

CROSSTABS

```
CROSSTABS [TABLES=] {varlist} BY varlist [BY...]  
           {ALL  
           [/{TABLES=} varlist...]  
  
           [/OPTIONS=option numbers]  
  
           [/STATISTICS={statistic numbers}]  
           {ALL
```

Options:

- | | |
|---|---|
| 1 Include user-missing values | 13 Suppress cell counts |
| 2 Suppress all labels | 14 Display expected frequencies |
| 3 Display row percentages | 15 Display chi-square residuals |
| 4 Display column percentages | 16 Display standardized chi-square residuals |
| 5 Display two-way table total percentages | 17 Display adjusted standardized chi-square residuals |
| 6 Suppress value labels | 18 Display all cell information value |
| 8 Order rows by descending value | 19 Suppress values |
| 12 Suppress tables | |

Statistics:

- | | |
|--|-------------------|
| 1 Chi-square | 6 Kendall's tau-b |
| 2 Phi for 2 × 2 tables, Cramér's V for larger tables | 7 Kendall's tau-c |
| 3 Contingency coefficient | 8 Gamma |
| 4 Lambda | 9 Somers' d |
| 5 Uncertainty coefficient | 10 Eta |
| | 11 Pearson's r |

Example:

```
CROSSTABS TABLES=WVAR XVAR BY YVAR BY ZVAR /TABLES=AVAR BY BVAR  
/OPTIONS=3 4 14 15  
/STATISTICS=1.
```

- Overview** Procedure CROSSTABS (alias XTABS) produces tables showing the joint distribution of two or more variables that have a limited number of distinct values. The frequency distribution of one variable is subdivided according to the values of one or more variables. The unique combination of values for two or more variables defines a cell, the basic element of all tables. To analyze contingency tables using hierarchical log-linear models, see HILOGLINEAR.
- Defaults** By default, CROSSTABS produces two-way to *n*-way crosstabulations for variables that have a limited number of numeric or string values. The default table consists of cell counts and is labeled with the names of the variables, the values of the variables, and their variable and value labels if these have been defined. Values of the row variable are displayed in ascending order. Cases that have user-missing or system-missing values for the variables that define a table are omitted from that table and the number of missing cases is reported.
- Tailoring**
- Display Format.** Display format options include ordering the table rows in descending value order, suppressing display of variable and value labels, suppressing display of values, and suppressing display of the table itself.
- Cell Contents.** You can include row, column, and total percentages, expected chi-square frequencies, and chi-square residuals in the cells of the table.
- Statistical Display.** Measures of association and significance tests are available for each subtable.
- Missing Values.** You can include user-missing values in the tables and in the calculation of statistics.

Syntax

- The minimum specification is a single TABLES subcommand with a tables list. The actual keyword TABLES may be omitted.
- The minimum tables list specifies a list of row variables, the keyword BY, and a list of column variables.
- Subcommands are separated by slashes.
- Subcommands can be specified in any order.
- You can invoke procedure CROSSTABS with the XTABS command.

Operations

- CROSSTABS is a procedure and causes the data to be read.
- If a long string variable is used, only the short-string portion is tabulated.
- Statistics are calculated separately for each two-way table or two-way subtable. Missing values are reported for the table as a whole.
- If you request only percentages and cell counts, percent signs are not displayed for percentages, and zero values for counts and percentages are represented by blanks.
- If you request percentages and any expected values or residuals, percent signs and zeros are displayed.
- Value labels for column variables are displayed in two lines with eight characters per line. Row value labels are displayed with 16 characters on one line.
- Percentages, expected values, and residuals are displayed with one decimal place.
- The display uses the width defined on the SET command.
- The BOXSTRING subcommand on SET controls the characters used in the table display.

Limitations

- The number of variables allowed per CROSSTABS command is the same as the system limit.
- Maximum 250 nonempty rows or columns are displayed for each variable.
- Maximum 20 TABLES subcommands lists per CROSSTABS command.
- Maximum 1 each OPTIONS and STATISTICS subcommands.
- Maximum 10 dimensions (9 BY keywords) per tables list.
- Maximum 250 value labels are displayed on any single table.

Example

```
CROSSTABS TABLES=WVAR XVAR BY YVAR BY ZVAR/TABLES=AVAR BY BVAR  
/OPTIONS=3 4 14 15  
/STATISTICS=1.
```

- The first tables list generates bivariate subtables of WVAR by YVAR for each value of ZVAR, followed by subtables of XVAR by YVAR for each value of ZVAR.
- The second tables list produces a single table in which AVAR is the row variable and BVAR is the column variable.
- The OPTIONS subcommand requests row and column percentages, expected cell frequencies, and residuals.
- The STATISTICS subcommand requests the chi-square.

Example

```
XTABS TABLES=JOB CAT BY EDCAT BY SEX BY INCOME3.
```

- Assuming that SEX has values 1 and 2 and INCOME3 has values 1, 2, and 3, this tables list will produce a subtable of JOB CAT by EDCAT first for value 1 of SEX and value 1 of INCOME3, then for value 2 of SEX and value 1 of INCOME3, then for value 1 of SEX and value 2 of INCOME3, and so forth.
- This example uses the CROSSTABS alias, XTABS.

TABLES Subcommand

The TABLES subcommand specifies the tables lists. The actual keyword TABLES may be omitted.

- Variables named before the first BY in a tables list are row variables, and variables named after the first BY in a tables list are column variables.
- Variables named after the second (or subsequent) BY are control variables.
- Each subsequent use of the keyword BY in a tables list adds a new dimension to the tables requested and introduces a new order of control among the independent variables.
- You can name more than one variable in each dimension.
- You can use keyword ALL to include all user-defined variables in a dimension.
- When the tables list specifies two dimensions, a table is produced for the first variable before BY with each variable after BY, then the second variable before BY with each variable after BY, and so forth.
- When the tables list specifies more than two dimensions, a two-way subtable is produced for each combination of values of control variables.
- When the tables list specifies more than three dimensions, the value of the last variable mentioned changes the most slowly in determining the order in which the tables are displayed.
- You can specify multiple TABLES subcommands on a single CROSSTABS command. The slash between the subcommands is required; the keyword TABLES is not.

Display Format

By default, CROSSTABS displays values, variable labels, and value labels and orders rows from the lowest to the highest value. You can change these defaults by specifying the following on the OPTIONS subcommand:

- Option 2** *Suppress all labels.* Neither variable nor value labels are displayed.
- Option 6** *Suppress value labels.* Only variable labels are displayed on each table.
- Option 8** *Order rows by descending value.*
- Option 19** *Suppress values.* Values are displayed only if no value labels have been defined.

Cell Contents

By default, CROSSTABS table cells contain only counts. You can display additional information in the cells of each bivariate table or subtable. Optionally, you can control cell contents by specifying the following on the OPTIONS subcommand:

- Option 3** *Display row percentages.* The cell count as a percentage of cases in the row is included in the cell.
- Option 4** *Display column percentages.* The cell count as a percentage of cases in the column is included in the cell.
- Option 5** *Display two-way table total percentages.* The cell count as a percentage of all cases in the subtable is included in the cell.
- Option 13** *Suppress cell counts.* Unless you specify Option 3, 4, 5, 14, 15, 16, or 17, nothing is displayed. If you request both Options 13 and 18, Option 13 is ignored.
- Option 14** *Display expected frequencies.* The expected cell count if the two variables in the subtable were statistically independent is included in the cell.
- Option 15** *Display chi-square residuals.* The observed cell count minus the expected value is included in the cell.
- Option 16** *Display standardized chi-square residuals.* The standardized residual is included in the cell. (See Haberman, 1978.)

- Option 17** *Display adjusted standardized chi-square residuals.* The adjusted standardized residual is included in the cell. (See Haberman, 1978.)
- Option 18** *Display all cell information.* Includes cell count; row, column, and total percentages; expected value; residual; standardized residual; and adjusted standardized residual. If you request both Options 13 and 18 -Option 13 is ignored.

Statistical Display

By default, only the cell counts, marginal percentages, and number of missing cases are displayed for each table. You can request additional statistics by specifying the following on the STATISTICS subcommand:

- Statistic 1** *Chi-square.* Fisher's exact test is computed using the rounded values of the cell entries when there are fewer than 20 cases in a 2×2 table that does not result from missing rows or columns in a larger table; Yates' corrected chi-square is computed for all other 2×2 tables.
- Statistic 2** *Phi for 2×2 tables, Cramér's V for larger tables.*
- Statistic 3** *Contingency coefficient.*
- Statistic 4** *Lambda.*
- Statistic 5** *Uncertainty coefficient.*
- Statistic 6** *Kendall's tau-b.*
- Statistic 7** *Kendall's tau-c.*
- Statistic 8** *Gamma.* Zero-order gammas are displayed for 2-way tables and conditional gammas are displayed for the 2-way subtables of 3-way to 10-way tables.
- Statistic 9** *Somers' d.*
- Statistic 10** *Eta.* Ignored for string variables.
- Statistic 11** *Pearson's r.* Ignored for string variables.
- ALL** *Display all statistics.*

Missing Values

By default, a case missing on any of the variables that define a table is not used either in the table display or in the calculation of the statistics. You can change the handling of cases with missing values by specifying the following on the OPTIONS subcommand:

- Option 1** *Include cases with user-missing values.* Cases with user-missing values are displayed in tables and are included in the calculation of statistics.

References

Haberman, S. J. 1978. *Analysis of Qualitative Data*, Vol. 1. London: Academic Press.

DATA LIST: Fixed Format

DATA LIST [FILE='filename'] [FIXED]

```
/varlist columns {{{(0)}}} [varlist columns ...]  
                {(n)}  
                {(A)}
```

```
[/ ...] [/ ...]
```

Format Meaning

(n) Implied decimal places
for numeric variables
(A) String variable

Example:

```
DATA LIST /ID 1-3 SEX 5 (A) AGE 7-8 OPINION1 TO OPINION5 10-14.
```

Overview

The DATA LIST command assigns names to variables and provides information about the column location and format. Data can be inline (entered with SPSS/PC commands) or stored in an external file. The DATA LIST command with keyword FIXED defines data arranged in fixed format. In fixed format, the values for each variable are found in the same location on the same record for each case. Fixed format is the default for DATA LIST.

See DATA LIST: Freefield Format for defining data organized in freefield format, and DATA LIST: Matrix Materials for defining matrix materials.

Defaults

By default, SPSS/PC assumes that data are inline, entered interactively or within an INCLUDE file. By default, all variables are assumed to be numeric without implied decimal places.

Tailoring

You can use data from an external file. You can also define string variables and specify implied decimal places for numeric variables.

Syntax

- The minimum DATA LIST specification for fixed format is a slash followed by at least one variable name and its location.
- The keyword FIXED is optional.

Operations

- Variable names are stored in the active file dictionary.
- The order of the variables in the active file dictionary is the order in which they are defined on the DATA LIST command, not their sequence on the input data file. This order is important if you later use the TO convention.
- By default, variables are assumed to be numeric. Alphabetical and special characters, except the decimal point and leading plus and minus signs, are not valid numeric values and are set to system-missing if encountered in the data.
- Blanks to the left or right of a number in the default format are ignored; embedded blanks are invalid.
- The system-missing value is assigned to a completely blank field for numeric variables. The value assigned to blanks can be changed using the BLANKS specification on the SET command.
- Formats are stored in the active file dictionary and are used to print and write out the values. Use the FORMATS command to change formats of numeric variables defined on DATA LIST.

Limitations

- You cannot define more than 200 variables on a DATA LIST command.
- Each 8-character portion of a long string variable counts toward the 200-variable limit. For example, an 18-character long string counts as three short string variables (see Universals: Strings).
- The maximum number of format tokens is 600.

Example

```
DATA LIST /ID 1-3 SEX 5 (A) AGE 7-8 OPINION1 TO OPINIONS 10-14.
BEGIN DATA.
001 m 28 12212
002 f 29 21212
003 f 45 32145
lines of data
128 m 17 11194
END DATA.
```

- The data are assumed to be inline because no data file is specified.
- The data are in fixed format (the default).
- Variable definitions start with ID in columns 1 through 3.
- Variable SEX is a short string variable in column 5.
- AGE is a two-column variable in columns 7 and 8.
- Variables OPINION1, OPINION2, OPINION3, OPINION4, and OPINIONS are named using the TO convention (see Universals: Variable-Naming Conventions). Each is a one-column variable, with OPINION1 located in column 10 and OPINIONS located in column 14.
- The BEGIN DATA and END DATA commands enclose the inline data. Note that the values of SEX are in lower-case characters and must be specified as such on subsequent commands.

FILE Subcommand

- The FILE subcommand is required when data are contained in an external data file. It is not required when the data are contained in a file included with an INCLUDE command or when the data are inline (see INCLUDE and BEGIN DATA).
- The file specification must be enclosed in apostrophes.
- The file specification can be fully qualified, including directory, filename, and extension.
- The FILE subcommand can be specified before or after the optional keyword FIXED.
- The FILE subcommand and keyword FIXED must be separated by at least one blank or comma.

Example

```
DATA LIST FIXED FILE=' \INVENTORY\MARCH.DAT'
/NUTS3 1-2 BOLTS35 3-5 NAILS3P 6-10.
```

- Keyword FIXED indicates that the variables to be defined are on a fixed-format data file.
- The FILE specification directs SPSS/PC to read data from the file MARCH.DAT in directory INVENTORY.
- Three numeric variables, NUTS3, BOLTS35, and NAILS3P, are defined for use in subsequent SPSS/PC procedures.

Record Specification

- Records are indicated on the DATA LIST command by a slash, followed by the variables to be defined from that record.
- The first slash indicates the first (or only) record.
- The second and any subsequent slashes tell SPSS/PC to skip to a new record.
- You must specify a slash for every record, even if no variables are being defined from that record.
- Variables from each record can be named in any order, regardless of their sequence on the data file.
- All variables to be read from one record must be defined before proceeding to the next record.

Example

```
DATA LIST FILE='SOCSUR82.DAT'
/ ID 1-7 SEX 15 AGE 16-18
/ ANOMIA 15 LIKEPOL 17
// OPIN1 76 OPIN2 77
/.
```

- The DATA LIST command defines data in fixed format from file SOCSUR82.DAT in the current directory.
- Three variables, ID, SEX, and AGE, are defined from the first record.
- ANOMIA and LIKEPOL are defined from the second record.
- The third data record for each case is skipped; no variables are defined.
- The fourth record contains two variables, OPIN1 and OPIN2.
- The fifth record is skipped.
- The data file contains a total of five records per case. The DATA LIST command defines seven variables from three of these records.

Variable Names

- Variable names can contain up to eight characters.
- All variable names must begin with a letter. System variables (beginning with a \$) cannot be defined on DATA LIST.
- You can name a list of variables using the TO convention. For more information on the TO convention and other variable-naming rules, see Universals: Variable-Naming Conventions.

Variable Locations

- Each variable name is followed by its column location.
- If the variable is one column wide, specify the number of the column. If the variable is two or more columns wide, specify the number of the first column followed by a dash (-) and the number of the last column.
- The same column locations can be used to define different variables.
- If several variables are recorded in adjacent columns on the same record and have the same width and format type, you can use an abbreviated format for specifying column location. First list all variable names, then the beginning column location of the first variable in the list, a dash, and the ending column location of the last variable in the list. SPSS/PC divides the total number of columns specified equally among the variables. If the number of columns do not divide equally, an error message is issued.

Example

```
DATA LIST FILE='AGES.DAT' / BIRTHDA 1-2 BIRTHMO 3-4 BIRTHYR 5-8
  BIRTHDAY 1-8 PRSNT1 TO PRSNT5 11-15 CELEBRAT 65-68
  CAKES 50-52 / WISHES 10-11.
```

- The DATA LIST command defines variables from the fixed-format file AGES.DAT in the current directory.
- Three variables, BIRTHDA, BIRTHMO, and BIRTHYR, are read from the first eight columns on the first record. Variable BIRTHDAY is also read from the first eight columns as one variable.
- Variables PRSNT1, PRSNT2, PRSNT3, PRSNT4, and PRSNT5 are defined using the TO convention. Each of these variables is one column wide.
- Variable CELEBRAT is read from columns 65 through 68, and next CAKES is read from columns 50 through 52. The SPSS/PC active file dictionary will contain these variable names in the order they are defined on DATA LIST, even though this order differs from their order in the data file.
- Variable WISHES is read from columns 10 and 11 on the second data record.

Example

```
DATA LIST / LINENUM 1 ID 2-6 V1 TO V7 7-13
  OPINREL OPINSEX OPINDRUG OPINRAR 15-18
  LOCATN76 TO LOCATN83 20-35.
```

- The DATA LIST command defines inline data in fixed format.
- Variables V1, V2, V3, V4, V5, V6, and V7 are named using the TO convention. Each variable is one column wide.
- Four opinion variables, OPINREL, OPINSEX, OPINDRUG, and OPINRAR, are defined separately in columns 15 through 18. Each of these variables is one column wide.
- Seven location variables are defined by the LOCATN76 TO LOCATN83 specification. Each of these variables is two columns wide.

Variable Formats

- In the default format, variables are assumed to be numeric, either signed or unsigned integer or real numbers.
- String (alphanumeric) variables are indicated with an A in parentheses following the column specification.
- You can specify numeric and string formats for different variables on a single DATA LIST command.
- If a value is encountered which cannot be read according to the format type specified, it is assigned the system-missing value and a warning message is issued.

Numeric Formats

- When a decimal point is not actually coded in real data, the number of implied decimal places can be indicated in parentheses following the column specification.
- A coded decimal point in the data overrides the number of implied decimal places indicated on the DATA LIST command.
- The table below compares how values are interpreted for a four-column numeric variable when no decimal places are defined on DATA LIST and when two decimal places are defined.

Values in the data file	Default	Two defined decimal places
2001	2001	20.01
201	201	2.01
-201	-201	-2.01
2	2	.02
20	20	.20
2.2	2.2	2.2
.201	.201	.201
2 01	Undefined	Undefined

String Formats

- The values of string (alphanumeric) variables can contain any number, letter, or character, including special characters and embedded blanks. For further discussion of string variables, see *Universals: Strings*.
- String variables whose values contain eight characters or less are called *short string variables*.
- String variables with values longer than eight characters and up to 255 characters are called *long string variables*.

Example

```
DATA LIST FILE=' \SPSSDAT\FILEX.DAT'  
/XVAR 1 YVAR 10-15 ZVAR 3-9(2) // AVAR 25-30(A) BVAR 31-45(A).
```

- The data are defined from file FILEX.DAT in directory \SPSSDAT and are arranged in fixed format.
- Numeric variable XVAR is found in column 1 of the first record.
- The next variable defined is YVAR, found in columns 10 through 15.
- Variable ZVAR, found in columns 3 through 9, contains two implied decimal places, indicated by (2).
- No variables are defined on the second record for each case.
- Two variables are defined from the third record for each case.
- AVAR is a six-column short string variable.
- BVAR is a long string variable read from columns 31 through 45. BVAR counts as two variables toward the 200-variable system limit.

DATA LIST: Freefield Format

```
DATA LIST [FILE='filename'] FREE
```

```
    /variable [( {A } )] varlist  
              {Aw}
```

Format Meaning

(Aw) String of width w

Example:

```
DATA LIST FILE='MYFILE.DAT' FREE/XVAR YVAR.
```

Overview

The DATA LIST command assigns names to variables and provides information about their formats. The data can be inline or read from an external file. The DATA LIST command with keyword FREE identifies data arranged in freefield format. In freefield format, all variables are recorded in the same order for each case but not necessarily in the same column locations. Each value in the data file is separated by one or more blanks or by one comma.

See DATA LIST: Fixed Format for information on defining data in fixed format, and DATA LIST: Matrix Materials to define matrix materials.

Defaults

By default, the data are assumed to be inline, entered interactively or within an INCLUDE file. All data values are assumed to be numeric.

Tailoring

You can use data stored on an external file. You can also define string variables.

Syntax

- The minimum specification for freefield data on DATA LIST is the keyword FREE, a slash, and at least one variable name.
- Variables must be named in the order they are entered on the data file.
- There is no record or column specification for freefield format.

Operations

- FREE can read freefield-format data with multiple cases recorded on one record or with one case recorded on more than one record.
- Variable names are stored in the active file dictionary.
- In the default format, variables are assumed to be numeric. Alphabetical and special characters, except the decimal point and leading plus and minus signs, are not valid numeric values and are set to the system-missing value.
- You can use BASIC conventions of delimiting data values. Two commas together or two commas separated by a blank indicate either numeric system-missing or a string blank.

Limitations

- You cannot define more than 200 variables on a DATA LIST command.
- Each eight-character portion of a long string variable counts toward the 200-variable system limit. For example, an eleven-character string variable counts as two short string variables (see Universals: Strings).
- The maximum number of format tokens is 600.

Example

```
DATA LIST FREE/XVAR YVAR.  
BEGIN DATA.  
1 3 2 15 3 16 4  
156  
5 22 6 -3  
END DATA.
```

- The DATA LIST command indicates inline data in freefield format.
- Two variables, XVAR and YVAR, are named.
- The values for the first case are 1 for variable XVAR and 3 for variable YVAR. The second case has values of 2 and 15 for XVAR and YVAR. The third case has values 3 and 16, the fourth case has values 4 and 156, and so on. Note that

in freefield format, a single line of data can include values for more than one case. Also, the values for one case can be split across lines, as for the fourth case. Individual values cannot be split across lines.

FILE Subcommand

- The FILE subcommand is required when data are contained in an external file. It is not required when the data are included using an INCLUDE command or when the data are inline (see INCLUDE and BEGIN DATA).
- The file specification must be enclosed in apostrophes.
- The file specification can be fully qualified, including directory, filename, and extension.
- The FILE subcommand can be specified before or after keyword FREE.
- The FILE subcommand and keyword FREE must be separated by at least one blank or comma.

Example DATA LIST FILE='MYFILE.DAT' FREE, XVAR YVAR.

- The freefield-format data file is read from MYFILE.DAT in the current directory.
- Two numeric variables are defined.

Variable Names

- Variable names can contain up to eight characters.
- All variable names must begin with a letter. System variables (beginning with \$) cannot be defined on DATA LIST.
- Each variable name corresponds to one value per case.
- You can name a list of variables using the TO convention. For more information on the TO convention and other variable-naming rules, see Universals: Variable-Naming Conventions.

Example DATA LIST FREE / ID VAR1 TO VAR7.

- The DATA LIST command indicates inline data in freefield format.
- Eight variables are defined: ID, VAR1, VAR2, VAR3, VAR4, VAR5, VAR6, and VAR7.

Variable Formats

- In DATA LIST with keyword FREE, formats can be specified for string variables only.
- All numeric variables are automatically assigned print and write formats of F8.2. Use the FORMATS command to specify any other format for numeric variables.
- All numeric variable digits are read and stored by SPSS/PC.
- Numeric values with decimal points in the data preserve the decimal point and decimal digits.
- String variables are indicated by an A in parentheses after the variable name.
- By default, all string variables are assigned formats of A8 (width of eight characters). You can change the format of a string variable by specifying A and a width enclosed in parentheses, such as (A20).
- A format specification applies only to the variable immediately preceding it.
- If the string value in the data is longer than the specified length, the string is truncated and a warning message is printed. Thus, you must specify formats for long string variables.
- If the string in the data is shorter than the specified format, it is right-padded with blanks and no warning message is printed. All subsequent transformations require value specifications with the declared format width, including all padded values.

Example DATA LIST FREE FILE='MASTER\APRIL.DAT'
/ID SEX (A1) NAME (A15) AGE TENURE ETHNIC (A).
FORMATS AGE (F2.0) TENURE (F3.1).

- The DATA LIST command defines data in freefield format from file APRIL.DAT in directory \MASTER.
- Six variables, ID, SEX, NAME, AGE, TENURE, and ETHNIC, are defined.
- ID, AGE, and TENURE are numeric variables.
- SEX is defined as a one-column string variable. NAME is defined as a long string variable up to 15 columns wide. ETHNIC is a short string variable with an assumed width of eight columns.
- The FORMATS command changes the print and write formats of numeric variables AGE and TENURE.

Entering Freefield Data

- Values are read sequentially in the order variables are named.
- A value cannot be split across records.
- One data value is separated from another by any number of blanks or by a single comma.
- You can use both commas and blanks to distinguish data values in a single data file.
- Any number of consecutive blanks (except blanks specified within a string value) are interpreted as one delimiter.
- Multiple blank columns at the end of a record are interpreted as one delimiter.
- When commas are used as delimiters, two consecutive commas or commas separated by a blank indicate a system-missing numeric value or a blank string value.
- A blank field for a variable that is not delimited by commas causes values from that point on to be assigned to the wrong variable.
- String values which contain embedded blanks or commas must be delimited by apostrophes or quotation marks. The delimiters are not read as part of the string value (see Universals: Strings).
- You can include an apostrophe in string values by delimiting the value with quotation marks. You can include quotation marks by delimiting the value with apostrophes.
- You cannot use commas or blanks within numeric values.
- If there are not enough values to complete the last case, a warning is issued and the incomplete case is dropped.

Example

```
DATA LIST FREE/AVAR BVAR STATE (A) MAYOR(A10).
BEGIN DATA.
7500000 20000000 'S DAKOTA' "O'LEARY" 22222000 55000000
'INDIANA' 'JONES' 120000000 56000000 'NEW YORK' 'ALDRIDGE'
-1.2 2222.223 'MAINE' 'BURNS'
END DATA.
FORMATS AVAR (COMMA15.2) BVAR(DOLLAR14.2).
DISPLAY VAR=ALL.
FREQ VAR=ALL.
LIST VAR=ALL.
```

- The DATA LIST command defines inline data in freefield format.
- AVAR and BVAR are defined as numeric variables. STATE is defined as a string variable with the default width of eight columns. MAYOR is declared as a long string variable with a width of 10 columns (A10).
- The BEGIN DATA command indicates the beginning of data lines.
- The first case has a value of 7500000 for AVAR. BVAR has the value 20000000. STATE has the value "S DAKOTA" enclosed in apostrophes to preserve the embedded blank. The value "O'LEARY" is enclosed in quotation marks to preserve the embedded apostrophe. The format for MAYOR indicates a width of 10 columns, so O'LEARY is right-padded with three blanks.
- The second case is split across two records. With freefield format, you can split cases but not individual values across records.
- Decimal values are included for AVAR and BVAR for the fourth case.

- The END DATA command indicates the end of inline data.
- The FORMATS command changes the print and write formats of numeric variables AVAR and BVAR from the default format (eight characters with two decimal places) to the specified formats (see FORMATS).
- The DISPLAY command shows the current print and write formats of the variables.
- The FREQUENCIES procedure produces tables for each of the variables. Because FREQUENCIES uses the internal representation of values in tables, the DOLLAR and COMMA formats do not appear on these tables.
- The LIST procedure produces a listing of the values of each variable. The DOLLAR and COMMA formats are preserved in the listing.

Example

```
DATA LIST FREE/AVAR BVAR STATE (A) MAYOR(A10).
BEGIN DATA.
7500000,20000000,'S DAKOTA','O'LEARY",.22222000.,
'INDIANA','JONES',120000000,,,'ALDRIDGE'
-1.2 2222.223 'MAINE' 'BURNS'
END DATA.
```

- This example shows both commas and blanks used as delimiters.
- The two commas at the end of the first record indicate missing information for BVAR for the second case.
- In the second line of data, the three commas after value 120000000 indicate missing values for BVAR and STATE for the third case. A blank between the commas would also indicate missing information.
- The blanks after ALDRIDGE indicate the end of the value. The last line of data uses blanks as delimiters.

DATA LIST: Matrix Materials

```
DATA LIST [FILE='filename'] MATRIX { {FIXED} }  
                                         {FREE }  
                                         /varlist
```

Example:

```
DATA LIST FILE='REG.MAT' MATRIX AGE SUICIDE ANOMIE.  
N 488.  
REG VAR=AGE SUICIDE ANOMIE/READ CORR  
/DEP=SUICIDE/METHOD=ENTER.
```

Overview

The DATA LIST command with keyword MATRIX provides variable names and a dictionary for matrix materials used as input in CLUSTER, FACTOR, ONEWAY, and REGRESSION. The matrix materials can include correlation coefficients, covariance coefficients, a matrix of n 's, or group distance measures. Matrix materials can be read in fixed or freefield format but must conform to the requirements of the individual procedures (see each procedure for details). The matrix input can be inline or read from an external file.

For information on reading individual casewise data, see DATA LIST: Fixed Format or DATA LIST: Freefield Format.

Syntax

- The minimum specification is DATA LIST with keyword MATRIX, followed by a slash and a list of variable names.
- The slash (/) between keyword MATRIX and the variable names is required.
- Variable names must be eight characters or less and must begin with a letter.
- Format types are meaningless for matrix materials and cannot be specified.
- The FILE subcommand is required when the matrix materials are contained in an external file. The file specification must be enclosed in apostrophes.
- The FILE subcommand can be specified before or after keyword MATRIX.
- The FILE subcommand and keyword MATRIX must be separated by at least one blank or comma.
- You can specify files in directories other than the current directory by using fully qualified file specifications.

Operations

- DATA LIST with keyword MATRIX processes a variety of matrix materials arranged for specific procedures.
- Each procedure can read the matrix materials that it writes.
- DATA LIST with keyword MATRIX cannot read individual casewise data.
- You cannot use DATA LIST MATRIX with procedures that expect casewise data.
- You cannot perform any SPSS/PC transformations on matrix materials.
- The BASIC convention of using two commas in a row to indicate missing data is *not* supported with DATA LIST MATRIX.

Limitations

- You cannot assign more than 200 variables on the DATA LIST command.
- The maximum number of format tokens is 600.

Example

```
DATA LIST FILE='REG.MAT' MATRIX/ AGE SUICIDE ANOMIE.  
N 488.  
REG VAR=AGE SUICIDE ANOMIE/READ CORR/DEP=SUICIDE/METHOD=ENTER.
```

- DATA LIST reads matrix materials from file REG.MAT in the current directory. The matrix was written to this file by procedure CORRELATION in a previous SPSS/PC session.
- The matrix materials are in fixed format (the default). CORRELATION writes matrices in the same format as does REGRESSION.

- The variable names AGE, SUICIDE, and ANOMIE are assigned to the rows and columns of the correlation matrix.
- The N command indicates that the matrix input is based on 488 cases.
- The READ subcommand on REGRESSION indicates that a correlation matrix will be read.
- This example takes advantage of spelling permitted by three-character truncation of keywords.

Example

```
DATA LIST FREE MATRIX/AGE SUICIDE ANOMIE.
BEGIN DATA.
1.0 .5555555 .3333333
.5555555 1.0 .4555555
.3333333 .4555555 1.0
488
END DATA.
REGRESSION VARIABLES=AGE SUICIDE ANOMIE
/READ CORR N
/DEPENDENT=SUICIDE
/METHOD=ENTER.
```

- The DATA LIST command specifies matrix materials in freefield format.
- Because no file is specified on DATA LIST, the matrix data are assumed to be inline.
- The variable names AGE, SUICIDE, and ANOMIE are assigned to the rows and columns of the correlation matrix.
- The matrix materials are entered between the BEGIN DATA and END DATA commands. Each row vector begins on a new line, followed by a single number indicating the number of cases.
- The READ subcommand on REGRESSION indicates that a correlation matrix followed by a matrix of *n*'s (number of cases) will be read.

Matrix Data

The SPSS/PC procedures CORRELATION, CLUSTER, ONEWAY, FACTOR, and REGRESSION write matrix materials in a fixed format that automatically conforms to the requirements of the various procedures that read matrix materials. If you enter your own matrix materials, they must conform to these formats as well as to the requirements below.

- Matrix materials can be arranged in fixed or freefield format.
- Each cell of the matrix must contain a value.
- Each element in a row in freefield format matrix materials is separated by at least one space or a comma.
- Each row vector of a matrix begins on a new line.
- Each type of matrix material begins on a new line.
- In fixed format, there is a maximum number of elements that can be entered in a row (see discussion of individual procedures below).
- If the elements for a vector do not fit in one row, the elements can be continued on the next row. Each row must be filled before continuing to the next.
- Matrix elements cannot be split across input lines.
- Decimal points must be entered with the data. You cannot specify implied decimal places.

Matrix Input for Procedure CLUSTER

- Procedure CLUSTER reads matrix materials in both fixed and freefield format.
- Fixed-format matrix materials for CLUSTER must be arranged so that each matrix cell is 16 columns wide with up to 5 decimal places. You can have only 5 elements of a vector in each row.

- Freefield format matrix materials, such as a correlation matrix written by procedure CORRELATION, must conform to the requirements listed under "Matrix Data."

Example

```
DATA LIST MATRIX FREE /
  ABDEFECT ABHLTH ABNOMORE ABPOOR ABRAPE ABSINGLE.
BEGIN DATA.
1.0000000 .6118418 .3936668 .3743177 .6284106 .3820830
.6118418 1.0000000 .2870408 .3098805 .6097969 .2935045
.3936668 .2870408 1.0000000 .7658386 .3806726 .7881280
.3743177 .3098805 .7658386 1.0000000 .3847740 .7379326
.6284106 .6097969 .3806726 .3847740 1.0000000 .3909586
.3820830 .2935045 .7881280 .7379326 .3909586 1.0000000
END DATA.
CLUSTER ABDEFECT ABHLTH ABNOMORE ABPOOR ABRAPE ABSINGLE/
  READ=SIMILAR.
```

- The DATA LIST command specifies inline matrix materials in freefield format. The active file dictionary contains six variable names.
- In this case, the matrix is a correlation matrix. Each row vector begins on a new line.
- The READ subcommand on CLUSTER indicates that a square matrix based on a measure of similarity will be read (see CLUSTER).

Example

```
DATA LIST MATRIX CASE1 TO CASE19
BEGIN DATA
0.0 19062.00391 17697.00781 17545.00781 19039.00781
19050.00781 17742.00781 17954.00781 19125.00391 111.99998
19230.00391 18041.00781 17693.00781 19023.00781 9635.00781
9860.00781 9899.00781 9901.00781 10028.00781
19062.00391 0.0 18485.00000 17867.00000 17904.00000
17890.00000 18484.00000 18569.00000 49.00000 18536.00391
109.99998 18537.00000 18191.00000 18009.00000 8427.00000
9228.00000 9657.00000 9865.00000 9726.00000

10028.00781 9726.00000 9929.00000 9859.00000 8502.00000
9435.99609 9589.99609 9717.99609 7527.00000 7739.00781
9476.00000 9660.99609 9449.00000 9261.00000 87.00000
46.00000 40.99998 57.99998
```

END DATA.
CLUSTER ALL MISSING INCLUDE READ.

- The DATA LIST commands specifies inline matrix materials in fixed format (the default). The active file dictionary contains 19 variable names.
- The distance matrix to be read was produced using procedure CLUSTER with the MISSING=INCLUDE and WRITE=DISTANCE subcommands. Here, cases, not variables, are going to be clustered.
- The data are automatically arranged in the format required by the CLUSTER. Each column vector is 16 characters wide with 5 decimal values. There are 4 rows for each vector. Only the first two and last vectors are shown.
- The READ subcommand on CLUSTER indicates that a distance matrix will be read (see CLUSTER).

Matrix Input for Procedure FACTOR

- Procedure FACTOR can use matrix materials in either fixed or freefield format.
- Fixed-format matrix materials for FACTOR must be arranged so that each column vector entry is 10 columns wide with up to 3 decimal places. You can enter up to 8 values in each row.
- Matrix materials in freefield format must conform to requirements noted above under "Matrix Data."

Example

```

DATA LIST MATRIX / X1 X2 X3 X4 X5.
N 100.
BEGIN DATA.
1.000
0.945      1.000
0.840      0.720      1.000
0.735      0.630      0.560      1.000
0.630      0.540      0.480      0.420      1.000
END DATA.
FACTOR READ=CORRELATION TRIANGLE/
VARIABLES=X1 TO X5/
ANALYSIS=X1 TO X5/
PRINT=ALL/
CRI=FAC(1)/
EXT=ULS.

```

- The DATA LIST command specifies inline matrix materials in fixed format (the default). The active file dictionary contains five variable names.
- The N command tells SPSS/PC that the matrix input is based on 100 cases.
- The matrix data conform to the fixed-format requirements of FACTOR. Each row vector starts on a new line and each column entry occupies 10 columns.
- The READ subcommand on FACTOR indicates that a lower-triangular correlation matrix will be read (see FACTOR: Matrix Materials).
- This example takes advantage of spelling permitted by three-character truncation of keywords.

Example

```

DATA LIST FREE MATRIX /X1 TO X5.
N 100.
BEGIN DATA.
1.000
0.945,1.000
0.840,0.720,1.000
0.735,0.630,0.560,1.000
0.630,0.540,0.480,0.420,1.000
END DATA.
FACTOR READ=CORRELATION TRIANGLE/
VARIABLES=X1 TO X5/
ANALYSIS=X1 TO X5/
PRINT=ALL/
CRI=FAC(1)/
EXT=ULS.

```

- The DATA LIST command specifies inline matrix materials in freefield format. The active file dictionary contains five variable names (X1, X2, X3, X4, and X5).
- The correlation matrix is entered with each coefficient separated by a comma.
- The READ subcommand on FACTOR indicates that a lower-triangular correlation matrix will be read (see FACTOR: Matrix Materials).

Matrix Input for Procedure ONEWAY

- Procedure ONEWAY reads matrix materials in either fixed or freefield format.
- If you use matrix materials in fixed format, you must specify Option 7 on the ONEWAY command.
- Each matrix cell entry has a width of 10 columns with up to 4 decimal places. You can enter up to 8 cells in each row.

Example

```

DATA LIST MATRIX / SCORE METHOD.
BEGIN DATA.
7          7
4.4286    7.5714    6.7143
1.2724    1.3973    .9512
END DATA.
ONEWAY SCORE BY METHOD(1,3)
/OPTION 7.

```

- The DATA LIST command specifies inline matrix materials in the default fixed format. The active file dictionary contains two variable names, SCORE and METHOD.
- The data are arranged with each vector element occupying 10 columns. The vector of counts does not require decimal places.
- Option 7 on ONEWAY indicates that a matrix with a vector of counts, a vector of means, and a vector of standard deviations will be read (see ONEWAY).

Example DATA LIST FREE MATRIX/YVAR XVAR.
 BEGIN DATA.
 65 95 181 82 40 37
 2.6462 2.7737 4.1796 4.5610 4.6625 5.2297
 6.2699
 494
 END DATA.
 ONEWAY VARIABLES=YVAR BY XVAR(1,6)
 /OPTIONS=8.

- The DATA LIST command specifies inline matrix materials in freefield format. Two variable names, YVAR and XVAR, are in the active file dictionary.
- The data are arranged to conform to the requirements of Option 8 in ONEWAY, with one row of counts, a row of means, an entry for the pooled variance estimate, and an entry for the degrees of freedom.
- Option 8 on the ONEWAY command indicates that matrix materials arranged as a vector of counts, a vector of means, the pooled variance estimate (a single entry), and the degrees of freedom (a single entry) will be read (see ONEWAY).

**Matrix Input for
 Procedure
 REGRESSION**

- Procedure REGRESSION reads matrix materials in either fixed or freefield format.
- Materials arranged in fixed format must have 10 columns for each vector entry with up to 7 decimal places. You can enter up to 8 entries per line for each vector.
- Materials arranged in freefield format must conform to the requirements described above under "Matrix Data."

Example DATA LIST MATRIX/AGE SUICIDE ANOMIE.
 BEGIN DATA.
 1.0 0.555555 0.333333
 0.555555 1.0 0.455555
 0.333333 0.455555 1.0
 488
 END DATA.
 REGRESSION VARIABLES=AGE SUICIDE ANOMIE
 /READ CORR N
 /DEPENDENT=SUICIDE
 /METHOD=ENTER.

- The DATA LIST command defines inline matrix materials in fixed format (the default). Three variable names are defined for the active file dictionary.
- The data are arranged to conform to the requirements of the REGRESSION procedure. Each row vector has 3 entries, each with a width of 10 columns. The last entry is the number of cases.
- The READ subcommand on REGRESSION reads the matrix of correlation coefficients and a value for the number of cases (see REGRESSION: Matrix Materials).

Example DATA LIST MATRIX FREE / X1 X2 X3 X4 X5 Y.
 BEGIN DATA.
 35.0825 2.7315 40.9060 3.1405 25.0690 6.2550
 5.8171 .4541 25.8985 9.6254 1.3138 .6543
 33.8381250 .5079382 113.502300 51.9130250 2.5499079 2.7898921
 .5079382 .2062029 2.1302800 1.0036676 .2998858 .0584658
 113.502300 2.1302800 670.734846 206.202997 1.7372853 15.7105368
 51.9130250 1.0036676 206.202997 92.6479839 2.3183584 5.1585079
 2.5499079 .2998858 1.7372853 2.3183584 1.7260832 .1064316
 2.7898921 .0584658 15.7105368 3.1585079 .1064316 .4281316
 20
 END DATA.
 VAR LABELS X1 'STAFF SALARIES PER PUPIL'
 X2 '6TH GRADE PER CENT WHITE-COLLAR FATHERS'
 X3 'SES COMPOSITE'
 X4 'MEAN TEACHER VERBAL TEST SCORE'
 X5 '6TH GRADE MEAN MOTHER EDUCATION'
 Y 'VERBAL MEAN TEST SCORE, ALL 6TH GRADERS'.
 REGRESSION READ=COV MEAN STDDEV N/DES DEF/
 VAR=Y,X1 TO X5/CRI TOL(.0001)/
 STATS ALL/DEP Y/ENT/.

- The **DATA LIST** command specifies inline matrix materials in freefield format. Six variable names are specified for the active file dictionary.
- The data are arranged to conform to the requirements of the **REGRESSION** procedure. The first six rows form a covariance matrix, with each row starting on a new line. The seventh row forms a vector of means for each variable named. The eighth row is a vector of standard deviations. The last entry is the number of cases.
- The **READ** subcommand on **REGRESSION** indicates that matrix materials with a covariance matrix, a vector of means, a vector of standard deviations, and an entry for the number of cases will be read (see **REGRESSION: Matrix Materials**).