GENLINMIXED

Syntax Chart

GENLINMIXED is available in the Advanced Models option.

```
GENLINMIXED
  [/DATA_STRUCTURE]
     SUBJECTS = fieldName[*fieldName[*...]]
     [REPEATED_MEASURES = fieldName[*fieldName[*...]]]
     [GROUPING = fieldName[*fieldName[*...]]]
     [COVARIANCE_TYPE = DIAGONAL** | AR1 | ARMA11 | COMPOUND_SYMMETRY | IDENTITY
                        TOEPLITZ | UNSTRUCTURED | VARIANCE_COMPONENTS]
  /FIELDS
   TARGET = fieldName
   [TRIALS = {NONE**          }]
            {VALUE(number)    }
            {FIELD(fieldName)}
   [ANALYSIS_WEIGHT = fieldName]
   [OFFSET = {NONE**          }]
            {VALUE(number)    }
            {FIELD(fieldName)}
  [/TARGET_OPTIONS]
   [REFERENCE = value]
   [DISTRIBUTION = NORMAL | BINOMIAL | MULTINOMIAL | GAMMA |
                  INVERSE_GAUSSIAN | NEGATIVE_BINOMIAL | POISSON]
   [LINK = IDENTITY | CAUCHIT | CLOGLOG | LOG | LOGC | LOGIT | NLOGLOG | POWER | PROBIT ]
   [LINK_PARAMETER = number]
  [/FIXED]
   [EFFECTS = effect [effect ...]]
   [USE_INTERCEPT = TRUE** | FALSE]
  [/RANDOM]
   [EFFECTS = effect [effect ...]]
   [USE_INTERCEPT = TRUE | FALSE**]
   [SUBJECTS = fieldName[*fieldName[*...]]]
   [GROUPING = fieldName[*fieldName[*...]]]
   [COVARIANCE_TYPE = DIAGONAL | AR1 | ARMA11 | COMPOUND_SYMMETRY | IDENTITY
                      TOEPLITZ | UNSTRUCTURED | VARIANCE_COMPONENTS**]
  [/RANDOM ...]
  [/BUILD_OPTIONS]
   [TARGET_CATEGORY_ORDER = {ASCENDING**}]
                           {DESCENDING }
                           {DATA       }
   [INPUTS_CATEGORY_ORDER = {ASCENDING**}]
                           {DESCENDING }
                           {DATA       }
   [MAX_ITERATIONS = {100** }]
                     {number}
   [CONFIDENCE_LEVEL = {95**   }]
                       {number}
   [DF_METHOD = {RESIDUAL**    }]
               {SATTERTHWAITE}
   [COVB = {MODEL**}]
          {ROBUST }
  [/EMMEANS
   [TABLES = factor[*factor...]]
   [COMPARE = factor]
   [CONTRAST = {NONE**    }]
              {PAIRWISE }
              {DEVIATION}
              {SIMPLE   }
  [/EMMEANS...]
  [/EMMEANS_OPTIONS]
   [CONTROL = fieldName (number) [fieldName (number) ...]]
   [SCALE = {ORIGINAL** }]
           {TRANSFORMED}
   [PADJUST = {LSD**        }]
             {SEQBONFERRONI}
             {SEQSIDAK     }
  [/OUTFILE MODEL = 'file']
  [/SAVE [PREDICTED_VALUES(fieldName)]]
   [PREDICTED_PROBABILITY(rootname)]
   [MAX_CATEGORIES({25**   })]
                   {integer}
```

```
[CONFIDENCE(fieldName)]
[CONFIDENCE_METHOD({PREDICTEDPROBABILITY**})]
                    {PROBABILITYIMPROVEMENT}
[CONFIDENCE_INTERVALS(rootname)]
[PEARSON_RESIDUALS(fieldName)].
```

** Default if the subcommand or keyword is omitted.

This command reads the active dataset and causes execution of any pending commands.

Release History

Release 19

■ Command introduced.

Release 20

■ Ordinal targets can be analyzed using the Multinomial distribution and the complementary log-log, cauchit, logit, negative log-log, or probit link functions.

Example

```
GENLINMIXED
  /FIELDS TARGET=y1.
```

Overview

The GENLINMIXED procedure fits generalized linear mixed models. Generalized linear mixed models extend the linear model so that:

■ The target is linearly related to the factors and covariates via a specified link function

■ The target can have a non-normal distribution

■ The observations can be correlated.

Generalized linear mixed models cover a wide variety of models, from simple linear regression to complex multilevel models for non-normal longitudinal data.

Options

Probability Distribution of Target. The probability distribution of the target can be specified as normal, binomial, gamma, inverse Gaussian, multinomial, negative binomial, or Poisson.

Events/Trials Target Specification. The typical target specification will be a single field, but for the binomial distribution the target can be specified using a number-of-events field and a number-of-trials field. Alternatively, if the number of trials is the same across all subjects, then trials can be specified using a fixed number instead of a field.

Link Function. The target is linearly related to the model effects via one of the following link functions: Identity, complementary log-log, log, log-complement, logit, negative log-log, power, and probit.

Repeated Measurements. You can specify the subject structure for repeated measurements and how the errors of the repeated measurements are correlated. Complete independence is assumed across subject blocks.

Random Effects. Random-effects are effects whose values in the data file can be considered a random sample from a larger population of values. They are useful for explaining excess variability in the target.

Estimated Marginal Means. Estimated marginal means can be computed for one or more crossed factors and can be based on either the response or the linear predictor.

Output. The procedure produces a model viewer object containing tables and charts for assessing the quality of the model; also you can save predicted values to the active dataset and the model to a file in PMML format.

Basic Specification

The basic specification is the GENLINMIXED command and the FIELDS subcommand with the TARGET keyword. This fits an intercept-only model that can be used as a basis for comparing more complicated models.

Note: Since measurement level can affect the results, if any variables (fields) have an unknown measurement level, an initial data pass will be performed to determine default measurement level for any variables with an unknown measurement level. For information on the criteria used to determine default measurement level, see SET SCALEMIN.

Syntax Rules

- The FIELDS subcommand is required. All other subcommands are optional.
- Subcommands can be specified in any order.
- There can be multiple RANDOM and EMMEANS subcommands; each RANDOM subcommand defines a separate random effects block, while each EMMEANS subcommand defines a separate set of estimated means. Only a single instance of any other subcommand is allowed.
- An error occurs if a keyword is specified more than once within a subcommand.
- Parentheses, equals signs, and slashes shown in the syntax chart are required.
- The command name, subcommand names, and keywords must be spelled in full.
- Empty subcommands are not allowed.
- Any split field defined on the SPLIT FILE command cannot be used on this command.
- Any field specfed on the WEIGHT command cannot be used on this command.

Limitations

- The SPLIT FILE setting is ignored with a warning by the procedure.

Examples

Common Distribution and Link Function Combinations

Linear model

```
GENLINMIXED
  /FIELDS TARGET=y
  /TARGET_OPTIONS DISTRIBUTION=NORMAL LINK=IDENTITY
  /FIXED EFFECTS=x1 x2 x3.
```

- The FIELDS subcommand specifies *y* as the target.

- The TARGET_OPTIONS subcommand that the target has a normal distribution and is linearly related to the model effects.

- The FIXED subcommand specifies a main effects model with fields *x1*, *x2*, and *x3*. If they are continuous, they will be treated as covariates, if categorical, they will be treated as factors.

Gamma regression

```
GENLINMIXED
  /FIELDS TARGET=claim_amount ANALYSISWEIGHT=number_of_claims
  /TARGET_OPTIONS DISTRIBUTION=GAMMA LINK=POWER(-1)
  /FIXED EFFECTS=holder_age vehicle_group vehicle_age
  /BUILD_OPTIONS INPUTS_CATEGORY_ORDER=DESCENDING.
```

- The procedure fits a model for the target *claim_amount*, using *holder_age*, *vehicle_group*, and *vehicle_age* as main effects. In order to account for the varying number of claims used to compute the average claim amounts, you specify *number_of_claims* as the analysis weight.

- The TARGET_OPTIONS specification assumes that *claim_amount* has a gamma distribution. A power link function with −1 as the exponent relates the distribution of *claim_amount* to a linear combination of the predictors, including an intercept term.

- The BUILD_OPTIONS subcommand specifies that the category order for factors is descending values of factor levels; thus, the first category of each categorical field is used as the reference category.

To compare this specification to the GENLIN procedure, see

(Poisson) Loglinear model

```
GENLINMIXED
  /FIELDS TARGET=damage_incidents OFFSET=log_months_service
  /TARGET_OPTIONS DISTRIBUTION=POISSON LINK=LOG
  /FIXED EFFECTS=type construction operation
  /BUILD_OPTIONS INPUTS_CATEGORY_ORDER=DESCENDING.
```

- The procedure fits a model for the target *damage_incidents*, using *type*, *construction*, and *operation* as main effects.

- The TARGET_OPTIONS specification assumes that *damage_incidents* has a Poisson distribution. A log link function relates the distribution of *damage_incidents* to a linear combination of the predictors, including an intercept term, and an offset equal to the values of *log_months_service*.

- The BUILD_OPTIONS subcommand specifies that the category order for factors is descending values of factor levels; thus, the first category of each categorical field is used as the reference category.

To compare this specification to the GENLIN procedure, see

Negative binomial regression

```
GENLINMIXED
  /FIELDS TARGET=damage_incidents OFFSET=log_months_service
  /TARGET_OPTIONS DISTRIBUTION=NEGATIVE_BINOMIAL LINK=LOG
  /FIXED EFFECTS=type construction operation.
```

- The negative binomial distribution is an alternative to the Poisson when the observations are overdispersed; that is, since the mean and variance of the Poisson distribution are the same, when the data show greater variability, the negative binomial distribution can provide a better fit.

(Nominal) Multinomial logistic regression

```
GENLINMIXED
  /FIELDS TARGET=bfast
  /TARGET_OPTIONS DISTRIBUTION=MULTINOMIAL LINK=LOGIT
  /FIXED EFFECTS=agecat gender active.
```

- The procedure fits a model for *bfast* using *agecat*, *gender*, and *active* as main effects.

- The TARGET_OPTIONS specification assumes that *bfast* has a (nominal) multinomial distribution. A logit link function relates the distribution of *bfast* to a linear combination of the predictors, including an intercept term.

To compare this specification to the NOMREG procedure, see

(Ordinal) Multinomial logistic regression

```
GENLINMIXED
  /FIELDS TARGET=chist
  /TARGET_OPTIONS DISTRIBUTION=MULTINOMIAL LINK=CLOGLOG
  /FIXED EFFECTS=numcred othnstal housng age duration.
```

- The procedure fits a model for *chist* using *numcred*, *othnstal*, *housing*, *age*, and *duration* as main effects. Because *numcred*, *othnstal*, and *housing* have categorical measurement level, they are treated as factors; *age*, and *duration* have continuous (scale) measurement level and are treated as covariates.

- The TARGET_OPTIONS specification assumes that *chist* has an (ordinal) multinomial distribution. A (cumulative) complementary log−log link function relates the distribution of *chist* to a linear combination of the predictors, including threshold terms for the categories of *chist* (except the last category).

To compare this specification to the PLUM procedure, see Example on p. 1573

Binary logistic regression

```
GENLINMIXED
  /FIELDS TARGET=default
  /TARGET_OPTIONS DISTRIBUTION=BINOMIAL LINK=LOGIT
  /FIXED EFFECTS=age ed employ address income debtinc creddebt othdebt.
```

■ The procedure fits a model for *default* using *age*, *ed*, *employ*, *address*, *income*, *debtinc*, *creddebt*, and *othdebt* as main effects.

■ The TARGET_OPTIONS specification assumes that *default* has a binomial distribution. A logit link function relates the distribution of *default* to a linear combination of the predictors, including an intercept term.

To compare this specification to the LOGISTIC REGRESSION procedure, see Examples on p. 1073

Binary probit model

```
GENLINMIXED
  /FIELDS TARGET=response TRIALS=nsubj
  /TARGET_OPTIONS DISTRIBUTION=BINOMIAL LINK=PROBIT
  /FIXED EFFECTS=site value.
```

■ The procedure fits a model for the number of responses *response* within trials *nsubj* using site and value as main effects.

■ The TARGET_OPTIONS specification assumes that *response* within *nsubj* has a binomial distribution. A probit link function relates the distribution of the target to a linear combination of the predictors, including an intercept term.

To compare this specification to the PROBIT procedure, see Examples on p. 1643

Interval censored survival

```
GENLINMIXED
  /FIELDS TARGET=result2
  /TARGET_OPTIONS DISTRIBUTION=BINOMIAL LINK=CLOGLOG
  /FIXED EFFECTS=duration treatment period age USE_INTERCEPT=FALSE
  /BUILD_OPTIONS TARGET_CATEGORY_ORDER=DESCENDING INPUTS_CATEGORY_ORDER=DESCENDING.
```

■ The procedure fits a model for the target *result2*, using *duration*, *treatment*, *period*, and *age* as main effects.

■ The BUILD_OPTIONS subcommand specifies that the category order for the target and all factors is descending values of factor levels; thus, the first category of each categorical field is used as the reference category.

■ The TARGET_OPTIONS specification assumes that *result2* has a binomial distribution. A complementary log-log link function relates the probability of *result2* to a linear combination of the predictors, excluding an intercept term.

To compare this specification to the GENLIN procedure, see Examples on p. 784

Linear mixed model

```
*Generalized Linear Mixed Models.
GENLINMIXED
  /DATA_STRUCTURE SUBJECTS=school*classroom*student_id
  /FIELDS TARGET=posttest TRIALS=NONE OFFSET=NONE
  /TARGET_OPTIONS DISTRIBUTION=NORMAL LINK=IDENTITY
  /FIXED  EFFECTS=school_setting school_type teaching_method n_student gender lunch pretest
    USE_INTERCEPT=TRUE
  /RANDOM USE_INTERCEPT=TRUE SUBJECTS=school COVARIANCE_TYPE=VARIANCE_COMPONENTS
  /RANDOM USE_INTERCEPT=TRUE SUBJECTS=school*classroom COVARIANCE_TYPE=VARIANCE_COMPONENTS
  /BUILD_OPTIONS TARGET_CATEGORY_ORDER=ASCENDING INPUTS_CATEGORY_ORDER=ASCENDING MAX_ITERA
    CONFIDENCE_LEVEL=95 DF_METHOD=RESIDUAL COVB=MODEL
  /EMMEANS_OPTIONS SCALE=ORIGINAL PADJUST=LSD.
```

- The DATA_STRUCTURE subcommand specifies that subjects are defined by *school*, *classroom*, and *student_id*.

- The FIELDS and TARGET_OPTIONS subcommands specify that the procedure creates a model for *posttest* using a normal distribution to fit the test score and an identity link to relate the target to a linear combination of the predictors.

- The FIXED subcommand speficies a model with *school_setting*, *school_type*, *teaching_method*, *n_student*, *gender*, *lunch*, and *pretest* as main effects.

- The first RANDOM subcommand specifies an intercept-only random effect block with *school* as the subject field. This should account for correlation between classrooms within the same school.

- The second RANDOM subcommand specifies an intercept-only random effect block with *school*classroom* as the subject field. This should account for correlation between students within the same classroom.

- All other options are set to their default values.

Poisson loglinear mixed model

```
*Generalized Linear Mixed Models.
GENLINMIXED
  /DATA_STRUCTURE SUBJECTS=patient_id
  /FIELDS TARGET=convulsions TRIALS=NONE OFFSET=NONE
  /TARGET_OPTIONS DISTRIBUTION=POISSON LINK=LOG
  /FIXED  EFFECTS=after_t treatment*after_t USE_INTERCEPT=TRUE
  /RANDOM EFFECTS=after_t USE_INTERCEPT=TRUE SUBJECTS=patient_id COVARIANCE_TYPE=UNSTRUCTU
  /BUILD_OPTIONS TARGET_CATEGORY_ORDER=ASCENDING INPUTS_CATEGORY_ORDER=ASCENDING MAX_ITERA
    CONFIDENCE_LEVEL=95 DF_METHOD=RESIDUAL COVB=MODEL
  /EMMEANS_OPTIONS SCALE=ORIGINAL PADJUST=LSD.
```

- The DATA_STRUCTURE subcommand specifies that subjects are defined by *patient_id*.

- The FIELDS and TARGET_OPTIONS subcommands specify that the procedure creates a model for *convulsions* using a Poisson distribution to fit the number of convulsions and a log link to relate the target to a linear combination of the predictors.

- The FIXED subcommand specifies a model with *after_t* and *treatment*after_t* as effects.

- The RANDOM subcommand specifies *after_t* and an intercept as effects in a random effect block with *patient_id* as the subject field. This should account for correlation between repeated observations of the same patient.

- All other options are set to their default values.

Multinomial logistic mixed model

```
*Generalized Linear Mixed Models.
GENLINMIXED
  /DATA_STRUCTURE SUBJECTS=customer_id
  /FIELDS TARGET=service_usage TRIALS=NONE OFFSET=NONE
  /TARGET_OPTIONS DISTRIBUTION=MULTINOMIAL LINK=LOGIT
  /FIXED  EFFECTS=edcat inccat reside service_type USE_INTERCEPT=TRUE
  /RANDOM USE_INTERCEPT=TRUE SUBJECTS=customer_id COVARIANCE_TYPE=VARIANCE_COMPONENTS
  /BUILD_OPTIONS TARGET_CATEGORY_ORDER=ASCENDING INPUTS_CATEGORY_ORDER=ASCENDING MAX_ITERA
    CONFIDENCE_LEVEL=95 DF_METHOD=RESIDUAL COVB=MODEL
  /EMMEANS_OPTIONS SCALE=ORIGINAL PADJUST=LSD.
```

- The DATA_STRUCTURE subcommand specifies that subjects are defined by *customer_id*.

- The FIELDS and TARGET_OPTIONS subcommands specify that the procedure creates a model for *service_usage* using a multinomial distribution and a logit link to relate the target to a linear combination of the predictors.

- The FIXED subcommand specifies a model with *edcat*, *inccat*, *reside*, and *service_type* as main effects.

- The RANDOM subcommand specifies an intercept-only random effect block with *customer_id* as the subject field. This should account for correlation between answers to the service usage questions across service types (tv, phone, internet) within a given survey responder's answers.

- All other options are set to their default values.

DATA_STRUCTURE Subcommand

The DATA_STRUCTURE subcommand specifies the subject structure for repeated measurements and how the errors of the repeated measurements are correlated. If the DATA_STRUCTURE subcommand is not specified, then the model assumes that all error terms are independent.

- Records with missing values for any field on the DATA_STRUCTURE subcommand are not used in the analysis.

SUBJECTS Keyword

The SUBJECTS keyword identifies subjects in the active dataset. Complete independence is assumed *across* subjects, but responses *within* subjects are assumed to be correlated.

- Specify a single categorical field or a list of categorical fields connected by asterisks (*) or the keyword BY.

- The number of subjects equals the number of distinct combinations of values of the fields.

- Any field specified on the SUBJECTS keyword cannot be used as a REPEATED_MEASURES field or on the FIELDS subcommand.

- The SUBJECTS keyword is required if the DATA_STRUCTURE subcommand is used.

REPEATED_MEASURES Keyword

The REPEATED_MEASURES keyword gives the repeated (or within-subject) effect. This effect defines the ordering of repeated measurements within subjects. If some measurements do not appear in the data for some subjects, then the existing measurements are ordered and the omitted

measurements are treated as missing values. If REPEATED_MEASURES is not specified, then no repeated measures model is built.

- Specify a single field or a list of fields connected by asterisks (*) or the keyword BY.

- Each distinct combination of the values of the fields defines a separate repeated measure.

- Any field specified on the REPEATED_MEASURES keyword cannot be used as a SUBJECTS keyword or on the FIELDS subcommand.

- Each distinct repeated measures value must occur only once within a subject. Generally speaking, you should not use a WEIGHT field if REPEATED MEASURES is specified.

GROUPING Keyword

The GROUPING keyword allows you to define independent sets of covariance parameters. All subjects have the same covariance type; subjects within the same covariance grouping will have the same values for the parameters.

- Specify a single categorical field or a list of categorical fields connected by asterisks (*) or the keyword BY.

- Each distinct combination of the values of the fields defines a separate covariance grouping.

- A field specified on the GROUPING keyword can also be used as a SUBJECTS or REPEATED_MEASURES field, the TARGET on the FIELDS subcommand, or on the EFFECTS keyword or the FIXED or RANDOM subcommand.

- The GROUPING keyword is optional.

- If the GROUPING keyword is used, there must also be a REPEATED_MEASURES specification.

COVARIANCE_TYPE Keyword

The COVARIANCE_TYPE keyword gives the covariance structure of the residual covariance matrix.

AR1. *First-order autoregressive.* The constraint $|\rho| \leq 1$ is imposed for stationarity.

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

ARMA11. *Autoregressive moving average (1,1).* The constraints $|\varphi| \leq 1$ and $|\rho| \leq 1$ are imposed for stationarity.

$$\sigma^2 \begin{bmatrix} 1 & \phi\rho & \phi\rho^2 & \phi\rho^3 \\ \phi\rho & 1 & \phi\rho & \phi\rho^2 \\ \phi\rho^2 & \phi\rho & 1 & \phi\rho \\ \phi\rho^3 & \phi\rho^2 & \phi\rho & 1 \end{bmatrix}$$

COMPOUND_SYMMETRY. This structure has constant variance and constant covariance.

$$\begin{bmatrix} \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 \end{bmatrix}$$

DIAGONAL. This is a diagonal structure with heterogenous variance. This is the default.

$$\begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}$$

IDENTITY. This is a scaled identity matrix.

$$\sigma^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

TOEPLITZ. ($|\rho_k| \le 1$).

$$\sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{bmatrix}$$

UNSTRUCTURED. This is a completely general covariance matrix.

$$\begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

VARIANCE_COMPONENTS. For repeated measures, the variance components structure is equivalent to the diagonal (DIAGONAL) structure.

FIELDS Subcommand

The FIELDS subcommand specifies the target, optional offset, and optional analysis (regression) weight.

- This subcommand, and the TARGET keyword, are required.
- A field can only be specified once on the FIELDS subcommand.
- A categorical TARGET can be specified as a GROUPING field on the DATA_STRUCTURE or RANDOM subcommand, but not in any other capacity. Any other field specified on the FIELDS subcommand cannot be used on any other subcommand.

TARGET. The target field is what the model tries to predict. Specify a single field.

TRIALS = NONE | VALUE(number) | FIELD(field).** If the model response can be expressed as the number of events occurring within a number of trials, then the TARGET keyword specifies the number of events and TRIALS specifies the number of trials. Use VALUE with the number of trials in parentheses if the number of trials is fixed for all subjects and FIELD with a field name in parentheses if the number of trials varies across subjects and there is a field containing the number of trials.

The procedure automatically computes the ratio of the events field over the trials field or number. Technically, the procedure treats the events field as the target in the sense that predicted values and residuals are based on the events field rather than the events/trials ratio.

ANALYSIS_WEIGHT. The model uses analysis (regression) weights to account for differences in variability across levels of the target field. Specify a single continuous field.

OFFSET = NONE** | VALUE(value) | FIELD(field). The offset is a constant added to the intercept. Specify a continuous field or number.

- Records with missing values on the OFFSET field are not used in the analysis.

- Specifying a number when /FIXED USE_INTERCEPT = YES is equivalent to adding a constant to the intercept.

- Specifying a number when /FIXED USE_INTERCEPT = NO is equivalent to fixing the intercept at the specified number.

TARGET_OPTIONS Subcommand

The TARGET_OPTIONS subcommand specifies the reference category, distribution, and link function for the target.

- This subcommand is optional. If the TARGET_OPTIONS subcommand is not specified, then GENLINMIXED assigns a default distribution and link function based on the measurement level of the target. A continuous target has NORMAL distribution and IDENTITY link. A categorical target has MULTINOMIAL distribution and LOGIT link.

REFERENCE. For categorical targets, this is the target value to use as the reference category for parameter estimation. No model parameters are assigned to the reference category. Specify a single value. Put the value inside a pair of quotes if it is formatted (such as date or time) or if the target is of string storage type. The REFERENCE value must be found in the dataset. If REFERENCE is not specified, the last category in ascending order is used as the reference.

- If the target has ordinal measurement level, the REFERENCE keyword is ignored.

DISTRIBUTION Keyword

The DISTRIBUTION keyword specifies the probability distribution of the target.

- The default probability distribution depends on the measurement level of the target. If the target is continuous, the default distribution is NORMAL. If the target is categorical, then the default is MULTINOMIAL. If an events/trials specification is used, then the default distribution is BINOMIAL.

- Caution must be exercised when the LINK but not the DISTRIBUTION keyword is used. The discussion of the LINK keyword below gives details about proper and improper combinations of DISTRIBUTION and LINK settings.

BINOMIAL. *Binomial probability distribution.* If the target is specified as a single field, then it can be numeric or string and it can have only two distinct valid data values.

If the events and trials options are specified, then the procedure automatically computes the ratio of the events field over the trials field or number. The events field—and the trials field if specified—must be numeric. Data values for the events field must be integers greater than or equal to zero. Data values for the trials field must be integers greater than zero. For each record, the trials value must be greater than or equal to the events value. If an events value is noninteger, less than zero, or missing, then the corresponding record is not used in the analysis. If a trials value is noninteger, less than or equal to zero, less than the events value, or missing, then the corresponding record is not used in the analysis.

If the trials option specifies a number, then it must be a positive integer, and it must be greater than or equal to the events value for each record. Records with invalid values are not used in the analysis.

This is the default probability distribution if the target is specified using events/trials format.

GAMMA. *Gamma probability distribution.* The target must be numeric, with data values greater than zero. If a data value is less than or equal to zero, or missing, then the corresponding record is not used in the analysis.

INVERSE_GAUSSIAN. *Inverse Gaussian probability distribution.* The target must be numeric, with data values greater than zero. If a data value is less than or equal to zero, or missing, then the corresponding record is not used in the analysis.

MULTINOMIAL. *Multinomial probability distribution.* The target can be numeric or string, and it must have at least two distinct, valid data values. This is the default probability distribution if the target is categorical. Note that the form of the model will depend on the measurement level of the target.

- A **nominal** target will result in a nominal multinomial model in which a separate set of model parameters are estimated for each category of the target (except the reference category). The parameter estimates for a given predictor show the relationship between that predictor and the likelihood of each category of the target, relative to the reference category.

- An **ordinal** target will result in an ordinal multinomial model in which the traditional intercept term is replaced with a set of **threshold** parameters that relate to the cumulative probability of the target categories.

NEGATIVE_BINOMIAL. *Negative binomial probability distribution.* The target must be numeric, with data values that are integers greater than or equal to zero. If a data value is noninteger, less than zero, or missing, then the corresponding record is not used in the analysis.

NORMAL. *Normal probability distribution.* The target must be numeric. This is the default probability distribution if the target is continuous.

POISSON. *Poisson probability distribution.* The target must be numeric, with data values that are integers greater than or equal to zero. If a data value is noninteger, less than zero, or missing, then the corresponding record is not used in the analysis.

LINK Keyword

The `LINK` keyword specifies the link function. The following link functions are available.

IDENTITY. *Identity link function. f(x)=x.* This link function is valid with any distribution except the multinomial.

CAUCHIT. *Cauchit link function.* $f(x) = \tan(\pi (x - 0.5))$. This is only valid with ordinal multinomial models.

CLOGLOG. *Complementary log-log link function.* $f(x)=\ln(-\ln(1-x))$. This valid with binomial and ordinal multinomial models.

LOG.*Log link function.* $f(x)=\ln(x)$. This link function is valid with any distribution except the multinomial.

LOGC. *Log complement link function.* $f(x)=\ln(1-x)$. This is only valid with the binomial distribution.

LOGIT. *Logit link function.* $f(x)=\ln(x / (1-x))$. This is valid with the binomial and multinomial distributions.

NLOGLOG. *Negative log-log link function.* $f(x)=-\ln(-\ln(x))$. This valid with binomial and ordinal multinomial models.

POWER. *Power link function.* $f(x)=x^{\alpha}$, if $\alpha \neq 0$. $f(x)=\ln(x)$, if $\alpha=0$. α is a required number specified on the `LINKPARAMETER` keyword and must be a real number. If $|\alpha| < 2.2e\text{-}16$, α is treated as 0. This link function is valid with any distribution except the multinomial.

PROBIT. *Probit link function.* $f(x)=\Phi^{-1}(x)$, where Φ^{-1} is the inverse standard normal cumulative distribution function. This valid with binomial and ordinal multinomial models.

■ The default setting for `LINK` depends on the `DISTRIBUTION` setting, whether `DISTRIBUTION` is set explicitly or is the default, as shown in the following table.

DISTRIBUTION Setting	Default LINK Setting
NORMAL	IDENTITY
BINOMIAL	LOGIT
GAMMA	POWER(−1)
INVERSE_GAUSSIAN	POWER(−2)
MULTINOMIAL	LOGIT
NEGATIVE_BINOMIAL	LOG
POISSON	LOG

■ The procedure will fit a model if a permissible combination of `LINK` and `DISTRIBUTION` specifications is given. The table below indicates the permissible `LINK` and `DISTRIBUTION` combinations. Specifying an improper combination will yield an error message.

■ Note that the default setting for `DISTRIBUTION` is irrespective of the `LINK` specification, and that not all `LINK` specifications are valid for the default `DISTRIBUTION`. Thus, if `LINK` = `CLOGLOG` is specified for a continuous target but `DISTRIBUTION` is not, then the default

DISTRIBUTION = NORMAL yields an improper combination of DISTRIBUTION and LINK settings.

Table 98-1
Valid combinations of distribution and link function

Link	NORMAL	BINOMIAL	GAMMA	INVERSE GAUSSIAN	ORDINAL MULTINOMIAL	NOMINAL MULTINOMIAL	NEGATIVE BINOMIAL	POISSON
IDENTITY	X	X	X	X			X	X
CAUCHIT					X			
CLOGLOG		X			X			
LOG	X	X	X	X			X	X
LOGC		X						
LOGIT		X			X	X		
NLOGLOG		X			X			
PROBIT		X			X			
POWER	X	X	X	X			X	X

LINKPARAMETER Keyword

The LINKPARAMETER keyword specifies the value of α for the POWER link.

■ Specify a number. There is no default.

FIXED Subcommand

The FIXED subcommand is used to specify fixed model effects, including the intercept.

■ This subcommand is optional. If not specified, then the procedure builds an intercept-only model.

EFFECTS. The effect list includes all effects to be included in the fixed model except for the intercept, which is specified using the USE_INTERCEPT keyword.

■ To include a term for the main effect of a factor (categorical predictor) or covariate (continuous predictor), enter its field name. Whether a field is treated as a factor or covariate depends upon its measurement level. Fields with categorical (flag, nominal, or ordinal) measurement level are treated as factors while fields with continuous (scale) measurement level are treated as covariates.

■ To include a term for an interaction between factors, use the keyword BY or an asterisk (*) to join the factors involved in the interaction. For example, A*B means a two-way interaction effect of A and B, where A and B are factors. A*A is not allowed because factors in an interaction effect must be distinct.

■ To include a term for nesting one effect within another, use a pair of parentheses. For example, A(B) means that A is nested within B.

■ Multiple nesting is allowed. For example, A(B(C)) means that B is nested within C, and A is nested within B(C). When more than one pair of parentheses is present, each pair of parentheses must be enclosed or nested within another pair of parentheses. Thus, A(B)(C) is not valid.

- Interactions between nested effects are not valid. For example, neither A(C)*B(C) nor A(C)*B(D) is valid.

- Covariates can be connected, but not nested, through the * operator to form another covariate effect. Interactions among covariates such as X1*X1 and X1*X2 are valid, but X1(X2) is not.

- Factor and covariate effects can be connected only by the * operator. Suppose A and B are factors, and X1 and X2 are covariates. Examples of valid factor-by-covariate interaction effects are A*X1, A*B*X1, X1*A(B), A*X1*X1, and B*X1*X2.

USE_INTERCEPT = TRUE** | FALSE. This keyword controls whether an intercept term is included in the model. Specifying USE_INTERCEPT=FALSE when there are no effects in the model fits a "null" model. The default is TRUE.

RANDOM Subcommand

The RANDOM subcommand specifies the random effects in the mixed model.

- Depending on the covariance type specified, random effects specified in one RANDOM subcommand may be correlated.

- One block of the covariance G matrix will be constructed for each RANDOM subcommand. The dimension of the random effect covariance G matrix is equal to the sum of the levels of all random effects across all RANDOM subcommands.

- When the variance components (VC) structure is specified, a scaled identity (ID) structure will be assigned to each of the effects specified. This is the default covariance type for the RANDOM subcommand.

- Use a separate RANDOM subcommand when a different covariance structure is assumed for a list of random effects. If the same effect is listed on more than one RANDOM subcommand, it must be associated with a different SUBJECT combination.

- No random effects are included in the mixed model unless a RANDOM subcommand is specified correctly.

EFFECTS. The effect list includes all effects to be included in the random effects model block except for the intercept, which is specified using the USE_INTERCEPT keyword.

- To include a term for the main effect of a factor (categorical predictor) or covariate (continuous predictor), enter its field name. Whether a field is treated as a factor or covariate depends upon its measurement level.

- To include a term for an interaction between factors, use the keyword BY or an asterisk (*) to join the factors involved in the interaction. For example, A*B means a two-way interaction effect of A and B, where A and B are factors. A*A is not allowed because factors in an interaction effect must be distinct.

- To include a term for nesting one effect within another, use a pair of parentheses. For example, A(B) means that A is nested within B.

- Multiple nesting is allowed. For example, A(B(C)) means that B is nested within C, and A is nested within B(C). When more than one pair of parentheses is present, each pair of parentheses must be enclosed or nested within another pair of parentheses. Thus, A(B)(C) is not valid.

- Interactions between nested effects are not valid. For example, neither A(C)*B(C) nor A(C)*B(D) is valid.

- Covariates can be connected, but not nested, through the * operator to form another covariate effect. Interactions among covariates such as X1*X1 and X1*X2 are valid, but X1(X2) is not.

- Factor and covariate effects can be connected only by the * operator. Suppose A and B are factors, and X1 and X2 are covariates. Examples of valid factor-by-covariate interaction effects are A*X1, A*B*X1, X1*A(B), A*X1*X1, and B*X1*X2.

USE_INTERCEPT. This keyword controls whether an intercept term is included in the model. It is invalid to specify USE_INTERCEPT=FALSE if there are no effects specified on the EFFECTS keyword. The default is FALSE.

Examples

```
GENLINMIXED
  /FIXED EFFECTS=a b
  /RANDOM SUBJECTS=id.

GENLINMIXED
  /FIXED EFFECTS=a b
  /RANDOM USE_INTERCEPT=TRUE SUBJECTS=id.

GENLINMIXED
  /FIXED EFFECTS=a b
  /RANDOM EFFECTS=c d SUBJECTS=id.
```

- The first command fails because the RANDOM subcommand has no effects or intercept. The next two commands succeed.

SUBJECTS Keyword

The SUBJECTS keyword identifies subjects in the active dataset. Complete independence is assumed *across* subjects, but responses *within* subjects are assumed to be correlated.

- Specify a single field or a list of fields connected by asterisks (*) or the keyword BY.

- The number of subjects equals the number of distinct combinations of values of the fields.

- The fields in the RANDOMSUBJECTS list must be a subset of the fields on the DATA_STRUCTURE SUBJECTS list.

- The SUBJECTS keyword is optional.

GROUPING Keyword

The GROUPING keyword allows you to define independent sets of covariance parameters. All subjects have the same covariance type; subjects within the same covariance grouping will have the same values for the parameters.

- Specify a single field or a list of fields connected by asterisks (*) or the keyword BY.

- Each distinct combination of the values of the fields defines a separate covariance grouping.

- A field specified on the `GROUPING` keyword can also be used as a `SUBJECTS` or `REPEATED_MEASURES` field, the `TARGET` on the `FIELDS` subcommand, or on the `EFFECTS` keyword or the `FIXED` or `RANDOM` subcommand.

- The `GROUPING` keyword is optional.

COVARIANCE_TYPE Keyword

The `COVARIANCE_TYPE` keyword gives the covariance structure of the random effect covariance matrix.

- Random effects are considered independent of each other, and a separate covariance matrix is computed for each effect.

AR1. *First-order autoregressive.* The constraint $|\rho| \leq 1$ is imposed for stationarity.

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

ARMA11. *Autoregressive moving average (1,1).* The constraints $|\varphi| \leq 1$ and $|\rho| \leq 1$ are imposed for stationarity.

$$\sigma^2 \begin{bmatrix} 1 & \phi\rho & \phi\rho^2 & \phi\rho^3 \\ \phi\rho & 1 & \phi\rho & \phi\rho^2 \\ \phi\rho^2 & \phi\rho & 1 & \phi\rho \\ \phi\rho^3 & \phi\rho^2 & \phi\rho & 1 \end{bmatrix}$$

COMPOUND_SYMMETRY. This structure has constant variance and constant covariance.

$$\begin{bmatrix} \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 \end{bmatrix}$$

DIAGONAL. This is a diagonal structure with heterogenous variance.

$$\begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}$$

IDENTITY. This is a scaled identity matrix.

$$\sigma^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

TOEPLITZ. ($|\rho_k| \leq 1$).

$$\sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{bmatrix}$$

UNSTRUCTURED. This is a completely general covariance matrix.

$$\begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

VARIANCE_COMPONENTS. *Variance components.* This is the default covariance structure for random effects. The variance components structure for random effects is a scaled identity (ID) structure assigned to each of the effects specified on the subcommand.

BUILD_OPTIONS Subcommand

The BUILD_OPTIONS subcommand specifies the criteria used to build the model.

TARGET_CATEGORY_ORDER. This determines the order of the target's categories for purposes of determining the "last" category, which is used as the reference category if the REFERENCE keyword is not used. This keyword is ignored if the target is not categorical.

- **ASCENDING.** *Target categories are sorted in ascending order, from the lowest value to the highest value.* The highest value is treated as the "last" category. This is the default.

- **DESCENDING.***Target categories are sorted in descending order, from the highest value to the lowest value.* The lowest value is treated as the "last" category.

- **DATA.***Target values are not sorted.* The first value encountered in the data defines the first category, the last value encountered defines the last category. This option is technically not valid if splits are defined on the SPLIT FILE command, but the procedure will run while it does not honor splits.

INPUTS_CATEGORY_ORDER. This determines the order of the categories for factors (categorical inputs) for purposes of determining the "last" category, which is used as the reference category. This keyword is ignored if there are no factors.

- **ASCENDING.** *Factor values are sorted in ascending order, from the lowest value to the highest value.* The highest value is treated as the "last" category. This is the default.

- **DESCENDING.***Factor values are sorted in descending order, from the highest value to the lowest value.* The lowest value is treated as the "last" category.

- **DATA.***Factor values are not sorted.* The first value encountered in the data defines the first category, the last value encountered defines the last category. This option is technically not valid if splits are defined on the SPLIT FILE command, but the procedure will run while it does not honor splits.

MAX_ITERATIONS = number. The model building algorithm stops after a certain number of iterations. By default, this is 100. Alternatively, specify a non-negative integer maximum number of iterations.

CONFIDENCE_LEVEL. This is the level of confidence used to compute interval estimates of the model coefficients. Specify a value greater than 0 and less than 100. The default is 95.

DF_METHOD. This specifies how degrees of freedom are computed for significance tests.

- **RESIDUAL.** The residual method has a fixed degrees of freedom for all tests. It is useful if your sample size is sufficiently large, or the data are balanced, or the model uses a simpler covariance type; for example, scaled identity or diagonal. This is the default.
- **SATTERTHWAITE.** The Satterthwaite method has a field degrees of freedom across tests. It is useful if your sample size is small, or the data are unbalanced, or the model uses a complicated covariance type; for example, unstructured.

COVB. This is the method for computing the parameter estimates covariance matrix.

- **MODEL.** The model-based estimate assumes that the model assumptions are correct. This is the default.
- **ROBUST.** The robust estimate is useful if you are concerned that the model assumptions are violated.

EMMEANS Subcommand

The EMMEANS subcommand displays estimated marginal means of the target for all level combinations of a set of factors. Note that these are predicted, not observed, means. Estimated marginal means can be computed based on the original scale of the target or the based on the link function transformation.

- Multiple EMMEANS subcommands are allowed. Each is treated independently.
- The EMMEANS subcommand can be specified with no additional keywords. The output for an empty EMMEANS subcommand is the overall estimated marginal mean of the response, collapsing over any factors.
- Estimated marginal means are not available if the multinomial distribution is used. If DISTRIBUTION = MULTINOMIAL on the MODEL subcommand and the EMMEANS subcommand is specified, then EMMEANS is ignored and a warning is issued.

TABLES Keyword

The TABLES keyword specifies the cells for which estimated marginal means are displayed.

- Valid options are all-factor effects appearing on the EFFECTS keyword of the FIXED subcommand, including main effects factors such as a b, all-factor interactions such as a*b, and all-factor nested effects such as a(b), if *a* and *b* are factors.
- If the TABLES keyword is specified, then the procedure collapses over any other factors specified on the command but not on the TABLES keyword before computing the estimated marginal means for the target.
- If the TABLES keyword is not specified, then the overall estimated marginal mean of the target, collapsing over any factors, is computed.

Examples

```
GENLINMIXED
  /FIXED EFFECTS=a b
  /EMMEANS TABLES=a*b.

GENLINMIXED
  /FIXED EFFECTS=a*b
  /EMMEANS TABLES=a
  /EMMEANS TABLES=b.

GENLINMIXED
  /FIXED EFFECTS=a b a*b
  /EMMEANS TABLES=a
  /EMMEANS TABLES=b
  /EMMEANS TABLES=a*b.
```

■ The first two commands fail because the effects listed on the `TABLES` keywords are not listed on the `EFFECTS` keyword. The third command succeeds.

COMPARE Keyword

The `COMPARE` keyword specifies a factor, the levels of which are compared using the contrast type specified on the `CONTRAST` keyword.

■ Valid options are factors appearing on the `TABLES` keyword.

■ The `COMPARE` keyword is valid only if the `TABLES` and `CONTRAST` keyword is also specified.

■ By default, the procedure sorts levels of the factors in ascending order and defines the highest level as the last level. (If the factor is a string variable, then the value of the highest level is locale-dependent.) However, the sort order can be modified using the `INPUTS_SORT_ORDER` keyword on the `BUILD_OPTIONS` subcommand.

CONTRAST Keyword

The `CONTRAST` keyword specifies the type of contrast to use for the levels of the factor on the `COMPARE` keyword. The `CONTRAST` keyword creates an **L** matrix (that is, a coefficient matrix) such that the columns corresponding to the factor match the contrast given. The other columns are adjusted so that the **L** matrix is estimable.

■ The `CONTRAST` keyword is valid only if the `COMPARE` keyword is also specified.

■ If the `COMPARE` keyword is specified without `CONTRAST`, then pairwise comparisons are performed for the factor on `COMPARE`.

■ `SIMPLE` contrasts are defined with respect to the last level as determined by the `INPUTS_CATEGORY_ORDER` specification on the `BUILD_OPTIONS` subcommand.

The following contrast types are available.

NONE. *No comparisons.* This is the default.

PAIRWISE. *Pairwise comparisons are computed for all levels of the specified or factor.* Pairwise contrasts are not orthogonal.

DEVIATION. *Each level of the factor is compared to the grand mean.* Deviation contrasts are not orthogonal.

SIMPLE. *Each level of the factor except the last is compared to the last level.* Simple contrasts are not orthogonal.

EMMEANS_OPTIONS Subcommand

The EMMEANS_OPTIONS subcommand contains specifications that are used across all EMMEANS subcommands.

CONTROL Keyword

The CONTROL keyword specifies the covariate values to use when computing the estimated marginal means.

- Specify one or more covariates appearing on the FIXED subcommand EFFECTS list, each of which must be followed by a numeric value.
- If a covariate is not specified on the CONTROL option, then its overall mean will be used in estimated marginal means calculations.
- Any covariate can occur only once on the CONTROL keyword.

SCALE Keyword

The SCALE keyword specifies whether to compute estimated marginal means based on the original scale of the target or based on the link function transformation.

ORIGINAL. *Estimated marginal means are based on the original scale of the target.* Estimated marginal means are computed for the target. This is the default. Note that when the target is specified using the events/trials option, ORIGINAL gives the estimated marginal means for the events/trials proportion rather than for the number of events.

TRANSFORMED. *Estimated marginal means are based on the link function transformation.* Estimated marginal means are computed for the linear predictor.

PADJUST Keyword

The PADJUST keyword indicates the method of adjusting the significance level.

LSD. *Least significant difference.* This method does not control the overall probability of rejecting the hypotheses that some linear contrasts are different from the null hypothesis value(s). This is the default.

SEQBONFERRONI. *Sequential Bonferroni.* This is a sequentially step-down rejective Bonferroni procedure that is much less conservative in terms of rejecting individual hypotheses but maintains the same overall significance level.

SEQSIDAK. *Sequential Sidak.* This is a sequentially step-down rejective Sidak procedure that is much less conservative in terms of rejecting individual hypotheses but maintains the same overall significance level.

OUTFILE Subcommand

The OUTFILE subcommand saves a file containing the scoring model.

MODEL = 'file'. Writes the fixed effects portion of the model to an external *.zip* file. You can use this model file to apply the model information to other data files for scoring purposes. Specify a unique, valid filename. If the file specification refers to an existing file, then the file is overwritten.

SAVE Subcommand

The SAVE subcommand writes optional temporary fields to the active dataset.

PREDICTED_VALUES(varname). *Predicted value of the target.* The default field name is *PredictedValue*.

PREDICTED_PROBABILITY(rootname). *Predicted probabilities of the target.* If the target is categorical, this keyword saves the predicted probabilities of the first *n* categories. The values are cumulative probabilities for ordinal targets. The value of *n* is specified by the MAX_CATEGORIES keyword.

- Specify a unique, valid variable name. The default root name is *PredictedProbability*.
- If the dependent variable is continuous, this keyword is ignored with a warning.

MAX_CATEGORIES. *The maximum number of categories for which predicted probabilities and confidence intervals (for a categorical target) should be saved.*

- Specify a positive integer. The default is 25.
- This keyword is ignored if predicted probabilities or confidence intervals (for a categorical target) are not saved.

CONFIDENCE(varname). *Confidence in predicted value for categorical target.* The computed confidence can be based on the probability of the predicted value (the highest predicted probability) or the difference between the highest predicted probability and the second highest predicted probability. The default field name is *Confidence*.

CONFIDENCE_METHOD. *Method for computing confidence.* The computed confidence can be based on the probability of the predicted value (the highest predicted probability) or the difference between the highest predicted probability and the second highest predicted probability.

CONFIDENCE_INTERVALS(rootname). *Upper and lower confidence intervals.* For all distributions except the multinomial, this creates two variables and the default root name is CI, with _Lower and _Upper as the suffixes. Specify a root name in parentheses to override the default.

For the multinomial distribution and a nominal target, one field is created for each dependent variable category. CONFIDENCE_INTERVALS saves the lower and upper bounds of the predicted probability for the first 25 categories by default. The default root name is *CI*, and the default field names are *CI_Lower_1*, *CI_Upper_1*, *CI_Lower_2*, *CI_Upper_2*, and so on, corresponding to the order of the target categories. Specify a root name in parentheses to override the default. The value of *n* is specified by the MAX_CATEGORIES keyword.

For the multinomial distribution and an ordinal target, one field is created for each dependent variable category except the last (see the TARGET_SORT_ORDER keyword). CONFIDENCE_INTERVALS saves the lower and upper bounds of the cumulative predicted probability for the first 25 categories, up to but not including the last, by default. The default root name is *CI*, and the default field names are *CI_Lower_1*, *CI_Upper_1*, *CI_Lower_2*, *CI_Upper_2*, and so on, corresponding to the order of the target categories. Specify a root name in parentheses to override the default. The value of *n* is specified by the MAX_CATEGORIES keyword.

PEARSON_RESIDUALS(varname). *Pearson residuals.* The default field name is *PearsonResidual*.