

MVA

```
MVA [VARIABLES=] {varlist}
      {ALL}

[/CATEGORICAL=varlist]

[/MAXCAT={25**}]
      {n}

[/ID=varname]
```

Description:

```
[/NOUNIVARIATE]

[/TTEST [PERCENT={5}] [{T} ] [{DF} ] [{PROB} ] [{COUNTS} ] [{MEANS} ]]]
      {n} {NOT} {NODF} {NOPROB} {NOCOUNTS} {NOMEANS}

[/CROSSTAB [PERCENT={5}]]
      {n}

[/MISMATCH [PERCENT={5}] [NOSORT]]
      {n}

[/DPATTERN [SORT=varname[({ASCENDING})] [varname ... ]]
      {DESCENDING}
      [DESCRIBE=varlist]]

[/MPATTERN [NOSORT] [DESCRIBE=varlist]]

[/TPATTERN [NOSORT] [DESCRIBE=varlist] [PERCENT={1}]]
      {n}
```

Estimation:

```
[/LISTWISE]

[/PAIRWISE]

[/EM [predicted_varlist] [WITH predictor_varlist]
      [( [TOLERANCE={0.001} ]
      {value}
      [CONVERGENCE={0.0001} ]
      {value}
      [ITERATIONS={25} ]
      {n}
      [TDF=n ]
      [LAMBDA=a ]
      [PROPORTION=b ]
      [OUTFILE='file' ]]]]
```

```

[/REGRESSION [predicted_varlist] [WITH predictor_varlist]
  [([TOLERANCE={0.001}
    {n}
    ]
  [FLIMIT={4.0}
    {N}
    ]
  [NPREDICTORS=number_of_predictor_variables]
  [ADDDTYPE={RESIDUAL*}
    {NORMAL}
    {T[({5})}
    {n}
    ]
    {NONE}
  ]
  [OUTFILE='file'
  ]
  ])]

```

* If the number of complete cases is less than half the number of cases, the default ADDTYPE specification is NORMAL.

** Default if the subcommand is omitted.

Examples:

```

MVA VARIABLES=populatn density urban religion lifeexpf region
  /CATEGORICAL=region
  /ID=country
  /MPATTERN DESCRIBE=region religion.

MVA VARIABLES=all
  /EM males msport WITH males msport gradrate facratio.

```

Overview

MVA (Missing Value Analysis) describes the missing value patterns in a data file (data matrix). It can estimate the means, the covariance matrix, and the correlation matrix by using listwise, pairwise, regression, and EM estimation methods. Missing values themselves can be estimated (imputed), and you can then save the new data file.

Options

Categorical variables. String variables are automatically defined as categorical. For a long string variable, only the first eight characters are used to define categories. Quantitative variables can be designated as categorical by using the CATEGORICAL subcommand.

MAXCAT specifies the maximum number of categories for any categorical variable. If any categorical variable has more than the specified number of distinct values, MVA is not executed.

Analyzing patterns. For each quantitative variable, the TTEST subcommand produces a series of *t* tests. Values of the quantitative variable are divided into two groups, based on the presence or absence of other variables. These pairs of groups are compared using the *t* test.

Crosstabulating categorical variables. The CROSSTAB subcommand produces a table for each categorical variable, showing, for each category, how many nonmissing values are in the other variables and the percentages of each type of missing value.

Displaying patterns. DPATTERN displays a case-by-case data pattern with codes for system-missing, user-missing, and extreme values. MPATTERN displays only the cases that have missing values and sorts by the pattern formed by missing values. TPATTERN tabulates the cases that have a common pattern of missing values. The pattern tables have sorting options. Also, descriptive variables can be specified.

Labeling cases. For pattern tables, an ID variable can be specified to label cases.

Suppression of rows. To shorten tables, the PERCENT keyword suppresses missing value patterns that occur relatively infrequently.

Statistics. Displays of univariate, listwise, and pairwise statistics are available.

Estimation. EM and REGRESSION use different algorithms to supply estimates of missing values, which are used in calculating estimates of the mean vector, the covariance matrix, and the correlation matrix of dependent variables. The estimates can be saved as replacements for missing values in a new data file.

Basic Specification

The basic specification depends on whether you want to describe the missing data pattern or estimate statistics. Often, description is done first, and then, considering the results, an estimation is done. Alternatively, both description and estimation can be done by using the same MVA command.

Descriptive analysis. A basic descriptive specification includes a list of variables and a statistics or pattern subcommand. For example, a list of variables and the subcommand DPATTERN would show missing value patterns for all cases with respect to the list of variables.

Estimation. A basic estimation specification includes a variable list and an estimation method. For example, if the EM method is specified, SPSS estimates the mean vector, the covariance matrix, and the correlation matrix of quantitative variables with missing values.

Syntax Rules

- A variables specification is required directly after the command name. The specification can be either a variable list or the keyword ALL.
- The CATEGORICAL, MAXCAT, and ID subcommands, if used, must be placed after the variables list and before any other subcommand. These three subcommands can be in any order.
- Any combination of description and estimation subcommands can be specified. For example, both the EM and REGRESSION subcommands can be specified in one MVA command.
- Univariate statistics are displayed unless the NOUNIVARIATE subcommand is specified. Thus, if only a list of variables is specified, with no description or estimation subcommands, univariate statistics are displayed.
- If a subcommand is specified more than once, only the last one is honored.
- The following words are reserved as keywords or internal commands in the MVA procedure: VARIABLES, SORT, NOSORT, DESCRIBE, and WITH. They cannot be used as variable names in MVA.
- The tables *Summary of Estimated Means* and *Summary of Estimated Standard Deviations* are produced if you specify more than one way to estimate means and standard deviations. The methods include univariate (default), listwise, pairwise, EM, and regression. For example, these tables are produced when you specify both LISTWISE and EM.

Symbols

The symbols displayed in the DPATTERN and MPATTERN table cells are:

- + Extremely high value
 - Extremely low value
 - S System-missing value
 - A First type of user-missing value
 - B Second type of user-missing value
 - C Third type of user-missing value
- An extremely high value is more than 1.5 times the interquartile range above the 75th percentile, if $(\text{number of variables}) \times n \log n \leq 150000$, where n is the number of cases.
 - An extremely low value is more than 1.5 times the interquartile range below the 25th percentile, if $(\text{number of variables}) \times n \log n \leq 150000$, where n is the number of cases.
 - For larger files—that is, $(\text{number of variables}) \times n \log n > 150000$ —extreme values are two standard deviations from the mean.

Missing Indicator Variables

For each variable in the VARIABLES list, a binary indicator variable is formed (internal to MVA), indicating whether a value is present or missing.

VARIABLES Subcommand

A list of variables or the keyword ALL is required.

- The order in which the variables are listed determines the default order in the output.
- The keyword VARIABLES is optional.
- If the keyword ALL is used, the default order is the order of variables in the working data file.
- String variables specified in the variable list, whether short or long, are automatically defined as categorical. For a long string variable, only the first eight characters of the values are used to distinguish categories.
- The list of variables must precede all other subcommands.
- Multiple lists of variables are not allowed.

CATEGORICAL Subcommand

The MVA procedure automatically treats all string variables in the variables list as categorical. You can designate numeric variables as categorical by listing them on the CATEGORICAL subcommand. If a variable is designated categorical, it will be ignored if listed as a dependent or independent variable on the REGRESSION or EM subcommand.

MAXCAT Subcommand

The MAXCAT subcommand sets the upper limit of the number of distinct values that each categorical variable in the analysis can have. The default is 25. This limit affects string variables in the variables list and also the categorical variables defined by the CATEGORICAL subcommand. A large number of categories can slow the analysis considerably. If any categorical variable violates this limit, MVA does not run.

Example

```
MVA VARIABLES=populatn density urban religion lifeexpf region
  /CATEGORICAL=region
  /MAXCAT=30
  /MPATTERN.
```

- The CATEGORICAL subcommand specifies that *region*, a numeric variable, is categorical. The variable *religion*, a string variable, is automatically categorical.
- The maximum number of categories in *region* or *religion* is 30. If either has more than 30 distinct values, MVA produces only a warning.
- Missing data patterns are shown for those cases that have at least one missing value in the specified variables.
- The summary table lists the number of missing and extreme values for each variable, including those with no missing values.

ID Subcommand

The ID subcommand specifies a variable to label cases. These labels appear in the patterns tables. Without this subcommand, the SPSS case numbers are used.

Example

```
MVA VARIABLES=populatn density urban religion lifeexpf region
  /CATEGORICAL=region
  /MAXCAT=20
  /ID=country
  /MPATTERN.
```

- The values of the variable *country* are used as case labels.
- Missing data patterns are shown for those cases that have at least one missing value in the specified variables.

NOUNIVARIATE Subcommand

By default, MVA computes univariate statistics for each variable—the number of cases with nonmissing values, the mean, the standard deviation, the number and percentage of missing values, and the counts of extreme low and high values. (Means, standard deviations, and extreme value counts are not reported for categorical variables.)

- To suppress the univariate statistics, specify NOUNIVARIATE.

Examples

```
MVA VARIABLES=populatn density urban religion lifeexpf region
/CATEGORICAL=region
/CROSSTAB PERCENT=0.
```

- Univariate statistics (number of cases, means, and standard deviations) are displayed for *populatn*, *density*, *urban*, and *lifeexpf*. Also, the number of cases, counts and percentages of missing values, and counts of extreme high and low values are displayed.
- The total number of cases and counts and percentages of missing values are displayed for *region* and *religion* (a string variable).
- Separate crosstabulations are displayed for *region* and *religion*.

```
MVA VARIABLES=populatn density urban religion lifeexpf region
/CATEGORICAL=region.
/NOUNIVARIATE
/CROSSTAB PERCENT=0.
```

- Only crosstabulations are displayed, no univariate statistics.

TTEST Subcommand

For each quantitative variable, a series of *t* tests is computed to test the difference of means between two groups defined by a missing indicator variable for each of the other variables (see “Missing Indicator Variables” on p. 4). For example, a *t* test is performed on *populatn* between two groups defined by whether their values are present or missing for *calories*. Another *t* test is performed on *populatn* for the two groups defined by whether their values for *density* are present or missing, and so on for the remainder of the variable list.

PERCENT=n *Omit indicator variables with less than the specified percentage of missing values.* You can specify a percentage from 0 to 100. The default is 5, indicating the omission of any variable with less than 5% missing values. If you specify 0, all rows are displayed.

Display of Statistics

The following statistics can be displayed for a *t* test:

- The ***t* statistic**, for comparing the means of two groups defined by whether the indicator variable is coded as missing or nonmissing (see “Missing Indicator Variables” on p. 4).

T *Display the t statistics.* This is the default.

NOT *Suppress the t statistics.*

- The **degrees of freedom** associated with the *t* statistic.

DF *Display the degrees of freedom.* This is the default.

NODF *Suppress the degrees of freedom.*

- The **probability** (two-tailed) associated with the t test, calculated for the variable tested without reference to other variables. Care should be taken when interpreting this probability.

PROB *Display probabilities.*

NOPROB *Suppress probabilities.* This is the default.

- The **number of values in each group**, where groups are defined by values coded as missing and present in the indicator variable.

COUNTS *Display counts.* This is the default.

NOCOUNTS *Suppress counts.*

- The **means** of the groups, where groups are defined by values coded as missing and present in the indicator variable.

MEANS *Display means.* This is the default.

NOMEANS *Suppress means.*

Example

```
MVA VARIABLES=populatn density urban religion lifeexpf region
  /CATEGORICAL=region
  /ID=country
  /TTEST.
```

- The TTEST subcommand produces a table of t tests. For each quantitative variable named in the variables list, a t test is performed, comparing the mean of the values for which the other variable is present against the mean of the values for which the other variable is missing.
- The table displays default statistics, including values of t , degrees of freedom, counts, and means.

CROSSTAB Subcommand

CROSSTAB produces a table for each categorical variable, showing the frequency and percentage of values present (nonmissing) and the percentage of missing values for each category as related to the other variables.

- No tables are produced if there are no categorical variables.
- Each categorical variable yields a table, whether it is a string variable assumed to be categorical or a numeric variable declared on the CATEGORICAL subcommand.
- The categories of the categorical variable define the columns of the table.
- Each of the remaining variables defines several rows—one each for the number of values present, the percentage of values present, and the percentage of system-missing values; and one each for the percentage of values defined as each discrete type of user-missing (if they are defined).

PERCENT=n *Omit rows for variables with less than the specified percentage of missing values.* You can specify a percentage from 0 to 100. The default is 5, indicating the omission of any variable with less than 5% missing values. If you specify 0, all rows are displayed.

Example

```
MVA VARIABLES=age income91 childs jazz folk
  /CATEGORICAL=jazz folk
  /CROSSTAB PERCENT=0.
```

- A table of univariate statistics is displayed by default.
- In the output are two crosstabulations, one for *jazz* and one for *folk*. The table for *jazz* displays, for each category of *jazz*, the number and percentage of present values for *age*, *income91*, *childs*, and *folk*. It also displays, for each category of *jazz*, the percentage of each type of missing value (system-missing and user-missing) in the other variables. The second crosstabulation shows similar counts and percentages for each category of *folk*.
- No rows are omitted, since PERCENT=0.

MISMATCH Subcommand

MISMATCH produces a matrix showing percentages of cases for a pair of variables in which one variable has a missing value and the other variable has a nonmissing value (a mismatch). The diagonal elements are percentages of missing values for a single variable, while the off-diagonal elements are the percentage of mismatch of the indicator variables (see “Missing Indicator Variables” on p. 4). Rows and columns are sorted on missing patterns.

PERCENT=n *Omit patterns involving less than the specified percentage of cases.* You can specify a percentage from 0 to 100. The default is 5, indicating the omission of any pattern found in less than 5% of the cases.

NOSORT *Suppress sorting of the rows and columns.* The order of the variables in the variables list is used. If ALL was used in the variables list, the order is that of the data file.

DPATTERN Subcommand

DPATTERN lists the missing values and extreme values for each case symbolically. For a list of the symbols used, see “Symbols” on p. 4.

By default, the cases are listed in the order in which they appear in the file. The following keywords are available:

SORT=varname [(order)] *Sort the cases according to the values of the named variables.* You can specify more than one variable for sorting. Each sort variable can be in ASCENDING or DESCENDING order. The default order is ASCENDING.

DESCRIBE=varlist *List values of each specified variable for each case.*

Example

```
MVA VARIABLES=populatn density urban religion lifeexpf region
  /CATEGORICAL=region
  /ID=country
  /DPATTERN DESCRIBE=region religion SORT=region.
```

- In the data pattern table, the variables form the columns, and each case, identified by its country, defines a row.
- Missing and extreme values are indicated in the table, and, for each row, the number missing and percentage of variables that have missing values are listed.
- The values of *region* and *religion* are listed at the end of the row for each case.
- The cases are sorted by *region* in ascending order.
- Univariate statistics are displayed.

MPATTERN Subcommand

The MPATTERN subcommand symbolically displays patterns of missing values for cases that have missing values. The variables form the columns. Each case that has any missing values in the specified variables forms a row. The rows are sorted by missing value patterns. For use of symbols, see “Symbols” on p. 4.

- The rows are sorted to minimize the differences between missing patterns of consecutive cases.
- The columns are also sorted according to missing patterns of the variables.

The following keywords are available:

NOSORT *Suppress the sorting of variables.* The order of the variables in the variables list is used. If ALL was used in the variables list, the order is that of the data file.

DESCRIBE=varlist *List values of each specified variable for each case.*

Example

```
MVA VARIABLES=populatn density urban religion lifeexpf region
  /CATEGORICAL=region
  /ID=country
  /MPATTERN DESCRIBE=region religion.
```

- A table of missing data patterns is produced.
- The *region* and the *religion* are named for each case listed.

TPATTERN Subcommand

The TPATTERN subcommand displays a tabulated patterns table, which lists the frequency of each missing value pattern. The variables in the variables list form the columns. Each pattern of missing values forms a row, and the frequency of the pattern is displayed.

- An X is used to indicate a missing value.
- The rows are sorted to minimize the differences between missing patterns of consecutive cases.
- The columns are sorted according to missing patterns of the variables.

The following keywords are available:

NOSORT	<i>Suppress the sorting of the columns.</i> The order of the variables in the variables list is used. If ALL was used in the variables list, the order is that of the data file.
DESCRIBE=varlist	<i>Display values of variables for each pattern.</i> Categories for each named categorical variable form columns in which the number of each pattern of missing values is tabulated. For quantitative variables, the mean value is listed for the cases having the pattern.
PERCENT=n	<i>Omit patterns that describe fewer than 1% of the cases.</i> You can specify a percentage from 0 to 100. The default is 1, indicating the omission of any pattern representing less than 1% of the total cases. If you specify 0, all patterns are displayed.

Example

```
MVA VARIABLES=populatn density urban religion lifeexpf region
  /CATEGORICAL=region
  /TPATTERN NOSORT DESCRIBE=populatn region.
```

- Missing value patterns are tabulated. Each row displays a missing value pattern and the number of cases having that pattern.
- DESCRIBE causes the mean value of *populatn* to be listed for each pattern. For the categories in *region*, the frequency distribution is given for the cases having the pattern in each row.

LISTWISE Subcommand

For each quantitative variable in the variables list, the LISTWISE subcommand computes the mean, the covariance between the variables, and the correlation between the variables. The cases used in the computations are listwise nonmissing; that is, they have no missing value in any variable listed in the VARIABLES subcommand.

Example

```
MVA VARIABLES=populatn density urban religion lifeexpf region
  /CATEGORICAL=region
  /LISTWISE.
```

- Means, covariances, and correlations are displayed for *populatn*, *density*, *urban*, and *lifeexpf*. Only cases that have values for all of these variables are used.

PAIRWISE Subcommand

For each pair of quantitative variables, the PAIRWISE subcommand computes the number of pairwise nonmissing values, the pairwise means, the pairwise standard deviations, the pairwise covariance, and the pairwise correlation matrices. These results are organized as matrices. The cases used are all cases having nonmissing values for the pair of variables for which each computation is done.

Example

```
MVA VARIABLES=populatn density urban religion lifeexpf region
  /CATEGORICAL=region
  /PAIRWISE.
```

- Frequencies, means, standard deviations, covariances, and the correlations are displayed for *populatn*, *density*, *urban*, and *lifeexpf*. Each calculation uses all cases that have values for both variables under consideration.

EM Subcommand

The EM subcommand uses an EM (expectation-maximization) algorithm to estimate the means, the covariances, and the Pearson correlations of quantitative variables. This is an iterative process, which uses two steps for each iteration. The E step computes expected values conditional on the observed data and the current estimates of the parameters. The M step calculates maximum likelihood estimates of the parameters based on values computed in the E step.

- If no variables are listed in the EM subcommand, estimates are performed for all quantitative variables in the variables list.
- If you want to limit the estimation to a subset of the variables in the list, specify a subset of quantitative variables to be estimated after the subcommand name EM. You can also list, after the keyword WITH, the quantitative variables to be used in estimating.
- The output includes tables of means, correlations, and covariances.
- The estimation, by default, assumes that the data are normally distributed. However, you can specify a multivariate *t* distribution with a specified number of degrees of freedom or a mixed normal distribution with any mixture proportion (PROPORTION) and any standard deviation ratio (LAMBDA).
- You can save a data file with the missing values filled in. You must specify a filename and its complete path in single or double quotation marks.
- Criteria keywords and OUTFILE specifications must be enclosed in a single pair of parentheses.

The criteria for the EM subcommand are as follows:

TOLERANCE=value *Numerical accuracy control.* The tolerance helps eliminate predictor variables that are highly correlated with other predictor variables and would reduce the accuracy of the matrix inversions involved in the calculations. The smaller the tolerance, the more inaccuracy is tolerated. The default value is 0.001.

CONVERGENCE=value *Convergence criterion.* Determines when iteration ceases. If the relative change in the likelihood function is less than this value, convergence is assumed. The value of this ratio must be between 0 and 1. The default value is 0.0001.

ITERATIONS=n *Maximum number of iterations.* Limits the number of iterations in the EM algorithm. Iteration stops after this many iterations even if the convergence criterion is not satisfied. The default value is 25.

Possible distribution assumptions:

TDF=n *Student's t distribution with n degrees of freedom.* The degrees of freedom must be specified if you use this keyword. The degrees of freedom must be an integer greater than or equal to 2.

LAMBDA=a *Ratio of standard deviations of a mixed normal distribution.* Any positive real number can be specified.

PROPORTION=b *Mixture proportion of two normal distributions.* Any real number between 0 and 1 can specify the mixture proportion of two normal distributions.

The following keyword produces a new data file:

OUTFILE='file' *Specify the name of the file to be saved.* Missing values for predicted variables in the file are filled in by using the EM algorithm. Specify the complete path in single or double quotation marks.

Examples

```
MVA VARIABLES=males to tuition
  /EM (OUTFILE='c:\colleges\emdata.sav').
```

- All variables on the variables list are included in the estimations.
- The output includes the means of the listed variables, a correlation matrix, and a covariance matrix.
- A new data file named *emdata.sav* with imputed values is saved in the *c:\colleges* directory.

```
MVA VARIABLES=all
  /EM males mspport WITH males mspport gradrate facratio.
```

- For *males* and *msport*, the output includes a vector of means, a correlation matrix, and a covariance matrix.
- The values in the tables are calculated using imputed values for *males* and *msport*. Existing observations for *males*, *msport*, *gradrate*, and *facratio* are used to impute the values that are used to estimate the means, correlations, and covariances.

```
MVA VARIABLES=males to tuition
  /EM verbal math WITH males mspport gradrate facratio
  (TDF=3 OUTFILE='c:\colleges\emdata.sav').
```

- The analysis uses a *t* distribution with three degrees of freedom.
- A new data file named *emdata.sav* with imputed values is saved in the *c:\colleges* directory.

REGRESSION Subcommand

The REGRESSION subcommand estimates missing values using multiple linear regression. It can add a random component to the regression estimate. Output includes estimates of means, a covariance matrix, and a correlation matrix of the variables specified as predicted.

- By default, all of the variables specified as predictors (after WITH) are used in the estimation, but you can limit the number of predictors (independent variables) by NPREDICTORS.
- Predicted and predictor variables, if specified, must be quantitative.
- By default, REGRESSION adds the observed residuals of a randomly selected complete case to the regression estimates. However, you can specify that the program add random normal, t , or no variates instead. The normal and t distributions are properly scaled, and the degrees of freedom can be specified for the t distribution.
- If the number of complete cases is less than half the total number of cases, the default ADDTYPE is NORMAL instead of RESIDUAL.
- You can save a data file with the missing values filled in. You must specify a filename and its complete path in single or double quotation marks.
- The criteria and OUTFILE specifications for the REGRESSION subcommand must be enclosed in a single pair of parentheses.

The criteria for the REGRESSION subcommand are as follows:

TOLERANCE=value	<i>Numerical accuracy control.</i> The tolerance helps eliminate predictor variables that are highly correlated with other predictor variables and would reduce the accuracy of the matrix inversions involved in the calculations. If a variable passes the tolerance criterion, it is eligible for inclusion. The smaller the tolerance, the more inaccuracy is tolerated. The default value is 0.001.
FLIMIT=n	<i>F-to-enter limit.</i> The minimum value of the F statistic that a variable must achieve in order to enter the regression estimation. You may want to change this limit, depending on the number of variables and the correlation structure of the data. The default value is 4.
NPREDICTORS=n	<i>Maximum number of predictor variables.</i> This specification limits the total number of predictors in the analysis. The analysis uses the stepwise selected n best predictors, entered in accordance with the tolerance. If $n = 0$, it is equivalent to replacing each variable with its mean.

ADDDTYPE *Type of distribution from which the error term is randomly drawn. Random errors can be added to the regression estimates before the means, correlations, and covariances are calculated. You can specify one of the following types:*

RESIDUAL. Error terms are chosen randomly from the observed residuals of complete cases to be added to the regression estimates.

NORMAL. Error terms are randomly drawn from a distribution with the expected value 0 and the standard deviation equal to the square root of the mean squared error term (sometimes called the **root mean squared error**, or RMSE) of the regression.

T(n). Error terms are randomly drawn from the $t(n)$ distribution and scaled by the RMSE. The degrees of freedom can be specified in parentheses. If T is specified without a value, the default degrees of freedom is 5.

NONE. Estimates are made from the regression model with no error term added.

The following keyword produces a new data file:

OUTFILE *Specify the name of the new data file to be saved. Missing values for the dependent variables in the file are imputed (filled in) by using the regression algorithm. Specify the complete path in single or double quotation marks.*

Examples

```
MVA VARIABLES=males to tuition
  /REGRESSION (OUTFILE='c:\colleges\regdata.sav').
```

- All variables in the variables list are included in the estimations.
- The output includes the means of the listed variables, a correlation matrix, and a covariance matrix.
- A new data file named *regdata.sav* with imputed values is saved in the *c:\colleges* directory.

```
MVA VARIABLES=males to tuition
  /REGRESSION males verbal math WITH males verbal math faculty
  (ADDDTYPE = T(7)).
```

- The output includes the means of the listed variables, a correlation matrix, and a covariance matrix.
- A t distribution with 7 degrees of freedom is used to produce the randomly assigned additions to the estimates.