ODA Command List

ATTRIBUTE

Syntax ATTRIBUTE *variable list* ;

Alias ATTR

Remarks The ATTRIBUTE command specifies the attribute variable or variables to be used in the analysis. This command is required unless TABLE input is specified. If more than one variable is named, a separate analysis will be run for each variable. The TO keyword may be used to define multiple variables in the variable list. For example, the command

ATTRIBUTE A1 TO A4 ;

indicates that the variables A1, A2, A3, A4 will be treated as attributes. Further exposition of the TO keyword may be found in the discussion for VARS.

CATEGORICAL

Syntax CATEGORICAL {ON | OFF} ;
CATEGORICAL *variable list* ;

Alias CAT

Remarks The CATEGORICAL command specifies that categorical analysis will be used: It is required when the attribute to be analyzed is categorical. Using the ON keyword indicates that all variables in the variable list are categorical. CATEGORICAL with no parameters is the same as CATEGORICAL ON. The TO keyword may be used in the variable list (see the discussion under VARS). When using TABLE input, a CATEGORICAL analysis is assumed, and it is not necessary to specify this command.

CLASS

Syntax CLASS *variable list* ;
CLASS {ROW | COL} ;

Remarks The CLASS command specifies the class variable to be used in the analysis. This command is mandatory. If more than one variable is named, a separate analysis will be run for each variable. ROW and COL are used for TABLE input to indicate whether the rows or columns of the table are to be used for the class variable. Otherwise, the TO keyword may be used in the variable list (see the discussion under VARS).

DATA

Syntax DATA ;

Remarks The DATA command indicates that the entries that follow the command are data to be used in the current analysis. The END statement terminates the data block. For example, the following commands enter hypothetical data on two variables for each of two observations:

```
DATA;
1  2
3  4
END;
```

Each line should correspond to a single observation. The END (or END DATA) statement must be the only command on the line on which it appears.

DEGEN

Syntax DEGEN {ON | OFF} ;
DEGEN *variable list* ;

Alias DEGENERATE

Remarks DEGEN specifies whether degenerate cutpoints are allowed. If DEGEN OFF is specified, the resulting ODA solution must have at least one observation assigned to each predicted class. DEGEN allows flexibility in data sets which have small or no representation in some classes. The default is OFF. DEGEN with no parameters is the same as DEGEN ON. The TO keyword may be used in the variable list (see the discussion under VARS).

DIRECTION

Syntax DIRECTION { < | LT | > | GT | OFF} *value list* ;

Aliases DIR, DIRECTIONAL

Remarks The DIRECTION command specifies the presence and nature of a directional (i.e., an a priori or one-tailed) hypothesis. The parameter < or LT indicates that the class values in the value list are ordered in the "less-than" direction. The parameter > or GT indicates the class values are ordered in the "greater-than" direction. The value list must contain every value of the class variable currently defined. The default is OFF.

EXCLUDE

Syntax EXCLUDE *variable* {= | <> | < | > | <= | >= | OFF} *value* (,*value2*,. . .) . . . ;

Aliases EX, EXCL

Remarks This command excludes observations with the indicated *value* of *variable*. For example,

EXCLUDE C=3 ;

tells ODA to drop all observations with the value of 3 for variable C. Also, the command

EXCLUDE A=1 G>=902 ;

drops all observations with the values of 1 for variable A or values greater than or equal to 902 for variable G. Commas in the exclude string enable the user to exclude multiple values of a variable with a single command:

EXCLUDE B=1,3 ;

excludes all observations which have a value of 1 or 3 for variable B. Multiple EXCLUDE commands may be entered, up to a maximum of 100 clauses. The system will exclude observations that satisfy any of the EXCLUDE clauses. EXCLUDE is not allowed with TABLE input.

FREE

See VARS.

GO

Syntax GO ;

Remarks The GO command begins execution of the currently defined analysis.

GEN

Syntax GEN {OFF} *variable* ;
GEN TABLE *g* ;

Alias GROUP

Remarks The GEN command specifies the variable whose (integer) values indicate groups in a multisample (GEN) analysis. If TABLE has been specified, then GEN TABLE *g* indicates that *g* tables are present, corresponding to *g* GEN groups. The default is OFF.

HOLDOUT

Syntax HOLDOUT *path\file name* ;

Alias HOLD

Remarks HOLDOUT specifies the file name to be used for hold-out (validity) analysis. The variable list for the hold-out file must be in the same order as that for the main input file. OFF turns the hold-out indicator off.

ID

Syntax ID {OFF} *variable* ;

Remarks The ID command defines the ID variable that is to be printed in the long report. The default is OFF.

INCLUDE

Syntax INCLUDE *variable* {= | <> | < | > | <= | >= | OFF} *value* (,*value2*,. . .) . . . ;

Aliases IN, INCL

Remarks The INCLUDE command functions in the same way as the EXCLUDE command, except that ODA will keep only those observations with the indicated *value* for *variable*. If multiple INCLUDE statements exist, only those observations will be kept which satisfy all these INCLUDE statements. INCLUDE is not allowed with TABLE input.

LOO

Syntax LOO {ON | OFF} ;

Remarks The LOO command specifies that a leave-one-out (jackknife) analysis will be performed. LOO is not allowed in WEIGHTed CATEGORICAL problems. The default is OFF. LOO with no parameters is the same as LOO ON.

MCARLO

Syntax MCARLO {ITERATIONS *value* | SECONDS *value* | TARGET *value* | SIDAK *value* | STOP *value* | STOPUP *value* | ADJUST | OFF} ;

Alias MC

Remarks The MCARLO command controls Monte Carlo (simulation) analysis for estimating Type I error, or *p*. The keywords specify a number of stopping criteria; if any criterion is met, then the analysis stops. ITERATIONS (ITER) specifies the maximum number of Monte Carlo iterations, SECONDS (SEC) specifies the maximum number of seconds before the analysis terminates. TARGET specifies a target significance level. SIDAK adjusts the target to reflect a Sidak (Bonferroni) adjustment of the TARGET level, in which *value* is an integer that indicates the number of experiments involved in the adjustment (see chap. 3, this volume). STOP indicates the confidence level (in percent) that the estimated Type I error rate is *less* than the TARGET value, at which point the analysis stops. STOPUP indicates the confidence level (in percent) that the estimated Type I error rate is *greater* than the TARGET value, at which point the analysis stops. For example, the command

MCARLO ITER 1000 SEC 30 TARGET .01 STOP 99.9 STOPUP 99 ;

indicates that a Monte Carlo analysis will be conducted, and will stop when one of the following occurs: (1) 1,000 iterations have been executed, (2) 30 seconds have elapsed, (3) a confidence level of 99.9% has been obtained for $p < 0.01$, or (4) a confidence level of 99% has been obtained for $p > 0.01$. The default Monte Carlo method is conservative in the estimation of significance, in that a Monte Carlo iteration, whose optimal value is tied with the optimal value obtained from the original analysis, is always counted toward a higher significance level. Specifying ADJUST will adjust for this boundary by splitting these tied iterations in half. The default is OFF.

MISSING

Syntax MISSING {*variable list* | ALL} (*value*) ;

Alias MISS

Remarks The MISSING command tells ODA to treat observations with value (*value*) as missing for each variable on the list. For example, the command

MISSING A B C (–1) ;

indicates that observations with variables A, B, or C equal to –1 will be dropped if they are present in a CLASS, ATTRIBUTE, WEIGHT, GROUP, or ID variable. ALL specifies that the indicated missing value applies to all variables. The TO keyword may be used in the variable list (see the discussion under VARS).

OPEN

Syntax OPEN {*path\file name* | DATA} ;

Remarks The OPEN command specifies the data file to be processed by ODA. This file must be in ASCII format. DATA indicates that a DATA statement, with inline data following, appears in the command stream.

OUTPUT

Syntax OUTPUT *path\file name* {APPEND} ;

Remarks The OUTPUT command specifies the output file containing the results of the ODA run. The default is ODA.OUT. APPEND indicates that the report is to be appended to the end of an already existing output file.

PRIMARY

Syntax PRIMARY {MAXSENS | MEANSENS | SAMPLEREP | BALANCED | SENS *value* | DISTANCE | RANDOM | GENMEAN | GENSENS *value* | DEFAULT} ;

Alias PRI

Remarks The PRIMARY command specifies the primary criterion for choosing among multiple optimal solutions. MAXSENS (or MAXPAC) is maximum sensitivity. MEANSENS (MEANPAC) is the mean of the sensitivities of the separate classes. SAMPLEREP (SREP) selects the pattern of predicted class membership most closely resembling the sample class membership. BALANCED (BAL) selects the solution in which the sensitivity of the actual classes is most similar amongst each other. SENS (PAC) selects the solution with maximum sensitivity of class *value*. DISTANCE (DIST) selects the solution with smallest maximum (over all cutpoints) distance between the cutpoints and their boundaries. RANDOM (RAND) selects a randomly chosen solution. GENMEAN is used only when GEN is in effect. It selects the solution with maximum mean (weighted) sensitivity over all GEN groups. GENSENS selects the solution with the maximum (weighted) sensitivity of group *value*. The default is MAXSENS when PRIORS is ON and MEANSENS otherwise.

PRIORS

Syntax PRIORS {ON | OFF} ;

Remarks The PRIORS command indicates whether the ODA criterion will be weighted by the reciprocal of sample class membership. The default is ON. PRIORS with no parameters is the same as PRIORS ON.

QUIT

Syntax QUIT ;

Remarks Use the QUIT command to exit from ODA immediately.

REPORT

Syntax REPORT {SHORT | LONG} ;

Alias REP

Remarks The REPORT command specifies whether the short or long report is to be generated. The LONG report additionally prints the predicted and actual class memberships for each observation (ordered analysis) or for each cell (categorical analysis). The default is SHORT.

RESET

Syntax RESET ;

Remarks Use this command to reset all parameters to their default values.

SECONDARY

Syntax SECONDARY {MAXSENS | MEANSENS | SAMPLEREP | BALANCED | SENS *value* | DISTANCE | RANDOM | GENMEAN | GENSENS *value* | DEFAULT} ;

Alias SEC

Remarks The SECONDARY command specifies the secondary criterion for choosing among multiple optimal solutions. The default is SAMPLEREP. See the entry for PRIMARY for definitions of the above criteria.

SEED

Syntax SEED {*value* | TIME | 0} ;

Remarks The SEED command supplies the seed value for random number generation. TIME or 0 indicate that the current time will be used for the seed. If this command is not present, the time at program initiation will be used.

TABLE

Syntax TABLE *row* (*col*) ;

Alias FREE TABLE

Remarks The TABLE command is used for categorical analysis only, and indicates that a *row*-by-*col* table is present in the input file. If only *row* is entered, a square *row*-by-*row* table is assumed. For the sake of illustration, imagine that the following 2-by-2 table constitutes the data one wishes to analyze:

	Column 1	Column 2
Row 1	5	6
Row 2	7	8

In the ODA script for this illustration, the statement TABLE 2 would be used to indicate that the 2-by-2 table was to be input. CLASS ROW would indicate that the rows were to be considered the class variable. If the table is rectangular, the CLASS command should reflect the smaller value of *row* or *col*. For example, if

TABLE 4 3 ;

was entered, the user should then enter

CLASS COL ; .

When using TABLE input, a CATEGORICAL analysis is assumed, and it is not necessary to specify this command.

TITLE

Syntax TITLE *title* ;

Remarks The TITLE command specifies the title to be printed in the report. TITLE with no parameters erases the currently defined title.

VARS

Syntax VARS *variable list* ;

Alias FREE

Remarks The VARS command specifies a list of variable names corresponding to fields in the input data set. The TO keyword may be used to define multiple variables in the variable list. For example, the command

VARS A B C X1 TO X5 ;

specifies that the input file contains, in order, variables A, B, C, X1, X2, X3, X4, and X5, and that there is at least one blank space separating all adjacent data. Alternatively, the data points may be separated by a single comma or tab (with no spaces).

The TO keyword may only be used to input a range of variables that have the same name except for the integer at the end of the name. The integers must be positive and ascending, increasing one unit per variable. Thus, VAR1 TO VAR10 is admissible (defining 10 variables). In contrast, VAR10 TO VAR1, VARA TO VARJ, or A TO X10, are not admissible.

The data for each observation may all exist on a single line of the data set, or may be spread on multiple adjacent lines. It is not recommended that a new observation be included on a line that contains data from the previous observation.

WEIGHT

Syntax WEIGHT {*variable* | OFF} ;

Alias RETURN

Remarks The optional WEIGHT command specifies the weight variable for the analysis. The data values for the WEIGHT variable supply the weight for the corresponding observation. The default is OFF.