loci [Weir 1996]

Standardized Composite LD (D') The ratio of D to D_{max} (Zaykin, 2004).

$$D' = \frac{D}{D_{max}}$$

LD Correlation Coefficient (*r*) A correlation coefficient assuming values from -1 to 1indicator variables indicating the presence of the two loci.

$$r = \frac{D}{(p_1 p_2 q_1 q_2)^{1/2}}$$

P-value associated measures 4 columns of *p*-value-associated measures (PValue, Logworth (log_{10} *p*-value), FDR PValue (False Discovery Rate *p*-value), and FDR Logworth) of the strength of the *chi*-square test for Hardy-Weinberg Equilibrium and/or linkage disequilibrium.

Hardy-Weinberg Equilibrium FDR Logworth Plot

The Hardy-Weinberg Equilibrium plot shows how close each of the markers are to equilibrium. The column number of each marker in the JMP table is plotted on the *x*-axis. The Logworth-adjusted *p*-value and -log_{10}(*p*-value) for HWE for each marker is plotted on the *y*-axis. Markers at or below the reference line are at equilibrium, markers above the line are not. The farther from the reference a marker is, the further the marker is from equilibrium.

LD Decay Plot

The LD Decay plot shows the relationship between LD Correlation Coefficient on the *y*-axis and the distance between marker pairs on the *x*-axis in order to understand the pattern of linkage disequilibrium (LD). Linkage disequilibrium generally declines exponentially as the distance between the markers increases.

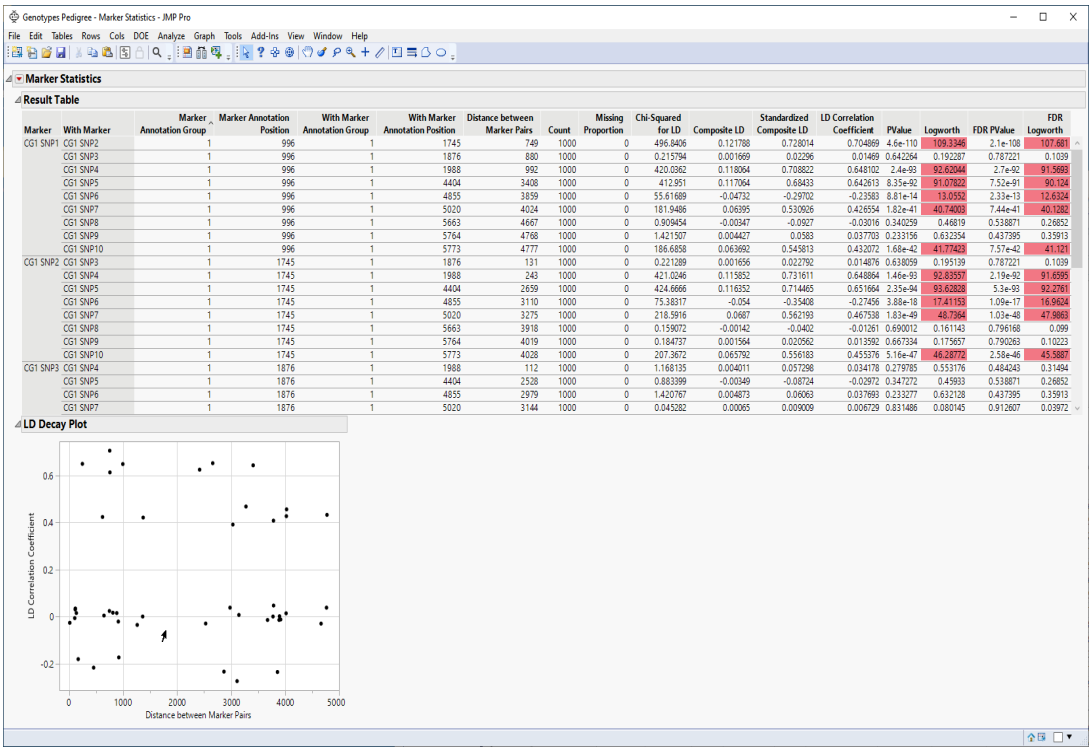
Additional Example of the Marker Statistics Platform

Learn how to calculate linkage statistics from your genetic data.

1. Select **Help** > **Sample Data Folder** > **Life Sciences** and open Genotypes Pedigree.jmp and Genotypes Pedigree Anno.jmp.
2. Select **Analyze > Genetics > Marker Statistics**.
3. Open the Markers column, select the first ten marker columns, and click **Marker**.
4. Select the same ten markers and click **With Marker**.
5. Check the **Use Annotation Table** check box.
6. Click **OK**.

7. In the Choose Annotation Table window, select the Genotypes Pedigree Anno.jmp annotation data table.
8. In the Annotation Specifications window, select Marker and click **Marker Variables**.
9. Select Gene and click **Annotation Group**.
10. Select Physical Position and click **Annotation Position**.
11. Click **OK**.

Figure 3.4 Analysis of Linkage Disequilibrium



Marker	With Marker	Marker Annotation Group	Marker Annotation Position	With Marker Annotation Group	With Marker Annotation Position	Distance between Marker Pairs	Count	Missing Proportion	Chi-Squared for LD	Composite LD	Standardized Composite LD	LD Correlation Coefficient	PValue	Logworth	FDR PValue	FDR Logworth
CG1 SNP1	CG1 SNP2	1	996	1	1745	749	1000	0	496.8406	0.121788	0.728014	0.704869	4.6e-110	109.3346	2.1e-108	107.681
	CG1 SNP3	1	996	1	1876	880	1000	0	0.215794	0.001669	0.02296	0.01469	0.642264	0.192287	0.787221	0.1039
	CG1 SNP4	1	996	1	1988	992	1000	0	420.0362	0.118064	0.708822	0.648102	2.4e-93	92.62044	2.7e-92	91.5693
	CG1 SNP5	1	996	1	4404	3408	1000	0	412.951	0.117064	0.68433	0.642613	8.35e-92	91.07822	7.52e-91	90.124
	CG1 SNP6	1	996	1	4855	3859	1000	0	55.61689	-0.04732	-0.29702	-0.23583	8.81e-14	13.0552	2.33e-13	12.6324
	CG1 SNP7	1	996	1	5020	4024	1000	0	181.9486	0.06395	0.530926	0.426554	1.82e-41	40.74003	7.44e-41	40.1282
	CG1 SNP8	1	996	1	5663	4667	1000	0	0.909454	-0.00347	-0.0927	-0.03016	0.340259	0.46819	0.538871	0.26852
	CG1 SNP9	1	996	1	5764	4768	1000	0	1.421507	0.004427	0.0583	0.037703	0.233156	0.632354	0.437395	0.35913
	CG1 SNP10	1	996	1	5773	4777	1000	0	186.6858	0.063692	0.545813	0.432072	1.68e-42	41.77423	7.57e-42	41.121
CG1 SNP2	CG1 SNP3	1	1745	1	1876	131	1000	0	0.221289	0.001656	0.022792	0.014876	0.638059	0.195139	0.787221	0.1039
	CG1 SNP4	1	1745	1	1988	243	1000	0	421.0246	0.115852	0.731611	0.648864	1.46e-93	92.83557	2.19e-92	91.6595
	CG1 SNP5	1	1745	1	4404	2659	1000	0	424.6666	0.116352	0.714465	0.651664	2.35e-94	93.62828	5.3e-93	92.2761
	CG1 SNP6	1	1745	1	4855	3110	1000	0	75.38317	-0.054	-0.35408	-0.27456	3.88e-18	17.41153	1.09e-17	16.9604
	CG1 SNP7	1	1745	1	5020	3275	1000	0	218.5916	0.0687	0.562193	0.467538	1.83e-49	48.7364	1.03e-48	47.9863
	CG1 SNP8	1	1745	1	5663	3918	1000	0	0.159072	-0.00142	-0.0402	-0.01261	0.690012	0.161143	0.796168	0.099
	CG1 SNP9	1	1745	1	5764	4019	1000	0	0.184737	0.001564	0.020562	0.013592	0.667334	0.175657	0.790263	0.10223
	CG1 SNP10	1	1745	1	5773	4028	1000	0	207.3672	0.065792	0.556183	0.455376	5.16e-47	46.28772	2.58e-46	45.5887
CG1 SNP3	CG1 SNP4	1	1876	1	1988	112	1000	0	1.168135	0.004011	0.057298	0.034178	0.279785	0.553176	0.484243	0.31494
	CG1 SNP5	1	1876	1	4404	2528	1000	0	0.883399	-0.00349	-0.08724	-0.02972	0.347272	0.45933	0.538871	0.26852
	CG1 SNP6	1	1876	1	4855	2979	1000	0	1.420767	0.004873	0.06063	0.037693	0.233277	0.632128	0.437395	0.35913
	CG1 SNP7	1	1876	1	5020	3144	1000	0	0.045282	0.00065	0.009009	0.006729	0.831486	0.080145	0.912607	0.03972

In the Result table, the values shown for each pair of markers are indicative of the strength of linkage disequilibrium between the pair. In the LD Decay Plot, you can see that linkage disequilibrium generally declines exponentially as the distance between the markers increases.

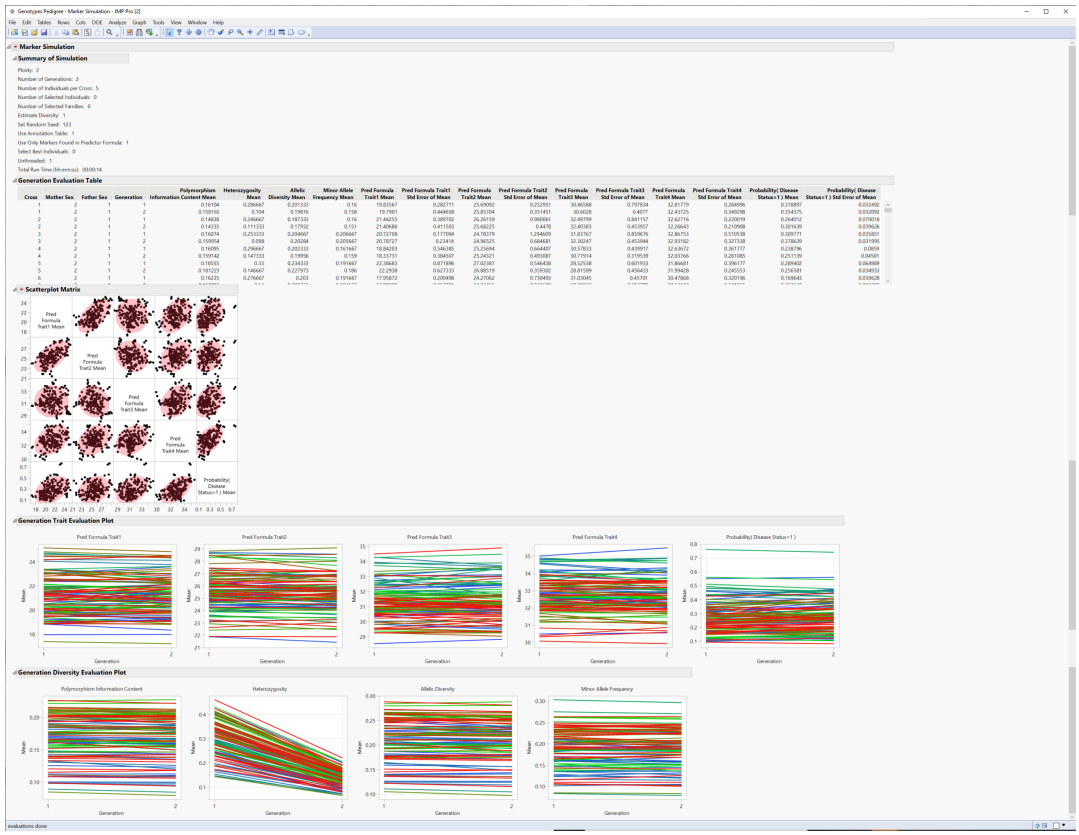
Chapter 4

JMP PRO Marker Simulation

Marker Simulation simulates the progeny from a specified set of crosses using biallelic markers and predictor formulas (predictive models generated using the Response Screening platform) saved in your data table. This process enables you to test various crosses to estimate which crosses generate progeny with the desired combinations of traits.

Note: Specifying a large number of crosses, progeny, or both can consume considerable CPU and disk resources.

Figure 4.1 Example of the Marker Simulation Platform



The Marker Simulation platform is available only in JMP Pro.

Contents

Example of Marker Simulation

Learn how to simulate genetic crosses from your genetic data.

1. Select **Help > Sample Data Folder** > **Life Sciences** and open Genotypes Pedigree.jmp and Genotypes Pedigree Anno.jmp.
2. Click on the Genotypes Pedigree.jmp table.
3. Select the first twenty rows in the table.

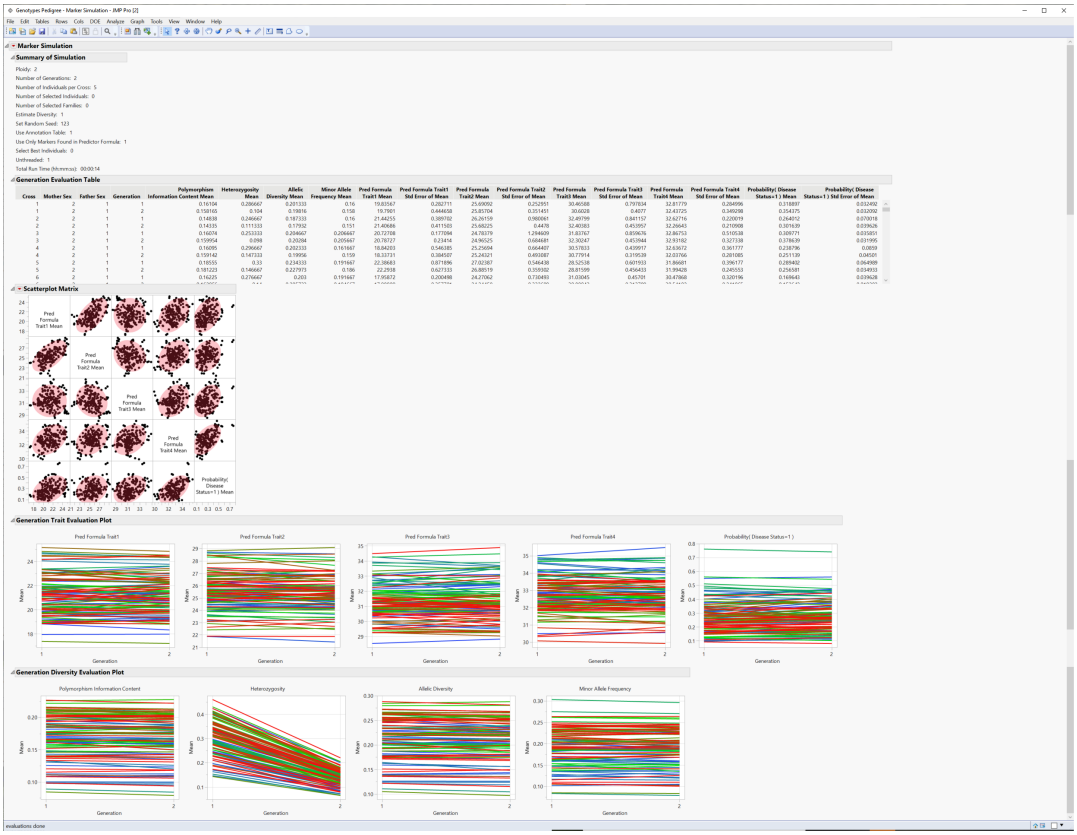
   Note: This example analyzes the first twenty rows only.

4. Select **Rows > Row Selection > Invert Row Selection**.
5. Select **Rows > Exclude/Unexclude** to exclude all but the first twenty rows from the analysis.
6. Select **Analyze > Genetics > Marker Simulation**.
7. Select Markers (60/0) and click **Marker**.
8. Select Pred Formula Trait1, Pred Formula Trait2, Pred Formula Trait3, Pred Formula Trait4, and Probability (Disease Status=1) and click **Predictor Formula**.
9. Select Sex and click **Cross**.
10. Enter 2 for **Ploidy**.
11. Enter 5 for the **Number of Individuals per Cross**.
12. Enter 2 for the **Number of Generations**.
13. Check the **Use Annotation Table**, **Use Only markers Found in the Predictor Formula**, and **Estimate Diversity** check boxes.
14. Under Advanced Options, enter 123 for **Set Random Seed**.
15. Check the **Unthreaded** check box.
16. Click **OK**.
17. In the Choose Annotation Table window, select the Genotypes Pedigree Anno.jmp annotation data table.
18. In the Annotation Specifications window, select Marker and click **Marker Variables**.
19. Select Gene and click **Annotation Group**.
20. Select Linkage Position and click **Annotation Position**.
21. Click **OK**.

The Marker Simulation report is generated.

22. Click the Marker Simulation red triangle and select **Show Evaluation Plot** and **Show Diversity Plot** to generate the report shown in Figure 4.1.

Figure 4.2 Results of the Simulated Crosses



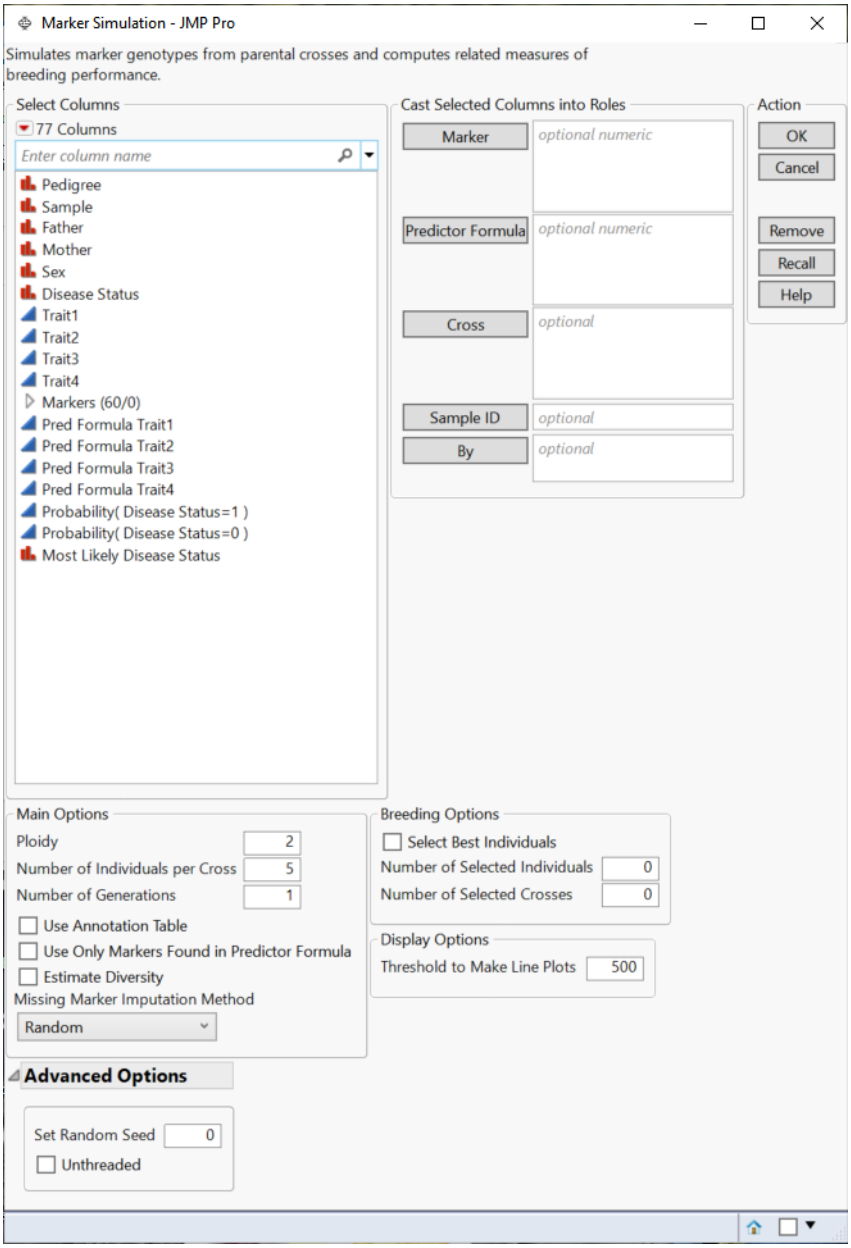
The Marker Simulation report contains the following information:

- The Summary of Simulation table summarizes the selections made on the platform and the time it took for the simulation to run.
- The Generation Evaluation Table provides genotypic and allelic statistics for each cross across generations along with predicted trait statistics.
- The Scatterplot matrix shows a pairwise comparison of the traits for each cross.
- The Evaluation Plot shows the change in the mean trait value across the generations for each cross.
- The Diversity Plot shows how polymorphism, heterozygosity, and allelic diversity and frequency change across generations.

Launch the Marker Simulation Platform

To simulate a genetic cross from your experimental data, launch the Marker Simulation platform by selecting **Analyze > Genetics > Marker Simulation**.

Figure 4.3 Marker Simulation Launch Window



Marker Select desired marker columns and click **Marker** to specify the markers that you want to analyze.

Predictor Formula Use this option to specify columns containing the predictor formulas. These formulas are developed on historical data where an event has been measured or inferred and are generated using one or more predictive modeling processes using predictive platforms in JMP, for example, Fit Model, Response Screening, XGBoost, etc. The predictive models are then applied to new data for which the attributes are known, but the event has not yet occurred.See Generating Predictor Formulas for Marker Simulation for details.

Note: Any trait column lacking a corresponding Predictor Formula column is ignored during the simulation.

Cross Use this option to specify the column to used differentiate the parents in the crosses. Specifying Sex, for example, directs the platform to cross parents of different sex (male with female) only.

Sample ID Use this option to specify one or more variables whose values that can, either singly or in combination, provide a unique identifier for each row.

By Produces a separate report for each level of the By variable. If more than one By variable is assigned, a separate report is produced for each possible combination of the levels of the By variables.

Ploidy Enables you to specify the ploidy level of the experimental organism under investigation. Note: This must be an even number

Number of Individuals per Cross Enables you to specify the number of replicates.

Number of Generations Specifies the number of generations.

Use Annotation Table Enables you to access annotation information contained in a separate data table. After you click **OK**, a window appears, prompting you to specify the name and location of the annotation table.

Use Only Markers Found in Predictor Formula Check this box to restrict the simulation to only those markers used to develop the predictor formulas.The algorithms used to generate the predictor formulas typically use some variable section method to select a subset of most significant markers in your data set. You can view the markers used by right-clicking on the column listing the trait predictors and selecting **Column Info**.

Estimate Diversity Check this box to calculate estimates of polymorphism, heterozygosity, and allelic diversity and frequency for the progeny of each cross.

Missing Marker Imputation Method This method does run when your data is missing marker data. Because of this, any missing data must be imputed. Use this option to specify how the missing values are to be imputed.

- Select **Random** to randomly assign one of the acceptable values (0, 1, 2, ..., *K* (where *K* is the ploidy level)).
- Select **Recessive Homozygous** to assign the minor allele value (0).
- Select **Dominant Homozygous** to assign the major allele value (2 for diploids, 4 for tetraploids, and in general, *K* (where *K* = ploidy level)).
- Select **Heterozygous** to assign the heterozygote value (1 for diploids, 1, 2 or 3 for tetraploids, and in general, any value expect 0 and *K* (where *K* = ploidy level)).

Select Best Individuals Check this box to select only the progeny that meet specified trait criteria, in each generation, for use in the subsequent cross. You must specify the selection criteria for each trait used for the selection. You can specify a lower limit, an upper limit, or a specific target value.

Specify a lower limit to select progeny with a trait value bigger than or equal to this limit to move to next generation. Select an upper limit to select progeny with a trait value lower than or equal to this limit to move to next generation. Specify a target value to select progeny with a trait value equal to this target to move to next generation.

Note: Specification of a target value is done when traits are non-continuous.

You can specify both an upper limit and a lower limit for any given trait to select only the progeny with trait values that fall within the interval formed by the upper and lower limit. Specification of a target value together with either an upper or a lower limit is not valid.

The final selection criterion is the intersection of all criteria specified for the traits. For example, if Spec Limits are such that, L1<= Trait1, L2 <= Trait2 <= U2, and Trait3 == T3, then the selection criterion is constructed to be L1<= Trait1 and L2 <= Trait2 <= U2 and Trait3 == T3. Any progeny that satisfies this criterion will be selected to next generation.

See Specifying Trait Selection Criteria for Marker Simulation for details about how to specify criteria for selecting progeny.

Note: This option is ignored unless Spec Limits have been specified for at least one of the Predictor Formula columns.

Number of Selected Individuals This option enables you to specify an upper limit of progeny meeting the trait selection criteria, in each generation, to use as parents in the subsequent cross. This limit is applied repeatedly for each subsequent generation.

Number of Selected Crosses This option enables you to specify an upper limit on the number of crosses meeting the trait selection criteria. Progeny from the previous cross are

assessed for the selection criteria and this limit is then applied, if needed, to the subsequent cross. This limit is applied repeatedly for each subsequent generation.

Threshold to Make Line Plots Generating line plots representing multi-generational crosses require substantial computer resources; too many can cause JMP to consume too much running time and the report to consume too much memory. Use this option to set an upper limit to the number of crosses used for generating the line plots. Should the number of crosses made exceed the specified value, JMP does not attempt to generate these plots.

Set Random Seed Use this option to specify a nonnegative integer to start the random number stream. Different values produce different outcomes of the algorithm.

Unthreaded Suppresses multi-threading. Deselect this option for improved computational speed.

Required Data Format for the Marker Simulation Platform

Most of the processes in JMP assume that the input table has a particular data structure. JMP distinguishes between tall and wide data sets. A tall data table has samples as columns and molecular entity (for example, marker, gene, clone, protein, or metabolite) as rows, whereas a wide data table is the transpose of the tall, having the samples as rows and molecular entity as columns.

When specifying the input data set for a process, it is important to know the required form. Marker Simulation requires a wide data table. The Transpose platform under the Tables menu enables you to transform your data tables between tall and wide forms.

Marker data must be encoded in the one-column, numeric format. Typically, in this format, diploid individuals homozygous for the least common, or minor allele, are represented in the table by a 2, whereas the heterozygotes are represented by a 1. Homozygotes for the most common allele are represented by a 0.

The Marker Simulation Report and Options

The Marker Simulation report enables you to explore the expected results of crosses of the different individuals in the study.

The Marker Simulation red triangle menu contains the following options:

Show Evaluation Plot Select this option to view the "Generation Trait Evaluation Plot". These plots show how the mean trait values compare and trend across generations.

Show Diversity Plot Select this option to view the "Generation Diversity Evaluation Plots". These plots show how the measures of diversity compare and trend across generations.

Save Simulated Table Use this option to save the simulated genotypes to a new data table.

Summary of Simulation

This section lists the selections made on the platform and the time it took for the simulation to run.

Generation Evaluation Table

The table lists various statistics for each cross:

Cross This column lists the cross numbers, starting at 1, up to the total of crosses.

Mother Sex This column lists the sex of the mother in the initial cross. This column name changes according to the name of column entered in **Cross** option. In this example, the column name was Sex, so Mother Sex appears. It always Mother + Name of column in **Cross** option.

Father Sex This column lists the sex of the father in the initial cross. This column name changes according to the name of column entered in **Cross** option. In this example, the column name was Sex, so Father Sex appears. It always Father + Name of column in **Cross** option.

Generation This column lists the generation the specific progeny is part of.

Minor Allele Frequency (MAF) Mean This column shows the average MAF across all markers. The MAF represents the proportion of the minor allele for each marker in the observed population. Assuming a biallelic marker locus M with alleles M_1 and M_2. A sample of N individuals of even ploidy k can therefore have $k + 1$ different genotypes at the locus. The number of individuals with i ($i = 0, 1, 2, \ldots, k$) copies of allele M_1 and j ($j = 0, 1, 2, \ldots, k$) copies of allele M_2 is denoted by N_{ij}. The number n_1 of copies of allele M_1 can be found directly by summation: $n_1 = 0 \times N_{00} + 1 \times N_{10} + 2 \times N_{20} + \ldots + k \times N_{k0}$. The sample frequency of allele M_1 is written as $p_1 = n_1 / (k \times \mathrm{n})$, frequency of allele M_2 is written as $p_2 = 1 - p_1$, and the sample frequency for each genotype carrying u copies of allele M_1 and v copies of allele M_2 is written as $P_{\mathrm{ij}} = N_{\mathrm{ij}} / n$.

Polymorphism Information Content (PIC) Mean This column shows the average PIC across all markers. The PIC (Botstein *et al*., 1980; Hilderbrand, Torney, and Wagner, 1992) measures the probability of differentiating the allele transmitted by a given parent to its child given the marker genotype of father, mother, and child.

$$PIC = 1 - \sum_{u=1}^{k} p_u^{\,2} - \sum_{u=1}^{k-1} \sum_{v=u+1}^{k} 2 p_u^{\,2} p_v^{\,2}$$

Heterozygosity (HET) Mean This column shows the average HET across all markers. The heterozygosity, sometimes called the observed heterozygosity, is simply the proportion of heterozygous individuals in the observed population.

$$Het = 1 - P_{ii} - P_{jj} (i < j)$$

Allelic Diversity Mean This column shows the average allelic diversity across all markers. The allelic diversity, sometimes called the expected heterozygosity, is the expected proportion of heterozygous individuals in the data set when Hardy-Weinberg equilibrium holds.

$$Div = 1 - p_1^2 - p_2^2$$

Pred Formula Trait Mean This column lists the mean of the predicted trait value for the progeny each cross at each generation.

Pred Formula Trait Standard Error of Mean This column lists the standard error of the mean of the predicted trait value for the progeny each cross at each generation. This value provides a measure of the expected variability.

Scatterplot Matrix

Shows a pairwise comparison of the traits. See *Discovering JMP*.

Generation Trait Evaluation Plot

These plots show how the mean trait values compare and trend across generations.

Note: These plots are not shown when the simulation is run for one generation only or when the total number of crosses exceeds the specified display threshold.

Generation Diversity Evaluation Plots

These plots show how the measures of diversity compare and trend across generations.

Note: You must check the **Estimate Diversity** check box to view these plots. These plots are not shown when the simulation is run for one generation only or when the total number of crosses exceeds the specified display threshold.

Appendix A

References

The following sources are referenced in the *Genetics* book.

Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). “Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms.” *American Journal of Human Genetics* 32:314–331.

Hildebrand, C. E., Torney, D. C., and Wagner, R. P. (1992). "Informativeness of Polymorphic DNA Markers." *Los Alamos Science* 20:100–102

Weir, B. S. (1979). “Inferences about Linkage Disequilibrium.” *Biometrics* 35:235–254.

Weir, B. S. (1996). *Genetic Data Analysis II*. Sunderland, MA: Sinauer Associates.

Zaykin, D. (2004). “Bounds and Normalization of the Composite Linkage Disequilibrium Coefficient.” *Genetic Epidemiology* 27:252–257.

Appendix B

Technology License Notices

- Scintilla is Copyright © 1998–2017 by Neil Hodgson <neilh@scintilla.org>.

  All Rights Reserved.

  Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation.

  NEIL HODGSON DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL NEIL HODGSON BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.
- Progressë® Telerik® UI for WPF: Copyright © 2008-2019 Progress Software Corporation. All rights reserved. Usage of the included Progress® Telerik® UI for WPF outside of JMP is not permitted.
- ZLIB Compression Library is Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.
- Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.
- Packages is Copyright © 2009–2010, Stéphane Sudre (s.sudre.free.fr). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  Neither the name of the WhiteBox nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED

WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- iODBC software is Copyright © 1995–2006, OpenLink Software Inc and Ke Jin (www.iodbc.org). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of OpenLink Software Inc. nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL OPENLINK OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- This program, "bzip2", the associated library "libbzip2", and all documentation, are Copyright © 1996–2019 Julian R Seward. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. The origin of this software must not be misrepresented; you must not claim that you wrote the original software. If you use this software in a product, an acknowledgment in the product documentation would be appreciated but is not required.

3. Altered source versions must be plainly marked as such, and must not be misrepresented as being the original software.

4. The name of the author may not be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE AUTHOR "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Julian Seward, jseward@acm.org

bzip2/libbzip2 version 1.0.8 of 13 July 2019

- R software is Copyright © 1999–2012, R Foundation for Statistical Computing.
- MATLAB software is Copyright © 1984-2012, The MathWorks, Inc. Protected by U.S. and international patents. See www.mathworks.com/patents. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.
- libopc is Copyright © 2011, Florian Reuter. All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and / or other materials provided with the distribution.
  – Neither the name of Florian Reuter nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR

ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- libxml2 - Except where otherwise noted in the source code (e.g. the files hash.c, list.c and the trio files, which are covered by a similar license but with different Copyright notices) all the files are:

  Copyright © 1998–2003 Daniel Veillard. All Rights Reserved.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

  THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL DANIEL VEILLARD BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of Daniel Veillard shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization from him.

- Regarding the decompression algorithm used for UNIX files:

  Copyright © 1985, 1986, 1992, 1993

  The Regents of the University of California. All rights reserved.

  THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

- Snowball is Copyright © 2001, Dr Martin Porter, Copyright © 2002, Richard Boulton.

  All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS \"AS IS\" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- Pako is Copyright © 2014–2017 by Vitaly Puzrin and Andrei Tuputcyn.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

- HDF5 (Hierarchical Data Format 5) Software Library and Utilities are Copyright 2006–2015 by The HDF Group. NCSA HDF5 (Hierarchical Data Format 5) Software Library and Utilities Copyright 1998-2006 by the Board of Trustees of the University of Illinois. All rights reserved. DISCLAIMER: THIS SOFTWARE IS PROVIDED BY THE HDF GROUP AND THE CONTRIBUTORS "AS IS" WITH NO WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED. In no event shall The HDF Group or the Contributors be liable for any damages suffered by the users arising out of the use of this software, even if advised of the possibility of such damage.
- agl-aglfn technology is Copyright © 2002, 2010, 2015 by Adobe Systems Incorporated. All Rights Reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  – Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
  – Neither the name of Adobe Systems Incorporated nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT

(INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- dmlc/xgboost is Copyright © 2019 SAS Institute.

  Licensed under the Apache License, Version 2.0 (the “License”); you may not use this file except in compliance with the License. You may obtain a copy of the License at

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libzip is Copyright © 1999–2019 Dieter Baron and Thomas Klausner.

  This file is part of libzip, a library to manipulate ZIP archives. The authors can be contacted at <libzip@nih.at>.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  3. The names of the authors may not be used to endorse or promote products derived from this software without specific prior written permission.

  THIS SOFTWARE IS PROVIDED BY THE AUTHORS ``AS IS'' AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- OpenNLP 1.5.3, the pre-trained model (version 1.5 of en-parser-chunking.bin), and dmlc/xgboost Version .90 are licensed under the Apache License 2.0 are Copyright © January 2004 by Apache.org.

You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

- You must give any other recipients of the Work or Derivative Works a copy of this License; and
- You must cause any modified files to carry prominent notices stating that You changed the files; and
- You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
- If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.
- You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

- LLVM is Copyright © 2003–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- clang is Copyright © 2007–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS", WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- lld is Copyright © 2011–2019 by the University of Illinois at Urbana-Champaign.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at:

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS", WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libcurl is Copyright © 1996–2021, Daniel Stenberg, daniel@haxx.se, and many contributors, see the THANKS file. All rights reserved.

  Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

  THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OF THIRD PARTY RIGHTS. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

  Except as contained in this notice, the name of a copyright holder shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization of the copyright holder.

- On the Windows operating system, JMP utilizes the OpenBLAS library. OpenBLAS is licensed under the 3-clause BSD license. Full license text follows: Copyright © 2011-2015, The OpenBLAS Project

  All rights reserved. Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of the OpenBLAS project nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE OPENBLAS PROJECT OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.