

# root

---

[Go Up](#)

Name	GLM
Version	3.0.1
Description	Generalized Linear Model implementation
License	<a href="http://www.apache.org/licenses/LICENSE-2.0">http://www.apache.org/licenses/LICENSE-2.0</a>
Copyright	Copyright (C) 2017 HPCC Systems®
Authors	HPCCSystems
DependsOn	ML_Core 3.1.1, PBblas
Platform	6.2.0

## Table of Contents

<a href="#">Apply2CellsBinary.ecl</a>	Iterate matrix and apply function to each pair of cells
<a href="#">BinomialConfusion.ecl</a>	Calculate the Binomial confusion matrix
<a href="#">Confusion.ecl</a>	Generate the confusion matrix, to compare actual versus predicted response variable values
<a href="#">Constants.ecl</a>	Constants used by the GLM bundle
<a href="#">DataStats.ecl</a>	Produce summary information about the datasets
<a href="#">Deviance_Analysis.ecl</a>	Analysis of Deviance Report
<a href="#">Deviance_Detail.ecl</a>	Deviance detail report
<a href="#">dimm.ecl</a>	Matrix multiply when either A or B is a diagonal and is passed as a vector
<a href="#">enum_workitems.ecl</a>	

Create an enumeration of string contents to be used as work items

[ExtractBeta.ecl](#)

Extract the beta values form the model dataset

[ExtractBeta\\_CI.ecl](#)

Extract the beta values and confidence intervals from the model dataset

[ExtractBeta\\_full.ecl](#)

Extract the coefficient information including confidence intervals, z and p values

[ExtractBeta\\_pval.ecl](#)

Extract the beta values including z and p value from the model

[ExtractReport.ecl](#)

Create a model report from a model

[Family.ecl](#)

Definitions of supported families of Linear Models

[GLM.ecl](#)

Main GLM regression module

[LogitPredict.ecl](#)

Predict the category values with the logit function and the the supplied beta coefficients

[LUCI\\_Model.ecl](#)

Create a LUCI model file description of the model(s) from the external version of the model

[Model\\_Deviance.ecl](#)

Model Deviance Report

[Named\\_Model.ecl](#)

Apply external labels for work items and field names to a model

[Null\\_Deviance.ecl](#)

Return Deviance information for the null model, that is, a model with only an intercept

[Predict.ecl](#)

Calculate the score using the appropriate mean function and the the supplied beta coefficients

[Types.ecl](#)

Type definitions for GLM bundle

# Apply2CellsBinary

---

[Go Up](#)

## IMPORTS

```
std.blas | std.BLAS.Types |
```

## DESCRIPTIONS

### **APPLY2CELLSBINARY Apply2CellsBinary**

/ EXPORT Types.matrix_t	Apply2CellsBinary
(Types.dimension_t m = 1, Types.dimension_t n = 1, Types.matrix_t x = [], Types.matrix_t y = [], ICellFuncBinary f = ICellFuncBinary)	

Iterate matrix and apply function to each pair of cells.

**PARAMETER** **m** ||| UNSIGNED4 — number of rows

**PARAMETER** **n** ||| UNSIGNED4 — number of columns

**PARAMETER** **x** ||| SET ( REAL8 ) — matrix

**PARAMETER** **y** ||| SET ( REAL8 ) — matrix

**PARAMETER** **f** ||| FUNCTION [ REAL8 , REAL8 , UNSIGNED4 , UNSIGNED4 ] ( REAL8 ) —  
function to apply

**RETURN** SET ( REAL8 ) — updated matrix

# BinomialConfusion

---

[Go Up](#)

## IMPORTS

Types | \_\_versions.ML\_Core.V3\_2\_2.ML\_Core.Types |

## DESCRIPTIONS

### **BINOMIALCONFUSION** BinomialConfusion

<code>DATASET(</code> <i>Types.Binomial_Confusion_Summary</i> <code>)</code>	BinomialConfusion
<code>(DATASET(</code> <i>Core_Types.Confusion_Detail</i> <code>)</code> <i>d</i> )	

Calculate the Binomial confusion matrix. Work items with multinomial responses are ignored by this function. The higher value lexically is considered to be the positive indication.

**PARAMETER** *d* ||| TABLE ( *Confusion\_Detail* ) — confusion detail for the work item and classifier.

**RETURN** TABLE ( { UNSIGNED2 *wi* , UNSIGNED4 *classifier* , UNSIGNED8 *true\_positive* , UNSIGNED8 *true\_negative* , UNSIGNED8 *false\_positive* , UNSIGNED8 *false\_negative* , UNSIGNED8 *cond\_pos* , UNSIGNED8 *pred\_pos* , UNSIGNED8 *cond\_neg* , UNSIGNED8 *pred\_neg* , REAL8 *prevalence* , REAL8 *accuracy* , REAL8 *true\_pos\_rate* , REAL8 *false\_neg\_rate* , REAL8 *false\_pos\_rate* , REAL8 *true\_neg\_rate* , REAL8 *pos\_pred\_val* , REAL8 *false\_disc\_rate* , REAL8 *false omit\_rate* , REAL8 *neg\_pred\_val* } ) — confusion matrix for a binomial classifier.

---

# Confusion

---

[Go Up](#)

## IMPORTS

```
_versions.ML_Core.V3_2_2.ML_Core |  
_versions.ML_Core.V3_2_2.ML_Core.Types | Types |
```

## DESCRIPTIONS

### **CONFUSION** Confusion

/ EXPORT DATASET(Confusion_Detail)	Confusion
(DATASET(DiscreteField) dependents, DATASET(DiscreteField) predicts)	

Generate the confusion matrix, to compare actual versus predicted response variable values.

**PARAMETER** dependents ||| TABLE ( DiscreteField ) — the original response values.

**PARAMETER** predicts ||| TABLE ( DiscreteField ) — the predicted responses.

**RETURN** TABLE ( { UNSIGNED2 wi , UNSIGNED4 classifier , INTEGER4 actual\_class  
, INTEGER4 predict\_class , UNSIGNED4 occurs , BOOLEAN correct , REAL8  
pctActual , REAL8 pctPred } ) — confusion matrix in Confusion\_Detail format.

**SEE** ML\_Core.Types.Confusion\_Detail

# Constants

---

[Go Up](#)

## **DESCRIPTIONS**

### **CONSTANTS** Constants

#### Constants

Constants used by the GLM bundle. Most of these are the nominal values used by the Model data set. A few are used to control behavior.

#### Children

1. [limit\\_card](#) : No Documentation Found
2. [default\\_epsilon](#) : No Documentation Found
3. [default\\_ridge](#) : No Documentation Found
4. [local\\_cap](#) : No Documentation Found
5. [id\\_base](#) : No Documentation Found
6. [id\\_iters](#) : No Documentation Found
7. [id\\_delta](#) : No Documentation Found
8. [id\\_mse](#) : No Documentation Found
9. [id\\_dispersion](#) : No Documentation Found
10. [id\\_stat\\_set](#) : No Documentation Found
11. [id\\_betas](#) : No Documentation Found
12. [id\\_betas\\_coef](#) : No Documentation Found
13. [id\\_betas\\_SE](#) : No Documentation Found
14. [base\\_builder](#) : No Documentation Found

15. [base\\_max\\_iter](#) : No Documentation Found
  16. [base\\_epsilon](#) : No Documentation Found
  17. [base\\_ind\\_vars](#) : No Documentation Found
  18. [base\\_dep\\_vars](#) : No Documentation Found
  19. [base\\_obs](#) : No Documentation Found
  20. [builder\\_irls\\_local](#) : No Documentation Found
  21. [builder\\_irls\\_global](#) : No Documentation Found
  22. [builder\\_softmax](#) : No Documentation Found
- 

## **LIMIT\_CARD limit\_card**

[Constants](#) \

<b>UNSIGNED2</b>	<b>limit_card</b>
------------------	-------------------

No Documentation Found

**RETURN** **UNSIGNED2** —

---

## **DEFAULT\_EPSILON default\_epsilon**

[Constants](#) \

<b>REAL8</b>	<b>default_epsilon</b>
--------------	------------------------

No Documentation Found

**RETURN** **REAL8** —

---

## **DEFAULT\_RIDGE** default\_ridge

Constants \

REAL8	default_ridge
-------	---------------

No Documentation Found

---

**RETURN** REAL8 —

---

## **LOCAL\_CAP** local\_cap

Constants \

UNSIGNED4	local_cap
-----------	-----------

No Documentation Found

---

**RETURN** UNSIGNED4 —

---

## **ID\_BASE** id\_base

Constants \

	id_base
--	---------

No Documentation Found

---

**RETURN** INTEGER8 —

---

## **ID\_ITERS** id\_iters

Constants \

	id_iters
--	----------

No Documentation Found

---

**RETURN** INTEGER8 —

---

## **ID\_DELTA** id\_delta

Constants \

	id_delta
--	----------

No Documentation Found

---

**RETURN** INTEGER8 —

---

## **ID\_MSE** id\_mse

Constants \

	id_mse
--	--------

No Documentation Found

---

**RETURN** INTEGER8 —

---

## **ID\_DISPERSION** `id_dispersion`

[Constants](#) \

	<code>id_dispersion</code>
--	----------------------------

No Documentation Found

---

**RETURN** INTEGER8 —

---

## **ID\_STAT\_SET** `id_stat_set`

[Constants](#) \

	<code>id_stat_set</code>
--	--------------------------

No Documentation Found

---

**RETURN** SET ( INTEGER8 ) —

---

## **ID\_BETAS** `id_betas`

[Constants](#) \

	<code>id_betas</code>
--	-----------------------

No Documentation Found

---

**RETURN** INTEGER8 —

---

## **ID\_BETAS\_COEF** id\_betas\_coef

Constants \

	id_betas_coef
--	---------------

No Documentation Found

---

**RETURN** INTEGER8 —

---

## **ID\_BETAS\_SE** id\_betas\_SE

Constants \

	id_betas_SE
--	-------------

No Documentation Found

---

**RETURN** INTEGER8 —

---

## **BASE\_BUILDER** base\_builder

Constants \

	base_builder
--	--------------

No Documentation Found

---

**RETURN** INTEGER8 —

---

## **BASE\_MAX\_ITER** `base_max_iter`

Constants \

	<code>base_max_iter</code>
--	----------------------------

No Documentation Found

---

**RETURN** INTEGER8 —

---

## **BASE\_EPSILON** `base_epsilon`

Constants \

	<code>base_epsilon</code>
--	---------------------------

No Documentation Found

---

**RETURN** INTEGER8 —

---

## **BASE\_IND\_VARS** `base_ind_vars`

Constants \

	<code>base_ind_vars</code>
--	----------------------------

No Documentation Found

---

**RETURN** INTEGER8 —

---

## **BASE\_DEP\_VARS** `base_dep_vars`

Constants \

	<code>base_dep_vars</code>
--	----------------------------

No Documentation Found

---

**RETURN** INTEGER8 —

---

## **BASE\_OBS** `base_obs`

Constants \

	<code>base_obs</code>
--	-----------------------

No Documentation Found

---

**RETURN** INTEGER8 —

---

## **BUILDER\_IRLS\_LOCAL** `builder_irls_local`

Constants \

	<code>builder_irls_local</code>
--	---------------------------------

No Documentation Found

---

**RETURN** INTEGER8 —

---

## **BUILDER\_IRLS\_GLOBAL** `builder_irls_global`

[Constants](#) \

	<code>builder_irls_global</code>
--	----------------------------------

No Documentation Found

---

**RETURN** `INTEGER8` —

---

## **BUILDER\_SOFTMAX** `builder_softmax`

[Constants](#) \

	<code>builder_softmax</code>
--	------------------------------

No Documentation Found

---

**RETURN** `INTEGER8` —

---

# DataStats

---

[Go Up](#)

## IMPORTS

Types | Constants | Family | \_\_versions.ML\_Core.V3\_2\_2.ML\_Core.Types |

## DESCRIPTIONS

### **DATASTATS** DataStats

/ EXPORT DATASET(Types.Data_Info)	DataStats
(DATASET(Core_Types.NumericField) indep, DATASET(Core_Types.NumericField) dep, BOOLEAN dep_details=TRUE, BOOLEAN ind_details=FALSE, Family.FamilyInterface fam=Family.Gaussian)	

Produce summary information about the datasets.

When dep\_details or ind\_details = FALSE, indicates the range for the x or y (independent or dependent) columns.

When dep\_details or ind\_details = TRUE, the cardinality, minimum, and maximum values are returned. A zero cardinality is returned when the field cardinality exceeds the Constants.limit\_card value.

Note that a column of all zero values cannot be distinguished from a missing column.

**PARAMETER** indep ||| TABLE ( NumericField ) — data set of independent variables.

**PARAMETER** dep ||| TABLE ( NumericField ) — data set of dependent variables.

**PARAMETER** dep\_details ||| BOOLEAN — Boolean directive to provide dependent field level info.

**PARAMETER** field\_details ||| — Boolean directive to provide independent field level info.

**PARAMETER** ind\_details ||| BOOLEAN — No Doc

**PARAMETER** fam ||| INTERFACE ( FamilyInterface ) — No Doc

**RETURN** TABLE ( { UNSIGNED2 wi , UNSIGNED4 dependent\_fields , UNSIGNED4 dependent\_records , UNSIGNED4 independent\_fields , UNSIGNED4 independent\_records , UNSIGNED4 dependent\_count , UNSIGNED4 independent\_count , TABLE ( Field\_Desc ) dependent\_stats , TABLE ( Field\_Desc ) independent\_stats } ) — a data set of information on each work item in Data\_Info format.

**SEE** Types.Data\_Info

**SEE** Constants.limit\_card

---

# Deviance\_Analysis

---

[Go Up](#)

## IMPORTS

Types | \_versions.ML\_Core.V3\_2\_2.ML\_Core.Math |

## DESCRIPTIONS

### **DEVIANCE\_ANALYSIS** Deviance\_Analysis

<b>PARAMETER</b> proposed     TABLE ( Deviance_Record ) — deviance records of the proposed model.	Deviance_Analysis
(DATASET(Types.Deviance_Record) proposed, DATASET(Types.Deviance_Record) base)	

Analysis of Deviance Report.

Compare deviance information between two models, a base and and proposed model.

Analysis of Deviance is analogous to the Analysis of Variance (ANOVA) used in least-squares modeling, but adapted to the general linear model (GLM). In this case it is adapted specifically to the logistic model.

The inputs are the deviance records for each model as obtained from a call to Model\_Deviance.

**PARAMETER** proposed ||| TABLE ( Deviance\_Record ) — deviance records of the proposed model.

**PARAMETER** base ||| TABLE ( Deviance\_Record ) — deviance records of the base model for comparison.

**RETURN** TABLE ( { UNSIGNED2 wi , UNSIGNED4 model , UNSIGNED8 residual\_df ,  
REAL8 df , REAL8 residual\_dev , REAL8 deviance , REAL8 p\_value } ) — the comparison of the deviance between the models in AOD\_Record format.

**SEE** Model\_Deviance

**SEE** Types.Deviance\_Record

**SEE** Types.AOD\_Record

---

# Deviance\_Detail

---

[Go Up](#)

## IMPORTS

```
_versions.ML_Core.V3_2_2.ML_Core |  
_versions.ML_Core.V3_2_2.ML_Core.Types | Types | IRLS | Family | Constants |
```

## DESCRIPTIONS

### **DEVIANCEDIETAIL Deviance\_Detail**

/ EXPORT DATASET(Types.Observation_Deviance)	Deviance_Detail
(DATASET(NumericField) dependents, DATASET(NumericField) predicts, DATASET(Layout_Model) model, Family.FamilyInterface fam)	

Deviance detail report.

Provides deviance information for each observation.

Analysis of Deviance is analogous to the Analysis of Variance (ANOVA) used in least-squares modeling, but adapted to the general linear model (GLM).

**PARAMETER dependents** ||| TABLE ( NumericField ) — original dependent records for the model.

**PARAMETER predicts** ||| TABLE ( NumericField ) — the predicted values of the response variable.

**PARAMETER model** ||| TABLE ( Layout\_Model ) — the fitted model object as returned from GetModel.

**PARAMETER** **fam** ||| INTERFACE ( FamilyInterface ) — a module defining the error distribution and link of the dependents

**RETURN** TABLE ( { UNSIGNED2 wi , UNSIGNED8 id , UNSIGNED4 model , REAL8 actual , REAL8 predicted , REAL8 mod\_ll , REAL8 mod\_dev\_component , REAL8 mod\_dev\_residual , REAL8 nil\_ll , REAL8 nil\_dev\_component , REAL8 nil\_dev\_residual } ) — the deviance information by observation and the log likelihood of the predicted result in Observation\_Deviance format.

**SEE** Types.Observation\_Deviance

---

# dimm

---

[Go Up](#)

## IMPORTS

std.blas | std.BLAS.Types |

## DESCRIPTIONS

### **DIMM dimm**

<code>/ EXPORT Types.matrix_t</code>	<code>dimm</code>
<code>(BOOLEAN transposeA, BOOLEAN transposeB, BOOLEAN diagonalA, BOOLEAN diagonalB, Types.dimension_t m, Types.dimension_t n, Types.dimension_t k, Types.value_t alpha, Types.matrix_t A, Types.matrix_t B, Types.value_t beta=0.0, Types.matrix_t C=[])</code>	

Matrix multiply when either A or B is a diagonal and is passed as a vector.

Computes:  $\text{alpha} * \text{op}(A) \text{ op}(B) + \text{beta} * C$  where  $\text{op}()$  is transpose.

**PARAMETER** `transposeA` ||| BOOLEAN — true when transpose of A is used.

**PARAMETER** `transposeB` ||| BOOLEAN — true when transpose of B is used.

**PARAMETER** `diagonalA` ||| BOOLEAN — true when A is the diagonal matrix.

**PARAMETER** `diagonalB` ||| BOOLEAN — true when B is the diagonal matrix.

**PARAMETER** `m` ||| UNSIGNED4 — number of rows in product.

**PARAMETER** `n` ||| UNSIGNED4 — number of columns in product.

**PARAMETER** k ||| UNSIGNED4 — number of columns/rows for the multiplier/multiplicand.

**PARAMETER** alpha ||| REAL8 — scalar used on A.

**PARAMETER** A ||| SET ( REAL8 ) — matrix A.

**PARAMETER** B ||| SET ( REAL8 ) — matrix B.

**PARAMETER** beta ||| REAL8 — scalar for matrix C.

**PARAMETER** C ||| SET ( REAL8 ) — matrix C or empty.

**RETURN** SET ( REAL8 ) — result matrix in matrix\_t format.

**SEE** Std.BLAS.Types.matrix\_t

---

# enum\_workitems

---

[Go Up](#)

## DESCRIPTIONS

### **ENUM\_WORKITEMS** enum\_workitems

/ EXPORT	enum_workitems
	(dsIn, dsOut, src_field, wi_name)

Create an enumeration of string contents to be used as work items.

This macro produces 2 external symbols, dsOut and dsOut\_Map.

The dsOut extends the input dataset with a numeric work-item number.

The dsOut\_Map dataset captures the relationship between the strings that name the work items and the nominal assigned in Workitem\_Mapping format.

**PARAMETER** dsIn ||| INTEGER8 — the input recordset.

**PARAMETER** dsOut ||| INTEGER8 — the symbol to use for the appended data.

**PARAMETER** src\_field ||| INTEGER8 — a field name to use to discriminate work-items.

**PARAMETER** wi\_name ||| INTEGER8 — the field name for the work item value assigned.

**RETURN** — Nothing. The macro creates the symbols 'dsOut' and 'dsOut\_Map' inline.

**SEE** Types.Workitem\_Mapping

---

# ExtractBeta

---

[Go Up](#)

## IMPORTS

Types | \_\_versions.ML\_Core.V3\_2\_2.ML\_Core.Types |

## DESCRIPTIONS

### **EXTRACTBETA ExtractBeta**

ExtractBeta
(DATASET(Core_Types.Layout_Model) mod_ds)

Extract the beta values form the model dataset.

**PARAMETER** mod\_ds ||| TABLE ( Layout\_Model ) — the model as returned from GetModel.

**RETURN** TABLE ( { UNSIGNED2 wi , UNSIGNED4 ind\_col , UNSIGNED4 dep\_nom , REAL8 w , REAL8 SE } ) — the beta values as Model\_Coef records, with zero as the constant term.

**SEE** Types.Model\_Coef

---

# ExtractBeta\_CI

---

[Go Up](#)

## IMPORTS

Types | \_\_versions.ML\_Core.V3\_2\_2.ML\_Core.Types |

## DESCRIPTIONS

### **EXTRACTBETA\_CI ExtractBeta\_CI**

<code>DATASET(</code> <i>Types</i> .Confidence_Model_Coef)	<code>) ExtractBeta_CI</code>
<code>(</code> <i>DATASET</i> ( <i>Core_Types</i> .Layout_Model) <i>mod_ds</i> , REAL8 <i>level</i> )	

Extract the beta values and confidence intervals from the model dataset.

**PARAMETER** *mod\_ds* ||| TABLE ( Layout\_Model ) — the model as returned from GetModel.

**PARAMETER** *level* ||| REAL8 — the significance value for the intervals.

**RETURN** TABLE ( { UNSIGNED2 *wi* , UNSIGNED4 *ind\_col* , UNSIGNED4 *dep\_nom* ,  
REAL8 *w* , REAL8 *SE* , REAL8 *upper* , REAL8 *lower* } ) — the beta values with  
confidence intervals in Confidence\_Model\_Coef format, with zero as the constant term.

**SEE** *Types*.Confidence\_Model\_Coef

---

# ExtractBeta\_full

---

[Go Up](#)

## IMPORTS

```
Types | _versions.ML_Core.V3_2_2.ML_Core.Math |
_versions.ML_Core.V3_2_2.ML_Core.Types |
```

## DESCRIPTIONS

### **EXTRACTBETA\_FULL ExtractBeta\_full**

<b>PARAMETER</b> <code>mod_ds</code>     TABLE ( Layout_Model ) — the model as returned from GetModel.	<b>ExtractBeta_full</b>
( <b>PARAMETER</b> <code>level</code>     REAL8 <code>level=0.05</code> )	

Extract the coefficient information including confidence intervals, z and p values.

**PARAMETER** `mod_ds` ||| TABLE ( Layout\_Model ) — the model as returned from GetModel.

**PARAMETER** `level` ||| REAL8 — the significance value for the intervals.

**RETURN** TABLE ( { UNSIGNED2 wi , UNSIGNED4 ind\_col , UNSIGNED4 dep\_nom ,  
REAL8 w , REAL8 SE , REAL8 z , REAL8 p\_value , REAL8 upper , REAL8 lower }  
) — the coefficient information for the model in Full\_Model\_Coef format, with zero as the constant term.

**SEE** Types.Full\_Model\_Coef

# ExtractBeta\_pval

---

[Go Up](#)

## IMPORTS

Types | \_\_versions.ML\_Core.V3\_2\_2.ML\_Core.Types |

## DESCRIPTIONS

### **EXTRACTBETA\_PVAL ExtractBeta\_pval**

/ EXPORT DATASET( Types.pval_Model_Coef )	ExtractBeta_pval
(DATASET(Core_Types.Layout_Model) mod_ds)	

Extract the beta values including z and p value from the model.

**PARAMETER** mod\_ds ||| TABLE ( Layout\_Model ) — the model as returned from GetModel.

**RETURN** TABLE ( { UNSIGNED2 wi , UNSIGNED4 ind\_col , UNSIGNED4 dep\_nom , REAL8 w , REAL8 SE , REAL8 z , REAL8 p\_value } ) — the beta values with p-values in pval\_Model\_Coef format, with zero as the constant term.

**SEE** Types.pval\_Model\_Coef

---

# ExtractReport

---

[Go Up](#)

## IMPORTS

Types | Constants | \_\_versions.ML\_Core.V3\_2\_2.ML\_Core.Types |

## DESCRIPTIONS

### **EXTRACTREPORT ExtractReport**

<b>DATASET(</b> Types.Model_Report <b>)</b>	ExtractReport
(	DATASET(Core_Types.Layout_Model) mod_ds)

Create a model report from a model.

**PARAMETER** mod\_ds ||| TABLE ( Layout\_Model ) — the model as returned from GetModel.

**RETURN** TABLE ( { UNSIGNED2 wi , UNSIGNED4 max\_iterations , REAL8 epsilon ,  
UNSIGNED4 dep\_vars , UNSIGNED4 ind\_vars , UNSIGNED8 obs , UNSIGNED2  
builder , TABLE ( Regressor\_Stats ) stats } ) — the model report in Model\_Report format.

**SEE** Types.Model\_Report

---

# Family

---

[Go Up](#)

## IMPORTS

```
_versions.ML_Core.V3_2_2.ML_Core |
```

## DESCRIPTIONS

### **FAMILY Family**

	Family
--	--------

Definitions of supported families of Linear Models.

Currently supported families are:

- Binomial
- Quasibinomial
- Poisson
- Quasipoisson
- Gamma
- Gaussian
- InvGaussian

In addition, FamilyInterface defines the interface needed to add new families.

Adding new families is fairly straightforward and involves overlaying a set of scalar functions that define the computations for that family. See FamilyInterface below.

## Children

1. [FamilyInterface](#) : Defines the interface to create new GLM Regression Families
  2. [Binomial](#) : The Binomial Regression Family models the response (dependent variable(s)) as a series of Bernoulli Trials, of one of two disjoint outcomes
  3. [Quasibinomial](#) : The Quasibinomial Regression Family is similar to the Binomial family (see Binomial above) except that it adjusts for situations where the variance of the distribution is greater or less than anticipated by the Binomial Distribution
  4. [Poisson](#) : Poisson Regression is generally used to model count data, where the dependent variable is a positive (or zero) integer
  5. [Quasipoisson](#) : Quasipoisson Regression is similar to Poisson Regression (see Poisson above) except that it adjusts for situations where the variance of the distribution is greater or less than anticipated by the Poisson Distribution
  6. [Gamma](#) : Gamma Regression is used to model continuous,non-negative, data with a right-skew
  7. [Gaussian](#) : Gaussian Regression is equivalent to Ordinary Least Squares (OLS) regression
  8. [InvGauss](#) : Inverse Gaussian Regression aka Wald Regression is similar to the Gamma Regression in that it is used to model continuous, positive heteroskedastic data
- 

## FAMILYINTERFACE [FamilyInterface](#)

[Family](#) \

	<a href="#">FamilyInterface</a>
--	---------------------------------

Defines the interface to create new GLM Regression Families. Each family defines a series of eleven attributes that describe the computations for that family within the overall GLM model.

## Children

1. [link](#) : This function defines the linkage between output of the linear function on independent data and the dependent data
2. [mu](#) : The Mean function is the inverse of the link function
3. [deta](#) : The derivative of the output of the linear function with respect to the expected value of the dependent variable

4. `var` : The variance as a function of the output value
  5. `init` : Initialization transform sets the initial value for Betas when running Iteratively Re-weighted Least Squares (IRLS)
  6. `ll` : Log Likelihood function
  7. `mu_LUCI` : The string representation of the mu function (see mu above) for use in LUCI
  8. `dispersion` : Flag indicating whether the error distribution should be adjusted for over-dispersion or under-dispersion
  9. `cardinality` : The minimum and maximum cardinality (i.e
  10. `values` : The range of values that the dependent data can take
  11. `isInteger` : Flag that indicates that the dependent variables can only take Integer values
- 

## **LINK link**

Family \ FamilyInterface \

<code>REAL8</code>	<code>link</code>
	<code>(REAL8 m)</code>

This function defines the linkage between output of the linear function on independent data and the dependent data.

**PARAMETER** `m` ||| `REAL8` — The output from the linear function (i.e. the mean)

**RETURN** `REAL8` — The value to be compared to the dependent data.

---

## **MU mu**

Family \ FamilyInterface \

<code>REAL8</code>	<code>mu</code>
	<code>(REAL8 v)</code>

The Mean function is the inverse of the link function. It maps the expected value of the dependent variable to the expected linear result.

**PARAMETER** `v` ||| REAL8 — The expected value of the dependent variable.

**RETURN** REAL8 — The expected output from the linear function.

---

## DETA data

Family \ FamilyInterface \

REAL8	<b>data</b>
(REAL8 m)	

The derivative of the output of the linear function with respect to the expected value of the dependent variable.

**PARAMETER** `m` ||| REAL8 — The value of the output.

**RETURN** REAL8 — The derivative at m.

---

## VAR var

Family \ FamilyInterface \

REAL8	<b>var</b>
(REAL8 m)	

The variance as a function of the output value. This is used for heteroskedastic distributions, otherwise 1.

**PARAMETER** `m` ||| REAL8 — The value of the output.

**RETURN** REAL8 — The expected variance when output is at m

---

## INIT init

Family \ FamilyInterface \

REAL8	init
(REAL8 y, REAL8 w)	

Initialization transform sets the initial value for Betas when running Iteratively Re-weighted Least Squares (IRLS).

**PARAMETER** y ||| REAL8 — the dependent value.

**PARAMETER** w ||| REAL8 — the current weight.

**RETURN** REAL8 — the initial weight value to use.

---

## LL ll

Family \ FamilyInterface \

REAL8	ll
(REAL8 y, // log-likelihood function REAL8 m, REAL8 disp)	

Log Likelihood function.

**PARAMETER** y ||| REAL8 — The dependent variable.

**PARAMETER** m ||| REAL8 — The output value.

**PARAMETER** disp ||| REAL8 — The dispersion factor

**RETURN** REAL8 — The log likelihood of seeing m given y.

---

## MU\_LUCI mu\_LUCI

Family \ FamilyInterface \

STRING	mu_LUCI
--------	---------

The string representation of the mu function (see mu above) for use in LUCI. See LUCI guide for formatting of this ECL string.

**RETURN** STRING —

**RETURNS** An ECL string representation of the mu function.

---

## DISPERSION dispersion

Family \ FamilyInterface \

BOOLEAN	dispersion
---------	------------

Flag indicating whether the error distribution should be adjusted for over-dispersion or under-dispersion.

**RETURN** BOOLEAN —

---

## CARDINALITY cardinality

Family \ FamilyInterface \

SET OF UNSIGNED4	cardinality
------------------	-------------

The minimum and maximum cardinality (i.e. number of unique values) for dependent data.

**RETURN** SET ( UNSIGNED4 ) — SET([min\_cardinality, max\_cardinality])

---

## **VALUES** values

Family \ FamilyInterface \

SET OF REAL8	values
--------------	--------

The range of values that the dependent data can take.

**RETURN** SET ( REAL8 ) — SET([min\_value, max\_value])

---

## **ISINTEGER** isInteger

Family \ FamilyInterface \

BOOLEAN	isInteger
---------	-----------

Flag that indicates that the dependent variables can only take Integer values. If FALSE, then REAL values are supported.

**RETURN** BOOLEAN — Boolean indicating if output is restricted to Integer values.

---

## **BINOMIAL** Binomial

Family \

	Binomial
--	----------

The Binomial Regression Family models the response (dependent variable(s)) as a series of Bernoulli Trials, of one of two disjoint outcomes.

It is appropriate for modeling a binary result such as success / fail or true / false, which is typical in binary classification problems.

**PARENT** Family.FamilyInterface <Family.ecl.tex>

---

## **QUASIBINOMIAL** Quasibinomial

[Family](#) \

	Quasibinomial
--	---------------

The Quasibinomial Regression Family is similar to the Binomial family (see Binomial above) except that it adjusts for situations where the variance of the distribution is greater or less than anticipated by the Binomial Distribution. This is known as over-dispersion or under-dispersion.

The results are adjusted based on the dispersion of the data to better model the observations in these situations.

---

**PARENT** [Family.FamilyInterface](#) <Family.ecl.tex>

---

## **POISSON** Poisson

[Family](#) \

	Poisson
--	---------

Poisson Regression is generally used to model count data, where the dependent variable is a positive (or zero) integer.

It is also known as a log-linear model in that the logarithm of the dependent variables is assumed to be linear.

---

**PARENT** [Family.FamilyInterface](#) <Family.ecl.tex>

---

## **QUASIPOISSON** Quasipoisson

[Family](#) \

## Quasipoisson

Quasipoisson Regression is similar to Poisson Regression (see Poisson above) except that it adjusts for situations where the variance of the distribution is greater or less than anticipated by the Poisson Distribution. This is known as over-dispersion or under-dispersion.

The results are adjusted based on the dispersion of the data to better model the observations in these situations.

**PARENT** Family.FamilyInterface <Family.ecl.tex>

---

## GAMMA Gamma

Family \

### Gamma

Gamma Regression is used to model continuous, non-negative, data with a right-skew. Such data exhibits heteroskedacity, (i.e. inconsistent variance across the range). The gamma regression assumes that the variance is near constant on a log scale. Various types of financial and insurance data often have these characteristics.

**PARENT** Family.FamilyInterface <Family.ecl.tex>

---

## GAUSSIAN Gaussian

Family \

### Gaussian

Gaussian Regression is equivalent to Ordinary Least Squares (OLS) regression. It assumes that the error term is Normally distributed.

**PARENT** Family.FamilyInterface <Family.ecl.tex>

---

## **INVGAUSS InvGauss**

[Family](#) \

	<b>InvGauss</b>
--	-----------------

Inverse Gaussian Regression aka Wald Regression is similar to the Gamma Regression in that it is used to model continuous, positive heteroskedastic data. It differs from the Gamma Regression assumptions in that it has a wider tail (i.e. more frequent occurrence of higher numbers). The variance is assumed to be proportional to the cube of the mean.

**PARENT** [Family.FamilyInterface <Family.ecl.tex>](#)

---

# GLM

---

[Go Up](#)

## IMPORTS

```
Constants | irls | Family | __versions.ML_Core.V3_2_2.ML_Core.Interfaces |
__versions.ML_Core.V3_2_2.ML_Core.Types |
```

## DESCRIPTIONS

### **GLM** **GLM**

GLM
( DATASET(NumericField) X = DATASET([], NumericField), DATASET(NumericField) Y = DATASET([], NumericField), Family.FamilyInterface fam = Family.Gaussian, DATASET(NumericField) weights = DATASET([], NumericField), UNSIGNED max_iter = 200, REAL8 epsilon = Constants.default_epsilon, REAL8 ridge = Constants.default_ridge)

Main GLM regression module. Performs regressions using iteratively re-weighted least squares (IRLS).

**PARAMETER** **X** ||| TABLE ( NumericField ) — The observed explanatory values in NumericField format.

**PARAMETER** **Y** ||| TABLE ( NumericField ) — The observed values the model aims to fit in NumericField format.

**PARAMETER** **fam** ||| INTERFACE ( FamilyInterface ) — (Optional) A module defining the type of regression to perform. Default = Gaussian (i.e. ordinary least squares).

**PARAMETER** **weights** ||| TABLE ( NumericField ) — (Optional) A set of observation weights (one per dependent value), in NumericField format. Default = equal weights.

**PARAMETER** **max\_iter** ||| UNSIGNED8 — (Optional) Maximum number of iterations to try. Default = 200.

**PARAMETER** `epsilon` ||| REAL8 — (Optional) The minimum change in the Beta value estimate to continue.

**PARAMETER** `ridge` ||| REAL8 — (Optional) A value to populate a diagonal matrix that is added to a matrix help assure that the matrix is invertible.

**SEE** ML\_Core.Types.NumericField

**PARENT** \_\_versions.ML\_Core.V3\_2\_2.ML\_Core.Interfaces.IRegression

</home/lily/.HPCCSystems/bundles/\_\_versions/ML\_Core/V3\_2\_2/ML\_Core/Interfaces/IRegression.ecl

## Children

1. [GetModel](#) : Calculate a model to fit the observation data to the observed values
  2. [Predict](#) : Predict the observations using models trained by the GetModel function
- 

## GETMODEL GetModel

GLM \

<code>DATASET(Types.Layout_Model)</code>	<code>GetModel</code>
--	-----------------------

Calculate a model to fit the observation data to the observed values.

**RETURN** — The encoded model in Layout\_Model format.

**SEE** ML\_Core.Types.Layout\_Model

## OVERRIDE

---

## PREDICT Predict

GLM \

<b>PARAMETER</b>	<b>newX</b>     TABLE ( NumericField ) — Observations to be predicted.
<b>PARAMETER</b>	<b>model</b>     TABLE ( Layout_Model ) — The model as returned from GetModel.

Predict the observations using models trained by the GetModel function.

**PARAMETER** **newX** ||| TABLE ( NumericField ) — Observations to be predicted.

**PARAMETER** **model** ||| TABLE ( Layout\_Model ) — The model as returned from GetModel.

**RETURN** TABLE ( { UNSIGNED2 wi , UNSIGNED8 id , UNSIGNED4 number , REAL8 value } ) — Predictions in NumericField format.

**SEE** ML\_Core.Tyeps.NumericField

**OVERRIDE**

---

# LogitPredict

---

[Go Up](#)

## IMPORTS

Types | Family | \_\_versions.ML\_Core.V3\_2\_2.ML\_Core.Types |

## DESCRIPTIONS

### **LOGITPREDICT** LogitPredict

<b>PARAMETER</b> <code>coef</code>     TABLE ( Model_Coef ) — the model beta coefficients as returned from ExtractBeta.	<code>LogitPredict</code>
<code>(DATASET(Model_Coef) coef, DATASET(NumericField) independents)</code>	

Predict the category values with the logit function and the supplied beta coefficients.

**PARAMETER** `coef` ||| TABLE ( Model\_Coef ) — the model beta coefficients as returned from ExtractBeta.

**PARAMETER** `independents` ||| TABLE ( NumericField ) — the observations.

**RETURN** TABLE ( { UNSIGNED2 wi , UNSIGNED8 id , UNSIGNED4 number , INTEGER4 value , REAL8 conf } ) — the predicted category values and a confidence score in Classify\_Result format.

**SEE** ExtractBeta

**SEE** ML\_Core.Types.Classify\_Result

# LUCI\_Model

---

[Go Up](#)

## IMPORTS

Types | IRLS | Family | std.Str | std.system.ThorLib |

## DESCRIPTIONS

### **LUCI\_MODEL LUCI\_Model**

DATASET(	Types.LUCI_Rec)	LUCI_Model
	(DATASET(	Types.LUCI_Model_Rqst) rqst,
	DATASET(	Types.External_Model) mod, STRING
	wi_field='work_item', Family.FamilyInterface fam =	
	Family.Gaussian)	

Create a LUCI model file description of the model(s) from the external version of the model.

LUCI is a proprietary format used within LexisNexis.

The multi-score card per model case assumes that the score card selection is based solely upon the work item field. If this is not the case, the L1SE records will need to be patched.

The model id and name may have a "\$" character that is updated to match the work item when there are multiple models applied. If the strings do not have a "\$" character, the work item string is appended.

The score card name may have a "\$" character which is updated to match the work item. If the name is blank, the score card is named for the work item.

LUCI data fields may not contain comma characters. This function requires that the work item identification strings do not contain characters that need special handling for CSV data.

**PARAMETER** rqst ||| TABLE ( LUCI\_Model\_Rqst ) — the information to map work items to models in LUCI\_Model\_Rqst format.

**PARAMETER** mod ||| TABLE ( External\_Model ) — the model with the external field names applied in External\_Model format as returned from Named\_Model.

**PARAMETER** wi\_field ||| STRING — the field name holding the work item identification string.

**PARAMETER** fam ||| INTERFACE ( FamilyInterface ) — the family module for the distribution family on which the regression is based.

**RETURN** TABLE ( { STRING line } ) — The lines of the LUCI file in LUCI\_Rec format.

**SEE** Family

**SEE** Types.External\_Model

**SEE** Named\_Model

**SEE** Types.LUCI\_Model\_Rqst

**SEE** Types.LUCI\_Rec

---

# Model\_Deviance

---

[Go Up](#)

## IMPORTS

Types |

## DESCRIPTIONS

### **MODEL\_DEVIANCE** Model\_Deviance

/ EXPORT DATASET(Types.Deviance_Record)	Model_Deviance
(DATASET(Types.Observation_Deviance) od, DATASET(Types.Model_Coef) mod)	

Model Deviance Report.

Create a report of deviance information for a model.

Analysis of Deviance is analogous to the Analysis of Variance (ANOVA) used in least-squares modeling, but adapted to the general linear model (GLM). In this case it is adapted specifically to the logistic model.

**PARAMETER** **od** ||| TABLE ( Observation\_Deviance ) — observation-deviance records, as obtained from a call to Deviance\_Detail.

**PARAMETER** **mod** ||| TABLE ( Model\_Coef ) — model co-efficients records, as obtained from a call to ExtractBeta.

**RETURN** TABLE ( { UNSIGNED2 wi , UNSIGNED4 model , REAL8 df , REAL8 deviance , REAL8 AIC } ) — model deviance in Deviance\_Record format.

**SEE** Deviance\_Detail

**SEE** ExtractBeta

**SEE** Types.Deviance\_Record

---

# Named\_Model

---

[Go Up](#)

## IMPORTS

Types |

## DESCRIPTIONS

### **NAMED\_MODEL Named\_Model**

/ EXPORT DATASET(Types.External_Model)	Named_Model
(DATASET(Types.Layout_Model) mod_ds, DATASET(Types.FieldName_Mapping) expl_map, DATASET(Types.FieldName_Mapping) resp_map, DATASET(Types.WorkItem_mapping) wi_map=empty, REAL8 level=0.05)	

Apply external labels for work items and field names to a model.

Returns an expanded model that includes:

- coefficients
- z and p-values
- independent variable field names
- dependent variable field names
- work-item names

**PARAMETER** mod\_ds ||| TABLE ( Layout\_Model ) — the model as returned from GetModel.

**PARAMETER** expl\_map ||| TABLE ( FieldName\_Mapping ) — the relation of the explanatory or independent variables to the field names for those variables in FieldName\_Mapping format.

**PARAMETER** resp\_map ||| TABLE ( FieldName\_Mapping ) — the relation of the response variable column numbers to the field names in FieldName\_Mapping format.

**PARAMETER** wi\_map ||| TABLE ( WorkItem\_Mapping ) — (optional) mapping of workitem strings to workitem nominals in FieldName\_Mapping format.

**PARAMETER** level ||| REAL8 — (optional) value for confidence intervals. Default = 0.05.

**RETURN** TABLE ( { STRING work\_item , STRING response\_field , UNSIGNED2 wi , UNSIGNED4 dep\_nom , TABLE ( External\_Coef ) coef } ) — an expanded model in External\_Model format.

**SEE** Types.FieldName\_Mapping

**SEE** Types.External\_Model

---

# Null\_Deviance

---

[Go Up](#)

## IMPORTS

Types |

## DESCRIPTIONS

### **NULL\_DEVIANCE** Null\_Deviance

<b>DATASET</b> (Types.Deviance_Record)	Null_Deviance
(DATASET(Types.Observation_Deviance) od)	

Return Deviance information for the null model, that is, a model with only an intercept.

Analysis of Deviance is analogous to the Analysis of Variance (ANOVA) used in least-squares modeling, but adapted to the general linear model (GLM). In this case it is adapted specifically to the logistic model.

**PARAMETER** **od** ||| TABLE ( Observation\_Deviance ) — Observation Deviance record set as returned from Deviance\_Detail.

**RETURN** TABLE ( { UNSIGNED2 wi , UNSIGNED4 model , REAL8 df , REAL8 deviance , REAL8 AIC } ) — a data set of the null model deviances for each work item and classifier in Deviance\_Record format.

**SEE** Types.Observation\_Deviance

**SEE** Types.Deviance\_Record

**SEE** Deviance\_Detail



# Predict

---

[Go Up](#)

## IMPORTS

Types | IRLS | Family | \_\_versions.ML\_Core.V3\_2\_2.ML\_Core.Types |

## DESCRIPTIONS

### **PREDICT** Predict

<b>PARAMETER</b> <code>coef</code>     TABLE ( Model_Coef ) — the model beta coefficients.	<b>PARAMETER</b> <code>independents</code>     TABLE ( NumericField ) — the observations.
<code>(DATASET(Model_Coef) coef, DATASET(NumericField) independents, Family.FamilyInterface fam)</code>	

Calculate the score using the appropriate mean function and the the supplied beta coefficients.

**PARAMETER** `coef` ||| TABLE ( Model\_Coef ) — the model beta coefficients.

**PARAMETER** `independents` ||| TABLE ( NumericField ) — the observations.

**PARAMETER** `fam` ||| INTERFACE ( FamilyInterface ) — module defining the error distribution and link of the dependents.

**RETURN** TABLE ( { UNSIGNED2 wi , UNSIGNED8 id , UNSIGNED4 number , REAL8 value } ) — the prediction value.

---

# Types

---

[Go Up](#)

## IMPORTS

\_versions.ML\_Core.V3\_2\_2.ML\_Core.Types |

## DESCRIPTIONS

### **TYPES** Types

	Types
--	-------

Type definitions for GLM bundle

### Children

1. [AnyField](#) : No Documentation Found
2. [NumericField](#) : The NumericField layout defines a matrix of Real valued data-points
3. [DiscreteField](#) : The Discrete Field layout defines a matrix of Integer valued data-points
4. [Layout\\_Model](#) : No Documentation Found
5. [t\\_work\\_item](#) : No Documentation Found
6. [t\\_RecordID](#) : No Documentation Found
7. [t\\_FieldNumber](#) : No Documentation Found
8. [t\\_FieldReal](#) : No Documentation Found
9. [t\\_Discrete](#) : No Documentation Found
10. [t\\_Universe](#) : No Documentation Found

11. [Field\\_Desc](#) : Describe information about each field in a training set
  12. [Data\\_Info](#) : Describes information about a training dataset composed of independent and dependent columns
  13. [Data\\_Diagnostic](#) : Describes any errors in the data
  14. [NumericField\\_U](#) : Record structure to add a 'Universe Number' to a NumericField allowing multiple independent NumericField matrixes within a work-item
  15. [DiscreteField\\_U](#) : Record structure to add a 'Universe Number' to a DiscreteField allowing multiple independent DiscreteField matrixes within a work-item
  16. [Layout\\_Column\\_Map](#) : Layout for a column map record that is used to remap column numbers
  17. [Regressor\\_Stats](#) : Summary information about a regressor
  18. [Model\\_Report](#) : Statistical information about a model
  19. [Binomial\\_Confusion\\_Summary](#) : Accuracy stats for binomial classifications
  20. [Model\\_Coef](#) : Model Coefficients
  21. [Confidence\\_Model\\_Coef](#) : Model Coefficients with confidence intervals
  22. [pval\\_Model\\_Coef](#) : Model coefficients with z and p-value
  23. [Full\\_Model\\_Coef](#) : Model coefficients with confidence intervals and p-value
  24. [External\\_Coef](#) : Model coefficients, confidence intervals, and p-value, plus independent field names, for each coefficient
  25. [External\\_Model](#) : Expanded version of a model with statistics and field names
  26. [Observation\\_Deviance](#) : Record to contain deviance information about each observation
  27. [Deviance\\_Record](#) : Record to hold deviance summary information about a model
  28. [AOD\\_Record](#) : Record to hold Analysis of Deviance (AOD) information for a model
  29. [FieldName\\_Mapping](#) : Layout used to hold the mapping between a field's number and its name
  30. [WorkItem\\_Mapping](#) : Layout used to hold the mapping between a work-item number and a textual name for that work-item
  31. [LUCI\\_Rec](#) : Layout to store the lines of a generated LUCI model file
  32. [LUCI\\_Model\\_Rqst](#) : Format for information to guide the generation of a LUCI file
-

## **ANYFIELD** AnyField

[Types](#) \

	AnyField
--	----------

No Documentation Found

---

## **NUMERICFIELD** NumericField

[Types](#) \

	NumericField
--	--------------

The NumericField layout defines a matrix of Real valued data-points. It acts as the primary Dataset layout for interacting with most ML Functions. Each record represents a single cell in a matrix. It is most often used to represent a set of data-samples or observations, with the 'id' field representing the data-sample or observation, and the 'number' field representing the various fields within the observation.

**FIELD** wi ||| — The work-item id, supporting the Myriad style interface. This allows multiple independent matrixes to be contained within a single dataset, supporting independent ML activities to be processed in parallel.

**FIELD** id ||| — This field represents the row-number of this cell of the matrix. It is also considered the record-id for observations / data-samples.

**FIELD** number ||| — This field represents the matrix column number for this cell. It is also considered the field number of the observation

**FIELD** value ||| — The value of this cell in the matrix.

---

## **DISCRETEFIELD** DiscreteField

[Types](#) \

	DiscreteField
--	---------------

The Discrete Field layout defines a matrix of Integer valued data-points. It is similar to the NumericField layout above, except for only containing discrete (integer) values. It is typically used to convey the class-labels for classification algorithms.

**FIELD** wi ||| — The work-item id, supporting the Myriad style interface. This allows multiple independent matrixes to be contained within a single dataset, supporting independent ML activities to be processed in parallel.

**FIELD** id ||| — This field represents the row-number of this cell of the matrix. It is also considered the record-id for observations / data-samples.

**FIELD** number ||| — This field represents the matrix column number for this cell. It is also considered the field number of the observation

**FIELD** value ||| — The value of this cell in the matrix.

---

## LAYOUT\_MODEL Layout\_Model

Types \

	Layout_Model
--	--------------

No Documentation Found

---

## T\_WORK\_ITEM t\_work\_item

Types \

	t_work_item
--	-------------

No Documentation Found

---

**RETURN** UNSIGNED2 —

---

## **T\_RECORDID** t\_RecordID

[Types](#) \

	t_RecordID
--	------------

No Documentation Found

---

**RETURN** UNSIGNED8 —

---

## **T\_FIELDNUMBER** t\_FieldNumber

[Types](#) \

	t_FieldNumber
--	---------------

No Documentation Found

---

**RETURN** UNSIGNED4 —

---

## **T\_FIELDREAL** t\_FieldReal

[Types](#) \

	t_FieldReal
--	-------------

No Documentation Found

---

**RETURN** REAL8 —

---

## T\_DISCRETE t\_Discrete

Types \

t_Discrete
------------

No Documentation Found

---

**RETURN** INTEGER4 —

---

## T\_UNIVERSE t\_Universe

Types \

t_Universe
------------

No Documentation Found

---

**RETURN** UNSIGNED1 —

---

## FIELD\_DESC Field\_Desc

Types \

Field_Desc
------------

Describe information about each field in a training set.

**FIELD** number ||| UNSIGNED4 — the column (feature) number.

**FIELD** cardinality ||| UNSIGNED4 — the number of unique values in the field.

**FIELD** min\_value ||| REAL8 — the minimum value for the field.

**FIELD** max\_value ||| REAL8 — the maximum value for the field.

**FIELD** is\_integer ||| BOOLEAN — No Doc

---

## **DATA\_INFO** Data\_Info

Types \

	Data_Info
--	-----------

Describes information about a training dataset composed of independent and dependent columns.

**FIELD** wi ||| UNSIGNED2 — the work-item number.

**FIELD** dependent\_fields ||| UNSIGNED4 — the number of fields in the dependent data.

**FIELD** dependent\_records ||| UNSIGNED4 — the number of records in the dependent data.

**FIELD** independent\_fields ||| UNSIGNED4 — the number of fields in the independent data.

**FIELD** independent\_records ||| UNSIGNED4 — the number of records in the independent data.

**FIELD** dependent\_stats ||| TABLE ( Field\_Desc ) — dataset of Field\_Desc records describing each of the fields of the dependent data.

**FIELD** independent\_stats ||| TABLE ( Field\_Desc ) — dataset of Field\_Desc records describing each of the fields of the independent data.

**FIELD** dependent\_count ||| UNSIGNED4 — No Doc

**FIELD** independent\_count ||| UNSIGNED4 — No Doc

**SEE** Field\_Desc

---

## **DATA\_DIAGNOSTIC** Data\_Diagnostic

Types \

## Data\_Diagnostic

Describes any errors in the data.

**FIELD** wi ||| UNSIGNED2 — The work-item number.

**FIELD** valid ||| BOOLEAN — Boolean TRUE indicates that the data is valid, FALSE indicates problems with the data.

**FIELD** message\_text ||| SET ( VARSTRING ) — A textual description of any errors in the data.

---

## NUMERICFIELD\_U NumericField\_U

Types \

### NumericField\_U

Record structure to add a 'Universe Number' to a NumericField allowing multiple independent NumericField matrixes within a work-item.

**FIELD** u ||| UNSIGNED1 — the 'universe' number identifying a distinct matrix within a NumericField dataset and work-item.

**FIELD** wi ||| UNSIGNED2 — No Doc

**FIELD** id ||| UNSIGNED8 — No Doc

**FIELD** number ||| UNSIGNED4 — No Doc

**FIELD** value ||| REAL8 — No Doc

---

## DISCRETEFIELD\_U DiscreteField\_U

Types \

### DiscreteField\_U

Record structure to add a 'Universe Number' to a DiscreteField allowing multiple independent DiscreteField matrixes within a work-item.

**FIELD** [u](#) ||| UNSIGNED1 — the 'universe' number identifying a distinct matrix within a DiscreteField dataset and work-item.

**FIELD** [wi](#) ||| UNSIGNED2 — No Doc

**FIELD** [id](#) ||| UNSIGNED8 — No Doc

**FIELD** [number](#) ||| UNSIGNED4 — No Doc

**FIELD** [value](#) ||| INTEGER4 — No Doc

---

## LAYOUT\_COLUMN\_MAP Layout\_Column\_Map

Types \

	<u><a href="#">Layout_Column_Map</a></u>
--	--

Layout for a column map record that is used to remap column numbers.

**FIELD** [wi](#) ||| UNSIGNED2 — the work-item number.

**FIELD** [orig\\_number](#) ||| UNSIGNED4 — the original field number.

**FIELD** [remap\\_number](#) ||| UNSIGNED4 — the mapped-to field number.

---

## REGRESSOR\_STATS Regressor\_Stats

Types \

	<u><a href="#">Regresso_Stats</a></u>
--	---------------------------------------

Summary information about a regressor.

**FIELD** [column](#) ||| UNSIGNED4 — the regressor field number.

**FIELD** [max\\_delta](#) ||| REAL8 — the max\_delta value for the regressor.

**FIELD** [iterations](#) ||| UNSIGNED4 — the number of iterations used to train the regressor.

**FIELD** mse ||| REAL8 — the mean square error of the regressor.

**FIELD** dispersion ||| REAL8 — the dispersion of the regressor.

---

## MODEL\_REPORT Model\_Report

Types \

Model_Report
--------------

Statistical information about a model.

One record is generated per work-item.

**FIELD** wi ||| UNSIGNED2 — the work-item

**FIELD** max\_iterations ||| UNSIGNED4 — the maximum iterations used to train the model.

**FIELD** epsilon ||| REAL8 — the 'epsilon' value used within the model.

**FIELD** dep\_vars ||| UNSIGNED4 — the number of dependent variables (i.e. classifiers).

**FIELD** ind\_vars ||| UNSIGNED4 — the number of independent variables (i.e. features).

**FIELD** obs ||| UNSIGNED8 — the number of observations (i.e. records) in the training data.

**FIELD** builder ||| UNSIGNED2 — the identifier for the builder used to train the model.

**FIELD** stats ||| TABLE ( Regressor\_Stats ) — child dataset of Regressor\_Stats, one for each regressor in the work-item.

**SEE** Regressor\_Stats

---

## BINOMIAL\_CONFUSION\_SUMMARY Binomial\_Confusion\_Summary

Types \

Binomial_Confusion_Summary
----------------------------

Accuracy stats for binomial classifications.

One record per work-item and classifier.

**FIELD** wi ||| UNSIGNED2 — the work-item number.

**FIELD** classifier ||| UNSIGNED4 — the classifier field number (i.e. dependent field number).

**FIELD** true\_positive ||| UNSIGNED8 — the count of true positive results (i.e. predicted = TRUE, actual = TRUE).

**FIELD** true\_negative ||| UNSIGNED8 — the count of true negative results (i.e. predicted = FALSE, actual = FALSE).

**FIELD** false\_positive ||| UNSIGNED8 — the count of false\_positive results (i.e. predicted = TRUE, actual = FALSE).

**FIELD** false\_negative ||| UNSIGNED8 — the count of false\_negative results (i.e. predicted = FALSE, actual = TRUE).

**FIELD** cond\_pos ||| UNSIGNED8 — the count of results where actual = TRUE.

**FIELD** pred\_pos ||| UNSIGNED8 — the count of results where predicted = TRUE.

**FIELD** cond\_neg ||| UNSIGNED8 — the count of results where actual = FALSE.

**FIELD** pred\_neg ||| UNSIGNED8 — the count of results where predicted = FALSE.

**FIELD** prevalence ||| REAL8 — cond\_pos / total.

**FIELD** accuracy ||| REAL8 — (true\_positive + true\_negative) / total.

**FIELD** true\_pos\_rate ||| REAL8 — true\_positive / cond\_pos.

**FIELD** false\_pos\_rate ||| REAL8 — false\_positive / cond\_neg.

**FIELD** true\_neg\_rate ||| REAL8 — true\_negative / cond\_neg.

**FIELD** pos\_pred\_val ||| REAL8 — true\_positive / pred\_pos.

**FIELD** false\_disc\_rate ||| REAL8 — false\_positive / pred\_pos.

**FIELD** false omit rate ||| REAL8 — false\_negative / pred\_neg.

**FIELD** neg\_pred\_val ||| REAL8 — true\_negative / pred\_neg.

**FIELD** false\_neg\_rate ||| REAL8 — No Doc

## **MODEL\_COEF** Model\_Coef

Types \

Model_Coef
------------

Model Coefficients.

**FIELD** wi ||| UNSIGNED2 — the work-item number.

**FIELD** ind\_col ||| UNSIGNED4 — the independent column number (i.e feature number).

**FIELD** dep\_nom ||| UNSIGNED4 — the dependent column number (i.e. classifier number).

**FIELD** w ||| REAL8 — the learned weight (i.e. coefficient).

**FIELD** SE ||| REAL8 — the Standard Error of the coefficient.

---

## **CONFIDENCE\_MODEL\_COEF** Confidence\_Model\_Coef

Types \

Confidence_Model_Coef
-----------------------

Model Coefficients with confidence intervals.

**FIELD** upper ||| REAL8 — the upper range of the confidence interval

**FIELD** lower ||| REAL8 — the lower range of the confidence interval

**FIELD** wi ||| UNSIGNED2 — No Doc

**FIELD** ind\_col ||| UNSIGNED4 — No Doc

**FIELD** dep\_nom ||| UNSIGNED4 — No Doc

**FIELD** w ||| REAL8 — No Doc

**FIELD** se ||| REAL8 — No Doc

---

## PVAL\_MODEL\_COEF pval\_Model\_Coef

Types \

pval_Model_Coef
-----------------

Model coefficients with z and p-value.

**FIELD** z ||| REAL8 — the z value.

**FIELD** p\_value ||| REAL8 — the p\_value of the coefficient.

**FIELD** wi ||| UNSIGNED2 — No Doc

**FIELD** ind\_col ||| UNSIGNED4 — No Doc

**FIELD** dep\_nom ||| UNSIGNED4 — No Doc

**FIELD** w ||| REAL8 — No Doc

**FIELD** se ||| REAL8 — No Doc

---

## FULL\_MODEL\_COEF Full\_Model\_Coef

Types \

Full_Model_Coef
-----------------

Model coefficients with confidence intervals and p-value

**FIELD** z ||| REAL8 — the z value.

**FIELD** p\_value ||| REAL8 — the p\_value of the coefficient.

**FIELD** upper ||| REAL8 — the upper range of the confidence interval

**FIELD** lower ||| REAL8 — the lower range of the confidence interval

**FIELD** wi ||| UNSIGNED2 — No Doc

**FIELD** ind\_col ||| UNSIGNED4 — No Doc

**FIELD** dep\_nom ||| UNSIGNED4 — No Doc

**FIELD** w ||| REAL8 — No Doc

**FIELD** se ||| REAL8 — No Doc

---

## **EXTERNAL\_COEF** External\_Coef

Types \

External_Coef
---------------

Model coefficients, confidence intervals, and p-value, plus independent field names, for each coefficient.

**FIELD** isIntercept ||| BOOLEAN — Boolean field is TRUE if this is the intercept coefficient, otherwise FALSE.

**FIELD** field\_name ||| STRING — the name of the independent field for this coefficient.

**FIELD** w ||| REAL8 — the coefficient value (weight)

**FIELD** SE ||| REAL8 — the Standard Error of the coefficient

**FIELD** z ||| REAL8 — the z value.

**FIELD** p\_value ||| REAL8 — the p-value.

**FIELD** upper ||| REAL8 — the upper bound of the confidence interval.

**FIELD** lower ||| REAL8 — the lower bound of the confidence interval.

**FIELD** ind\_col ||| UNSIGNED4 — the field number of the independent field for this coefficient.

---

## **EXTERNAL\_MODEL** External\_Model

Types \

External_Model
----------------

Expanded version of a model with statistics and field names.

Field names include independent data field names, dependent data field names and work-item names.

- FIELD** work\_item ||| STRING — the work-item's name.
- FIELD** response\_field ||| STRING — the name of the classifier field (i.e. dependent field name).
- FIELD** wi ||| UNSIGNED2 — the work-item number.
- FIELD** dep\_nom ||| UNSIGNED4 — the field number of the classifier (i.e. dependent field number).
- FIELD** coef ||| TABLE ( External\_Coef ) — child dataset of External\_Coef format. One record per model coefficient.

**SEE** External\_Coef

---

## OBSERVATION\_DEVIANCE Observation\_Deviance

Types \

	Observation_Deviance
--	----------------------

Record to contain deviance information about each observation.

- FIELD** wi ||| UNSIGNED2 — the work-item number.
- FIELD** id ||| UNSIGNED8 — the record id (i.e. observation number).
- FIELD** classifier ||| — the dependent field number.
- FIELD** actual ||| REAL8 — the actual (i.e. ground truth value).
- FIELD** predicted ||| REAL8 — the value predicted by the model.
- FIELD** mod\_ll ||| REAL8 — log likelihood of the model
- FIELD** mod\_dev\_component ||| REAL8 — the deviance explained by the model
- FIELD** mod\_dev\_residual ||| REAL8 — the deviance not explained by the model (i.e. the residual)
- FIELD** nil ||| — ll log likelihood of the nil model (i.e. model with only a constant term).
- FIELD** nil\_dev\_component ||| REAL8 — the deviance explained by the null model
- FIELD** nil\_dev\_residual ||| REAL8 — the deviance not explained by the null model (i.e. the residual)
- FIELD** model ||| UNSIGNED4 — No Doc
- FIELD** nil\_ll ||| REAL8 — No Doc

## **DEVIANCE\_RECORD** Deviance\_Record

Types \

### Deviance\_Record

Record to hold deviance summary information about a model.

**FIELD** wi ||| UNSIGNED2 — the work-item number

**FIELD** classifier ||| — the classifier number (i.e. field number of the dependent variable).

**FIELD** df ||| REAL8 — degrees-of-freedom of the chi squared distribution.

**FIELD** deviance ||| REAL8 — the total deviance for this classifier.

**FIELD** AIC ||| REAL8 — the Akaike Information Criteria value.

**FIELD** model ||| UNSIGNED4 — No Doc

---

## **AOD\_RECORD** AOD\_Record

Types \

### AOD\_Record

Record to hold Analysis of Deviance (AOD) information for a model.

**FIELD** wi ||| UNSIGNED2 — the work-item number

**FIELD** classifier ||| — the classifier number (i.e. field number of the dependent variable).

**FIELD** df ||| REAL8 — degrees of freedom of the chi squared distribution.

**FIELD** residual\_dev ||| REAL8 — the deviance not explained by the model.

**FIELD** deviance ||| REAL8 — the total deviance.

**FIELD** p ||| — value the probability that the null hypothesis is correct.

**FIELD** model ||| UNSIGNED4 — No Doc

**FIELD** residual\_df ||| UNSIGNED8 — No Doc

**FIELD** p\_value ||| REAL8 — No Doc

---

## **FIELDNAME\_MAPPING** FieldName\_Mapping

Types \

	FieldName_Mapping
--	-------------------

Layout used to hold the mapping between a field's number and its name.

**FIELD** orig\_name ||| STRING — typically the field number as a text string (e.g. '2').

**FIELD** assigned\_name ||| STRING — the textual name of the field (e.g. 'age').

---

## **WORKITEM\_MAPPING** WorkItem\_Mapping

Types \

	WorkItem_Mapping
--	------------------

Layout used to hold the mapping between a work-item number and a textual name for that work-item.

**FIELD** wi ||| UNSIGNED2 — the work-item number.

**FIELD** orig\_wi ||| STRING — the work-item name.

---

## **LUCI\_REC** LUCI\_Rec

Types \

	LUCI_Rec
--	----------

Layout to store the lines of a generated LUCI model file.

**FIELD** line ||| STRING — the text for a single line for the LUCI file.

---

## LUCI\_MODEL\_RQST LUCI\_Model\_Rqst

[Types](#) \

	LUCI_Model_Rqst
--	-----------------

Format for information to guide the generation of a LUCI file.

**FIELD** model\_id ||| STRING — a short textual name for the model as used in the LUCI L1MD format.

**FIELD** model\_name ||| STRING — an expanded name for the model as used in the LUCI L1MD format.

**FIELD** response\_field ||| STRING — name of the dependent field (aka classifier name).

**FIELD** wi\_list ||| SET ( STRING ) — can be set to ['ALL'], or can be a list of work-item names.

**FIELD** score\_card\_name ||| STRING — the score card name pattern (see LUCI\_Model.ecl for details).

---