***Categorical into Contrast***

SPSS macros by Kirill Orlov

kior@akado.ru, ttnphns@gmail.com

<https://www.spsstools.net/en/KO-spssmacros>

All rights reserved

*Categorical variables into contrast variables.* Creates contrast variables from categorical variables, of several types, and their interaction variables. Contrast variables are needed first of all when one has to analyze influence of qualitative factors by methods designed for quantitative input (e.g., linear regression).

*Read “*[*About SPSS macros*](https://www.spsstools.net/en/KO-aboutmacros)*” what are they and how to run them.*

*The “Protected directory” error.* Some of the macros described in the current document write temporary files to hard disc. If you don't have full Administrator rights of your computer, it may cause error saying, among things: *“SPSS Statistics cannot access a file... specifies a protected directory...”*, meaning that the default directory the macro wants to use is protected on your PC. To solve the problem, in Syntax window issue command: CD 'myfolder'., where 'myfolder' is the path/name of some folder where you are allowed to save files to.

**MACRO !KO\_CATCONT: RECODING CATEGORICAL VARIABLES INTO CONTRAST VARIABLES**

Version 4, Feb 2022 (Version 1, Oct 2002). Tested on SPSS Statistics 20, 22, 26.

!KO\_catcont factors= *v1 v2* /\*One or more categorical variables (factors) name-by-name

/types= DEV SIM /\*For each factor, contrast type: IND, SIM, DEV, WDEV, HEL, WHEL,

/\*DIF, WDIF, REP, POL, WPOL, { }, file

/spacing= /\*For POL WPOL, spacing: EQUAL (default) or ASIS

/inter= ALL /\*If interactions are needed: integer or UP integer or ALL

/\*e.g. 3 means "all 3-way", UP 3 means "all up to 3-way"

/seq= EFFECT /\*Sequence among interaction variables: combinatorial (COMBIN, default)

/\*or by effect (EFFECT)

/id= *id*  /\*Optional: case identifier variable

/print= /\*Printing: full (LONG, default), short (SHORT), minimal (NONE)

/savelmx= *'d:\exercise\lmx.sav'* /\*Output constructed L matrix for factors and (if

/\*requested) interactions: file path/name.

Minimal specification FACTORS, TYPES.

What’s new July 2024.

Weighted types WDEV, WHEL, WDIF added.

The macro creates, out of categorical or discrete quantitative variables, contrast variables of the needed type and, optionally, their interactions. It also returns matrices of contrast coefficients (**L** matrices). If a factor consists of *k* levels, *k*-1 contrast variables are formed from it. The output dataset containing the contrast variables is unnamed. Use Data > Merge Files > Add Variables to unite it with the input dataset.

Contrast variables are numeric, quantitative variables into which there arises need to recode between-subject factors (categorical variables) in order the factors can mathematically play a role of predictors in various linear or nonlinear prediction models. Out of a factor, a set of contrast variables is created, the variables together replacing (substituting) the between-subject factor. Famous dummy variables are one type of contrast variables.

Each contrast variable is the embodiment of a specific *contrast*. Contrast is a weighted linear combination of levels of a factor. Due to the weights summing to 0, a contrast is an instrument of comparison between selected separate levels or mixtures thereof, in a factor. And polynomial contrasts – treating a factor as discrete quantitative one – is an instrument of examining linear and nonlinear dependencies from such factor. The said weights in a contrast are called contrast coefficients, and they are directly tied with the values of the contrast variable embodying the contrast.

SPSS procedures such as General Linear Model, Logistic Regression and many other taking between-subject factors, offer to choose among different types of contrasts for them (these types are also known as categorical encoding schemes). From the type you choose there depends the assigning, by the program, of these or those contrast coefficients, and through this – the *meaning* of the parameters in the prediction model depends. However, the mentioned flexible procedures in SPSS do not present to the user the respective contrast variables as the data: everything takes place behind the scene. Macro !KO\_CATCONT has the purpose to give the user contrast variables themselves as data, physically. Having contrast variables on hands one can use them as continuous regressors (covariates) in such procedures that are not suited for categorical predictors (factors), recognizing only quantitative predictors. For example, this way one can obtain in Linear Regression procedure some results – assuming factors – that are available in One-way Analysis of Variance or in General Linear Model. You may use one and the same contrast variables in models with or without intercept, in models with or without covariates.

Limitations. Up to 26 input factors. Up to 36 levels in a factor. Up to 4-way interactions. If s/c SAVELMX is specified, then up to 4 input factors and order of interactions not above 3.

Naming of output variables

Contrast variable’s name consists of a symbol and an index. The symbol designates the factor of which the variable was made, and the index is the ordinal number of the contrast variable from this factor. Symbols used for factors (in the order of their listing in FACTORS): *a*, *b*, ..., *z*. Indices used: *1*, *2*, ..., *9*, *0*, *A*, *B*, ..., *Z*. Thus, the first contrast variable for the first factor will be named *a1*, the second for the same factor will be *a2*, and so on. Those same indices used in contrast variables names are also used to enumerate a factor’s levels (by ascending order of codes in a factor). This is convenient, because contrast variables of multinomial types (see s/c TYPE) correspond to levels. With polynomial types, contrast variables correspond not to levels but to exponential components, for which indices *1*, *2*, ..., again, are convenient.

Contrast variables of interactions inherit this naming system, being the cohesion. *a1b3*, for instance, is the 2-way interaction “a1 by b3”, or “a1\*b3”.

Factor variables. Besides contrast variables, the output dataset contains the input factors themselves, only as string variables now and in the above notation. The factors bear names *a*, *b*, ..., and their levels (values) are indices *1*, *2*, ..., etc. Take attention that such values will be always, whatever way the input factor variables were coded.

DATA used in the examples.

id y y2 f1 f2 f3 f4 g1 g2 g3 g4 covar1

1 3 0 1 10 1 1 1 10 1 1 3

2 4 0 1 10 1 2 1 10 1 1 5

3 6 0 1 10 1 2 1 10 3 1 8

4 6 0 1 10 2 1 1 10 1 1 4

5 3 0 1 10 2 1 1 10 1 2 1

6 7 1 1 10 2 1 1 10 1 2 4

7 6 0 1 10 2 2 1 10 2 1 1

8 3 0 1 10 2 2 1 10 2 1 4

9 7 1 1 10 3 1 1 10 2 2 8

10 5 0 1 10 3 1 1 10 2 2 5

11 6 0 1 10 3 1 1 10 2 2 4

12 6 0 1 10 3 2 1 10 3 1 3

13 6 0 1 10 3 2 1 10 3 2 3

14 3 0 1 10 3 2 1 10 3 2 4

15 6 0 1 10 4 1 1 25 1 1 1

16 6 0 1 10 4 1 1 10 3 1 4

17 7 1 1 10 4 2 1 25 1 2 7

18 5 0 1 10 4 2 1 25 1 1 2

19 7 1 1 25 1 1 1 25 1 2 6

20 4 0 1 25 2 1 1 25 1 2 6

21 6 0 1 25 2 1 1 25 2 1 5

22 4 0 1 25 2 2 1 25 2 1 8

23 6 0 1 25 2 2 1 25 2 1 3

24 6 0 1 25 2 2 1 25 2 2 4

25 6 0 1 25 3 1 1 25 2 2 5

26 7 1 1 25 3 1 1 25 3 1 8

27 7 1 1 25 3 1 1 25 3 2 6

28 6 0 1 25 3 1 1 25 3 2 4

29 8 1 1 25 3 1 1 25 3 1 5

30 8 1 1 25 3 2 1 60 1 1 2

31 6 0 1 25 3 2 1 60 1 2 4

32 6 0 1 25 3 2 1 60 1 2 3

33 6 0 1 25 3 2 1 60 1 1 5

34 6 0 1 25 3 2 1 25 3 2 3

35 6 0 1 25 4 1 1 60 2 1 3

36 4 0 1 25 4 1 1 60 2 1 4

37 7 1 1 25 4 1 1 60 2 2 7

38 6 0 1 25 4 2 1 60 1 1 7

39 5 0 1 25 4 2 1 60 2 2 8

40 6 0 1 60 1 1 1 60 2 2 3

41 5 0 1 60 1 1 1 60 3 1 6

42 4 0 1 60 1 2 1 60 3 1 5

43 5 0 1 60 2 1 2 10 1 1 7

44 7 1 1 60 2 2 1 60 3 2 6

45 6 0 1 60 2 2 1 60 3 2 5

46 4 0 1 60 3 1 2 10 1 1 4

47 6 0 1 60 3 2 2 10 1 2 5

48 1 0 1 60 3 2 2 10 1 2 5

49 7 1 1 60 4 1 2 10 1 2 6

50 10 1 1 60 4 1 1 60 3 1 5

51 5 0 1 60 4 2 2 10 2 1 3

52 5 0 2 10 1 1 2 25 1 1 7

53 5 0 2 10 1 1 2 10 2 1 0

54 9 1 2 10 1 2 2 10 2 1 7

55 5 0 2 10 1 2 2 10 3 1 4

56 4 0 2 10 2 1 2 10 2 2 6

57 7 1 2 10 2 1 2 10 2 2 7

58 9 1 2 10 2 1 2 10 3 1 3

59 6 0 2 10 2 2 2 25 1 1 3

60 5 0 2 10 2 2 2 10 3 2 4

61 7 1 2 10 3 1 2 25 1 1 7

62 5 0 2 10 3 1 2 10 3 2 4

63 4 0 2 10 3 2 2 25 1 2 9

64 9 1 2 10 3 2 2 10 3 2 10

65 6 0 2 10 4 1 2 25 1 2 6

66 7 1 2 10 4 2 2 25 2 1 4

67 5 0 2 25 1 1 2 25 2 1 5

68 3 0 2 25 1 2 2 25 2 2 7

69 7 1 2 25 2 1 2 25 2 2 7

70 7 1 2 25 2 1 2 25 2 2 1

71 9 1 2 25 2 2 2 25 3 1 8

72 6 0 2 25 3 1 2 25 3 1 6

73 4 0 2 25 3 1 2 25 3 2 5

74 6 0 2 25 3 1 2 25 3 1 4

75 7 1 2 25 3 2 2 60 1 1 7

76 7 1 2 25 3 2 2 60 1 1 4

77 3 0 2 25 3 2 2 60 1 2 3

78 7 1 2 25 3 2 2 25 3 2 2

79 5 0 2 25 4 1 2 60 1 2 7

80 6 0 2 25 4 1 2 60 2 2 1

81 3 0 2 25 4 2 2 60 2 1 4

82 8 1 2 25 4 2 2 60 2 1 3

83 7 1 2 60 1 1 2 60 2 1 3

84 5 0 2 60 1 1 2 60 3 2 4

85 6 0 2 60 1 2 2 60 2 2 1

86 8 1 2 60 2 2 2 60 3 1 1

87 7 1 2 60 3 2 2 60 3 1 2

88 4 0 2 60 4 1 2 60 3 2 6

89 5 0 2 60 4 2 2 60 1 2 4

90 7 1 2 60 4 2 2 60 3 2 6

* F1, F2, F3, F4 are factors (unbalanced; and if all with F4, then with some empty cells in the design). G1, G2, G3, G4 is one more company of factors (balanced, if without G4).
* Factors F2 and G2 are coded with quantitative discrete values, not by arbitrary codes. This may play a role only with SPACING=ASIS.
* Y is a quantitative dependent variable. Y2 is a dichotomous dependent variable. COVAR1 is a quantitative covariate. Dependent variables or covariates are not needed for the macro’s job; they are only used in some of the examples in the present document.

EXAMPLE 1.

!KO\_catcont factors= f1 f2 f3 /types= DEV DEV 'd:\exercise\f3\_lmx.sav'.

* Factor F1 is requested to encode into a set of contrast variables of type DEV, factor F2 – in type DEV too. Factor F3 is requested to encode into contrast variables corresponding to a user contrast matrix, the latter specified as an external file.
* F1 (2 levels) will yield contrast variable a1. F2 (3 levels) will yield contrast variables b1, b2. F3 (4 levels) will yield contrast variables c1, c2, c3. Contrast variables are 1 less than there are levels in a factor. In each factor, the last level is being taken for reference one. The meaning of type DEV contrast variables please find in s/c TYPES.
* Contrast variables are output as new unnamed dataset. The three input factors are copied there too, under names a, b, c and in string format; their codes are set by the macro (these are always indices 1, 2, ..., 9, 0, A, B, ..., Z).

***Subcommands***

**FACTORS**

Name-by-name numeric variables (factors), minimum one, from which contrast variables should be produced. Each unique valid value in a variable counts as a factor level. Values are discrete, although don’t have to be integer. If SPACING=ASIS, they must be all positive. Names can repeat in the list. The order of levels in a factor corresponds to the ascending sort order of values in it.

Missing values are deleted by the macro listwise, i.e., from all the factors (plus ID variable) at one.

**TYPES**

List in correspondence to FACTORS: for each factor, indicate the type of contrast into which you want to recode (encode) that factor. You may also indicate one type and word ALL following it – then this type will be applied to all the factors. Cannot use word ALL if the type is user-defined.

To a factor of *k* levels, *k*-1 contrasts correspond, and to each contrast there corresponds own contrast variable that outputs. So, a set of *k*-1 numeric variables will be created out of each factor. Table of values of the contrast variables (*k* levels × *k*-1 contrasts) is called **C matrix** (= coding matrix, basis matrix, summarized design matrix). To it, the **L matrix** (contrast matrix, matrix of contrast coefficients) sized *k*-1 contrasts × *k* levels, corresponds. The macro displays both matrices under PRINT=LONG; and **L** matrix is shown transposed.

Mathematical relation between the matrices: **L+ =** **C+-1** and **C+ = L+-1**, where **C+** is **C** with the 1st column of units attached, and **L+** is **L** with the 1st row of 1/*k* attached. It is possible also to obtain one matrix from the other via pseudoinversion, but this is true not for all contrast types.

Number of contrasts is by 1 less the number of values in a factor for the multicollinearity reason: determinant of SSCP matrix (i.e., matrix **X´X**) of the set of contrast variables **X** would become 0 if the number of created contrast variables were *k*, not *k*-1[[1]](#footnote-1).

Examples of **C** and **L** matrix. Let there be factor *A* with three levels (*1*, *2*, *3*).

Coding matrix **C** and contrast matrix **L** for Indicator (dummy) type:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Contrast variables | |  |  | Levels | | |
| Levels | a1 | a2 |  | Contrasts | *A=1* | *A=2* | *A=3* (ref.) |
| *A=1* | 1 | 0 |  | a1 | 1 | 0 | -1 |
| *A=2* | 0 | 1 |  | a2 | 0 | 1 | -1 |
| *A=3* (ref.) | 0 | 0 |  |  |  |  |  |

Coding matrix **C** and contrast matrix **L** for Deviation type:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Contrast variables | |  |  | Levels | | |
| Levels | a1 | a2 |  | Contrasts | *A=1* | *A=2* | *A=3* (ref.) |
| *A=1* | 1 | 0 |  | a1 | 0.6666 | -0.3333 | -0.3333 |
| *A=2* | 0 | 1 |  | a2 | -0.3333 | 0.6666 | -0.3333 |
| *A=3* (ref.) | -1 | -1 |  |  |  |  |  |

In **L** matrix, the meaning of each contrast (contrast variable) is seen; specifically, what will regression parameter by the given contrast variable *tell*, if we inter all the set of contrast variables, obtained from a factor, as multiple quantitative predictors (covariates). Coefficients in each row of **L** matrix sum up to 0, and “contrasting” positive vs negative coefficients show which levels are being compared with which, by the given contrast (and hence, by the corresponding contrast variable). Contrast coefficients are the weights in a linear combination of levels or values of the factor: nonzero weights show what is being confronted with what. Examination of contrast coefficients does *not* allow to predict the result – the magnitude of the comparison (= the value of the regression parameter), but allows to understand its *content*.

Types of contrasts and the corresponding, to each type, meaning of regression parameters by the contrast variables – are expressed below.

a) **Multinomial types**. Factor is taken for categorical variable. Each of *k*-1 contrasts corresponds to one level of the factor, comparing it with a specific another level or a mix of levels.

IND - **Indicator** type. This is the so-called dummy variables; they are binary (1 vs 0). Regression parameter at *i*-th contrast variable will be the comparison of prediction function between: *i*-th level and *k*-th level. Intercept of the regression will be prediction function in *k*-th level. *k*-th level is reference, no contrast variable for it, and its parameter logically is 0. Contrasts of this type are nonorthogonal; under balanced factor the contrast variables correlate with r = -1/(*k*-1). Type IND should be interpreted somewhat differently from what is said here when there are interfactor interactions in the model (see EXAMPLE 8).

SIM - **Simple** type. **C** matrix of this type coincides with **C** matrix of type IND after centering columns in the latter. Their **L** matrices are identical. SIM is equivalent to IND and carries out those same comparisons, being different from IND only by intercept: intercept of the regression will be prediction function unweightedly averaged between all levels. The contrasts are nonorthogonal; under balanced factor the contrast variables correlate with r = -1/(*k*-1). SIM is equivalent to IND, however, only in the absence of interfactor interactions in the model. See EXAMPLE 8, explaining difference between SIM and IND in the presence of such interactions. When there are interactions, in typical case one should use SIM, not IND.

DEV - **Deviation** aka Effect type. These are trinary variables (1, 0, -1; and with *k*=2 only 1, -1). Regression parameter at *i*-th contrast variable will be the comparison of prediction function between: *i*-th level and unweighted averaging between all levels. Intercept of the regression will be that prediction function unweightedly averaged between all levels (as in SIM). *k*-th level is reference, no contrast variable for it, and its parameter is equal to the negated sum of the other *k*-1 parameters. Contrasts of this type are nonorthogonal; under balanced factor the contrast variables correlate with r = 0.5. This type frequently comes as default one for between-subject factors in ANOVA.

WDEV - **weighted Deviation** type. This version of the previous considers the fact of possible unbalancedness. Regression parameter at *i*-th contrast variable will be the comparison of prediction function between: *i*-th level and weighted (by level size) averaging between all levels. Intercept is as with DEV. Under balanced factor this type coincides with DEV.

HEL - **Helmert** type. Regression parameter at *i*-th contrast variable will be the comparison of prediction function between: *i*-th level and unweighted averaging between subsequent levels. Intercept of the regression will be prediction function unweightedly averaged between all levels. No contrast variable for *k*-th level. Contrasts of this type are orthogonal, and so if factor is balanced, contrast variables come uncorrelated.

WHEL - **weighted Helmert** type. This version of the previous considers the fact of possible unbalancedness. Regression parameter at *i*-th contrast variable will be the comparison of prediction function between: *i*-th level and weighted (by level size) averaging between subsequent levels. Intercept is as with HEL. Contrast variables come out always uncorrelated. But contrasts in **L** matrix are orthogonal only under balanced factor (then this type coincides with HEL).

DIF - **Difference** type, or reverse Helmert contrast. It is turned over previous type. No contrast variable for 1-st level. Regression parameter at *i*-th contrast variable will be the comparison of prediction function between: *i*+1-th level and unweighted averaging between preceding it levels. Intercept of the regression will be prediction function unweightedly averaged between all levels. Contrasts of this type are orthogonal, and so if factor is balanced, contrast variables come uncorrelated.

WDIF - **weighted Difference** type. This version of the previous considers the fact of possible unbalancedness. Regression parameter at *i*-th contrast variable will be the comparison of prediction function between: *i*+1-th level and weighted (by level size) averaging between preceding it levels. Intercept is as with DIF. Contrast variables come out always uncorrelated. But contrasts in **L** matrix are orthogonal only under balanced factor (then this type coincides with DIF).

REP - **Repeated** type. Regression parameter at *i*-th contrast variable will be the comparison of prediction function between: *i*-th level and *i*+1-th level. Intercept of the regression will be prediction function unweightedly averaged between all levels. No contrast variable for *k*-th level. Contrasts of this type are nonorthogonal; under balanced factor the contrast variables correlate positively with different strength.

*k*-th level in types IND, SIM, and DEV/WDEV is called reference one and its selection is arbitrary (on user). If you want to appoint another level than *k*-th to be reference, simply recode the input factor: assign the highest code to the level that you want to make reference. It will become *k*-th and reference. Types HEL/WHEL, DIF/WDIF, and REP order levels, therefore reference level is “fixed” in these types and should be planned in advance at coding of the factor.

Multinomial types, except WDEV, WHEL, WDIF, are unweighted: **L** and **C** matrices are computed identically with balanced and with unbalanced factor.

Multinomial contrast types exist in SPSS Statistics in GLM/UNIANOVA, MANOVA, LOGISTIC REGRESSION and some other procedures. Type IND is not everywhere there present (for a number of reasons); almost always type SIM is suitable instead of it.

EXAMPLE 2. Indicator type.

dataset name data.

!KO\_catcont factors= g1 g3 /types= IND ALL.

match files /file= data /file= \*.

execute.

dataset name merged.

* The macro encodes factors G1 (2 levels), G3 (3 levels) into type IND each.
* The output dataset is merged with the input one; further things are done on the merged dataset MERGED:

regression /dependent= y /method= enter a1 b1 b2.

unianova y with a1 b1 b2 /print parameter /design= a1 b1 b2.

* These two commands are equivalent, they run regression of Y on the contrast variables as quantitative predictors.
* Regression parameter at b1, for instance, equals the difference : , because corresponding coefficients in L matrix for factor B (i.e. G3) are 1 0 -1. It is the comparison of level B=1 with the reference level B=3 (contrast type IND). Of course, because there also the second factor is present in the model, it having 2 levels, each member of the above difference is actually a number that is the average of 2 numbers (one per each level of the second factor). I.e., : is composite: it is the mean of two values : ; and analogously – for : .
* Intercept with type IND equals of reference level. Because two factors in the model, the intercept is on the intersection of the reference levels; it equals .

unianova y by a b /print parameter /design= a b.

* ANOVA of Y by factors A and B (i.e. = G1 and G3), main effects model. Note that parameters being returned by UNIANOVA coincide with regression parameters obtained earlier on the contrast variables. That is because UNIANOVA (i.e. GLM) procedure internally is based on encoding factors into Indicator type.

EXAMPLE 3. Deviation and Repeated types.

dataset name data.

!KO\_catcont factors= f2 f3 /types= DEV REP.

match files /file= data /file= \*.

execute.

dataset name merged.

* The macro encodes factor F2 (3 levels) into type DEV and factor F3 (4 levels) into type REP.
* The output dataset is merged with the input one; further things are done on the merged dataset MERGED:

regression /dependent= y /method= enter a1 a2 b1 b2 b3.

unianova y with a1 a2 b1 b2 b3 /print parameter /design= a1 a2 b1 b2 b3.

* These two commands are equivalent, they run regression of Y on the contrast variables as quantitative predictors.
* Regression parameter at a1, for instance, equals the difference : , because corresponding coefficients in L matrix for factor A (i.e. F2) are .6667 -.3333 -.3333. And this difference is equal to: , what is the comparison of level A=1 with the unweighted mean of all three levels (contrast type DEV). Of course, because there also the second factor is present in the model (as was in the previous Example), and it having 4 levels, each member of the above difference is actually a number that is the average of 4 numbers (one per each level of the second factor).
* Regression parameter at b2, for instance, equals the difference : , because corresponding coefficients in L matrix for factor B (i.e. F3) are 0 1 -1 0. It is the comparison of level B=2 with level B=3 (contrast type REP). Of course, because there also the second factor is present in the model, and it having 3 levels, each member of the above difference is actually a number that is the average of 3 numbers (one per each level of the second factor).
* Intercept in types DEV and REP is the average of all the unique values of .

unianova y by a b /contrast(a)= deviation /contrast(b)= repeated /print parameter /design= a b.

* ANOVA of Y by factors A and B (i.e. = F2 and F3), main effects model. For factor A, type Deviation requested, and for factor B, type Repeated. Results – estimates of contrasts in table “Contrast Results (K Matrix)” – coincide with regression parameters obtained earlier on the contrast variables.
* Note that parameters in table “Parameter Estimates” don’t coincide with these results. That is because, as was remarked in EXAMPLE 2, this table in GLM always corresponds to Indicator encoding. But in MANOVA procedure, where algorithm is more straightforward than in GLM, parameter estimates always match the type of contrast you’ve requested:

autorecode a b /into a# b#.

manova y by a#(1 3) b#(1 4) /contrast(a#)= Deviation /contrast(b#)= Repeated

/print= parameters(estim) design(all) /design= a# b#.

* The same analysis via MANOVA procedure (A# and B# are simply numeric copies of A and B; you might have indicated the input F2 F3 with the same success). Parameters (table “Estimates”) coincide with contrast estimates in UNIANOVA and with regression parameters obtained on the contrast variables.
* Detailed report PRINT= DESIGN(ALL) shows, among other, coding matrices (Basis matrices) which !KO\_CATCONT displays too. “Between-Subjects Design Matrix” is the resume of values of all contrast variables together, that is, it is this:

dataset declare aggr.

aggregate /outfile= 'aggr' /BREAK= a b

/a1\_first=FIRST(a1)

/a2\_first=FIRST(a2)

/b1\_first=FIRST(b1)

/b2\_first=FIRST(b2)

/b3\_first=FIRST(b3).

* (but for the 1st column which in MANOVA printout corresponds to the intercept).

EXAMPLE 3a. Deviation unweighted and weighted types.

dataset name data.

!KO\_catcont factors= f3 f3 /types= DEV WDEV.

match files /file= data /file= \*.

execute.

dataset name merged.

* The macro encodes factor F3 (4 levels) ones into type DEV (variables a1 a2 a3) and another time into type WDEV (variables b1 b2 b3). The output dataset is merged with the input one; further things are done on the merged dataset MERGED:

regression /dependent= y /method= enter a1 a2 a3.

regression /dependent= y /method= enter b1 b2 b3.

* The first regression uses a1 a2 a3 as the predictors. Regressional parameter under type DEV is the difference of means between a level and the unweighted grand mean of all the levels. Parameter at a1, for example, is the difference : .
* The second regression uses b1 b2 b3 as the predictors. Regressional parameter under type WDEV is the difference of means between a level and the weighted grand mean of all the levels, i.e., the total mean in the sample. Parameter at b1, for example, is the difference : .

b) **Polynomial types**. Factor is taken not for categorical factor but for quantitative variable (see SPACING subcommand). There generated are linear, quadratic, and more high degree contrasts for exploring linear and curvilinear dependencies from such metric predictor. The 1-st contrast answers linear dependency from it, the 2-nd contrast – quadratic dependency above linear, the 3-rd contrast – cubic dependency above[[2]](#footnote-2) linear and quadratic, and so forth. *i*-th contrast variable corresponds, thereby, not to *i*-th level (as in multinomial types) but *i*-th degree, degree of nonlinearity. *k*-1 contrasts are generated, they all together comprehensively replace the factor. The macro offers two types – unweighted and weighted. These two types are identical in case the factor is balanced (equal-sized groups), they differ in situation of an unbalanced factor.

POL - **unweighted Polynomial**. This type ignores the fact of unbalancedness: factor levels are assigned equal significance. Unweighted contrasts are always orthogonal in **L** matrix (and **C** coincides with **L**´). Contrast variables come out uncorrelated only in case of balanced factor.

WPOL - **weighted Polynomial**. This type accounts for the fact of unbalancedness and contrast variables come out always uncorrelated. But contrasts in **L** matrix are orthogonal only in case of balanced factor.

Weighted polynomial contrast exists in SPSS Statistics’ One-way Analysis of Variance (ONEWAY). Unweighted type exists also in GLM/UNIANOVA, MANOVA, LOGISTIC REGRESSION and a number of other procedures.

EXAMPLE 4. Polynomial contrasts.

dataset name data.

!KO\_catcont factors= f3 f3 /types= POL WPOL.

match files /file= data /file= \*.

execute.

dataset name merged.

* In this example, one and same factor F3 (4 levels, unbalanced factor) is encoded once into unweighted polynomial type (yielding set of variables a1 a2 a3), second time into weighted polynomial type (yielding set of variables b1 b2 b3). The variables of the latter type are uncorrelated, despite of the unbalancedness.
* The output dataset is merged with the input one; further things are done on the merged dataset MERGED:

regression /dependent= y /method= enter a1 a2 a3.

unianova y by f3 /contrast(f3)= polynomial /design= f3.

* Regression on the contrast variables (unweighted polynomial type) gives the same results as ANOVA requesting polynomial type for factor F3.
* In the model with predictors a1, a2, a3 the regression coefficient at a1 expresses linear trend, coefficient at a2 expresses quadratic trend (above linear), coefficient at a3 expresses cubic trend (above linear and quadratic).

regression /dependent= y /method= enter b1 b2 b3.

oneway y by f3 /polynomial= 3.

* Regression on the contrast variables (weighted polynomial type) gives the same results as ANOVA requesting weighted polynomial type for factor F3. Weighted type is available in SPSS Statistics in ONEWAY command. This procedure does not display in its Anova table the values of contrasts (i.e., the parameter estimates), but displays their *p*-values. And these are the same as *p*-values of the parameters in the regression. In the Anova table you see both weighted polynomial type and unweighted polynomial type (results for that latter we’ve already obtained above).

c) **User-defined**, or special, contrast. Inline entering or filename.

*matrix/file* - **user-defined contrast**. Enter your matrix of contrast coefficients (**L** matrix) for the factor, size *k*-1 contrast × *k* levels. Normally the sum of coefficients in each contrast must be zero.

If the entering is inline, put the matrix in braces {}; row elements to separate with comma, rows to separate with semicolon. The matrix goes row-after-row. Carrying over to next line is arbitrary. An example of matrix with *k*=4:

{ 3, -1, -1, -1;

0, 2, -1, -1;

0, 0, 1, -1 }

Same matrix in one line: {3, -1, -1, -1; 0, 2, -1, -1; 0, 0, 1, -1}

Elements of matrix being entered need not be integer, but they must be numbers, not expressions (*not* permitted, in particular, to enter vulgar fraction, for example: -1/3).

And if you are specifying a .SAV file name (in quotes or apostrophes), then its data must be likewise *k*-1 × *k* dataset of numeric coefficients; column names play no role.

EXAMPLE 5.

!KO\_catcont factors= g2 g3 g4 /types= {1,-0.5,-0.5;0,1,-1} 'd:\exercise\g3\_lmx.sav' HEL /inter= ALL.

* For factor G2 (3 levels) a user-defined contrast matrix (2×3) is specified, entered inline. For factor G3 (3 levels) a user-defined contrast matrix (2×3) is specified, external file; that matrix was: {-1,0,1;-0.333333,-0.333333,0.666666}. Type Helmert was specified for factor G4.
* Creation of interaction variables was requested.

Orthogonality of contrasts. It was pointed out above that contrasts of type HEL, DIF, POL are orthogonal[[3]](#footnote-3). *Orthogonal* contrasts are called such ones for which sum of products of corresponding elements for every pair of rows in **L** matrix equals 0; in other words, **LL´** is diagonal matrix. When contrasts are orthogonal, then under the condition of balanced factor the contrast variables don’t correlate among each other, therefore their individual effects will not overlap and will sum up precisely to the total prediction by their entire set. While if a factor is unbalanced or if contrasts are nonorthogonal, then contrast variables correlate, so their individual effects will overlap and in sum turn to be less than the total prediction by their set. So takes place under simultaneous admission of all the variables of the set to be predictors (what corresponds to Type III SS aka *unique* method). You may inlet contrast predictors one at a time (Type I SS aka *sequential* method), but, as a rule, one should use all the set: contrast variables imply each other – because together they represent one whole factor and exhaust it.

Contrasts in the presence of interactions. When there are interaction terms in a model, the meaning of a contrast belonging to a factor (i.e., to the main effect) gets modified, incorporating the fact of present interactions. The contrast’s meaning doesn’t change fundamentally, rather, it is concretized, adjusting for the interactions. See EXAMPLE 8, 9. Specially to remark, that under interactions present, the meaning of contrasts of types IND and SIM, basically similar, gets concretized differently.

**SPACING**

This subcommand acts only for polynomial contrasts: TYPES mentioning keywords POL or WPOL. By default and with SPACING=EQUAL, levels of factor are taken to be separated with equal intervals, and actual coding in the input factors to whom polynomial type was prescribed, plays no role. That is, the factor is taken as ordinal and then the levels are assigned quantitative values (“metrics”) 1, 2, 3...

With SPACING=ASIS, coding of input factors to whom polynomial type was prescribed, is taken as is, i.e., the factor is used as a ready scale variable with existing in it ratio of intervals between levels. Only ratio among intervals, not magnitude of values, plays a role. Values must be positive.

EXAMPLE 6.

dataset name data.

!KO\_catcont factors= g3 g2 /types= POL POL /spacing= ASIS.

match files /file= data /file= \*.

execute.

dataset name merged.

* The macro encodes factors G3 and G2 into type POL. SPACING=ASIS demands to treat values of the input G3, G2 literally, as interval ones.
* The output dataset is merged with the input one; further things are done on the merged dataset MERGED:

unianova y with a1 a2 b1 b2 /print parameter /design= a1 a2 b1 b2.

unianova y by g3 g2 /contrast(g3)= polynomial(1 2 3) /contrast(g2)= polynomial(10 25 60) /design= g3 g2.

* These two commands are equivalent. The first runs regression of Y on contrast variables obtained from the factors. The second runs ANOVA encoding the factors into polynomial type. Regression coefficients after the first command coincide with contrast estimates in table “Contrast Results (K Matrix)” after the second command.
* Note that in the specification of the second command, in subcommands CONTRAST, user-defined intervals were requested between levels in a factor. In the present case, the values of the input factors were indicated as they are. Such specification corresponds to what we’ve done at the creation of contrast variables, by indicating SPACING=ASIS in !KO\_CATCONT.

**INTER**

By default, the macro creates only contrast variables of factors (i.e., of main effects). If you need contrast variables bringing up interactions between factors when factors are more than one, demand them here. These variables are created by cross-multiplication of factor contrast variables (the factor contrast variables themselves don’t depend on INTER subcommand). The highest order you have the right to request – 4-way interactions.

Specify the order by integer from 2 to 4. Keyword UP before the number means “all interactions up to” this order. For example, UP 3 is the request for 2-way and 3-way interactions. You may also write INTER=ALL, meaning “all possible interactions” (but not higher than 4). INTER=2 and INTER= UP 2 are equivalent, since 2 is the smallest order of interactions.

If factors have many levels, request of interactions considerably slows the job and loads memory resource. !KO\_CATCONT does almost all its job in RAM memory.

EXAMPLE 7.

dataset name data.

!KO\_catcont factors= g1 g2 g3 g4 /types= SIM IND SIM DEV /inter= UP 3 /seq= EFFECT /print= SHORT.

match files /file= data /file= \*.

execute.

dataset name merged.

* The macro encodes factors G1, G2, G3, G4 into requested types SIM, IND, SIM, DEV, respectively. It is requested to create, out of contrast variables, interaction contrast variables between the factors, up to 3-way.
* The output dataset is merged with the input one; further things are done on the merged dataset MERGED:

logistic regression y2

/method= enter a1 b1 b2 c1 c2 d1 a1b1 a1b2 a1c1 a1c2 a1d1 b1c1 b1c2 b2c1 b2c2 b1d1 b2d1 c1d1 c2d1

a1b1c1 a1b1c2 a1b2c1 a1b2c2 a1b1d1 a1b2d1 a1c1d1 a1c2d1 b1c1d1 b1c2d1 b2c1d1 b2c2d1 covar1.

* Logistic regression of Y2 on the contrast variables, including interaction ones, as quantitative predictors. Covariate COVAR1 was also included in the model.

logistic regression y2

/method= enter a b c d a\*b a\*c a\*d b\*c b\*d c\*d a\*b\*c a\*b\*d a\*c\*d b\*c\*d covar1

/categorical= a b c d

/contrast(a)= simple /contrast(b)= indicator /contrast(c)= simple /contrast(d)= deviation.

* Logistic regression of Y2 by factors A, B, C, D (i.e. = G1, G2, G3, G4), model with main effects and interactions up to 3-way; covariate COVAR1 also present. The requested contrast types, i.e. methods of factor encoding: Simple, Indicator, Simple, Deviation. Results – regression parameters – coincide with those obtained on the contrast variables before.

**SEQ**

This subcommand only affects the sequence in which contrast variables of interactions go in the output data. By default and with SEQ=COMBIN it is the combinatorial order. With SEQ=EFFECT, all variables making up an effect are gathered together. For example, if there are factors A (3 levels), B (3 levels), and C (2 levels), then the combinatorial order of their 2-way interaction contrast variables is:

a1b1 a1b2 *a1c1* a2b1 a2b2 *a2c1* b1c1 b2c1

And the order of the same variables under SEQ=EFFECT is:

a1b1 a1b2 a2b1 a2b2 *a1c1 a2c1* b1c1 b2c1

Variables making up each of effects A\*B, A\*C, B\*C are not parted.

**ID**

Optional numeric case identifier variable, indicate its name. Since the macro screens out nonvalid cases listwise, the output dataset may happen to have less cases than the input one has. ID-variable will allow to correctly merge the input dataset with the variables output by the macro.

**PRINT**

Report into Output Viewer:

LONG - (default/unspecifying) complete report, including **C** and **L** matrices for each factor, checking orthogonality of contrasts, etc.

SHORT - concise report: factor information: their symbols, levels, frequencies; determinant of correlation matrix of contrast variables[[4]](#footnote-4).

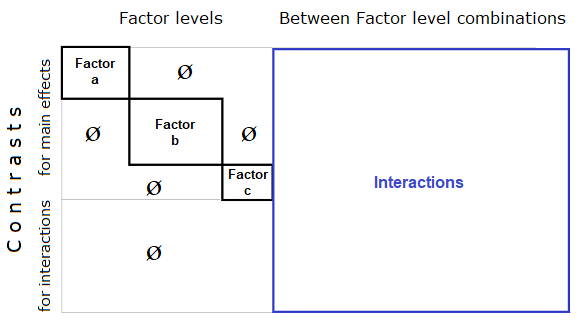
NONE - suppress report, except information on symbols of factors.

**SAVELMX**

Optional subcommand to compile and save **L** matrix (contrast coefficient matrix) for all the factors together and their interactions (if interactions were requested in INTER). You can later input this **L** matrix (whole or its selected rows, the contrasts) in such procedures as GLM/UNIANOVA, ONEWAY, MANOVA, LOGISTIC REGRESSION, COXREG, MIXED, and other. In these procedures there exist subcommands LMATRIX or CONTRAST or TEST or EMMEANS with option to specify user-defined contrasts through syntax. See EXAMPLE 10, where **L** matrix is input to LMATRIX subcommand of GLM procedure.

If interactions were requested whose design table contained empty cells, SAVELMX subcommand will not make the matrix and will issue warning.

Specify path/name of .SAV file to save, quoted or in apostrophes. S/c SAVELMX is permitted only if factors are no more than 4 and interactions, if requested, are not higher than order 3. The structure of **L** matrix is such:



Rows are contrasts and correspond to the contrast variables (sum of coefficients in each row normally is 0), and columns correspond to all the factor levels and (if interactions were requested) interfactor combinations of levels. In other words, columns are the detailed list of all terms of the design in the “overparameterized” style (i.e., without omitting “redundant” levels). Left-top part is occupied by per-factor **L** matrices joined block-diagonally. If interactions were not requested, the matrix is merely that part. Right part is occupied by coefficients that appeared due to the involvement of interactions to the model. Note that these coefficients for contrasts of main effects (“for main effects”) in that right part of the matrix are not zero; they are the “scheduling” of coefficients existing in the left part by cells of interfactor design.

EXAMPLE 8. Indicator and Simple types in the presence of interactions.

!KO\_catcont factors= g2 g3 /types= SIM ALL /inter= ALL /savelmx= 'd:\exercise\sim\_lmx.sav'.

get file 'd:\exercise\sim\_lmx.sav'.

format all (f6.4).

list.

* The macro encodes factors G2 (3 levels) and G3 (3 levels) into type SIM. Interactions between the factors are requested. Combined L matrix is saved. Columns 1–3 in it correspond to main effect of A (= G2), columns 4–6 correspond to main effect of B (= G3), columns 7–15 correspond to effect A\*B ( = G2\*G3).
* Let’s consider as example the 1-st row of L matrix (this contrast corresponds to contrast variable a1 and expresses the meaning of regression parameter at that variable in the situation when all 8 contrast variables serve to be regressors, i.e., in case of model A B A\*B).

This contrast is: A 1 0 -1 B 0 0 0 A\*B 1/3 1/3 1/3 0 0 0 -1/3 -1/3 -1/3, that is:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | B | | |
|  |  | 0 | 0 | 0 |
| A | 1 | 1/3 | 1/3 | 1/3 |
| 0 | 0 | 0 | 0 |
| -1 | -1/3 | -1/3 | -1/3 |

As we see, the 1-st contrast of type Simple compares the 1-st level of factor A with the 3-rd its level, and in the condition of present interactions each of the two nonzero coefficients is uniformly “scheduled” between levels of the second factor (B), with which A interacts. While in the absence of interaction effect in the model the comparison would be the difference : , ignoring factor B, in the presence of interaction this difference gets corrected: . The difference is uniformly “spread” across levels of B.

* Let’s consider the 5-th row of L matrix (this contrast corresponds to contrast variable a1b1 and expresses the meaning of regression parameter at this variable, which is a term in the interaction effect A\*B).

This contrast is: A 0 0 0 B 0 0 0 A\*B 1 0 -1 0 0 0 -1 0 1, that is:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | B | | |
|  |  | 0 | 0 | 0 |
| A | 0 | 1 | 0 | -1 |
| 0 | 0 | 0 | 0 |
| 0 | -1 | 0 | 1 |

As we see, the 1-st interaction contrast gives the difference : (; and regression coefficient at a1b1 will be just this comparison.

* Intercept with type SIM is the average of all the unique values of .

!KO\_catcont factors= g2 g3 /types= IND ALL /inter= ALL /savelmx= 'd:\exercise\ind\_lmx.sav'.

get file 'd:\exercise\ind\_lmx.sav'.

format all (f6.4).

list.

* This run of the macro is fully like the previous, but it encodes the factors into type IND.
* Let’s consider as example the 1-st row of L matrix (this contrast corresponds to contrast variable a1 and expresses the meaning of regression parameter at that variable in the situation when all 8 contrast variables serve to be regressors, i.e., in case of model A B A\*B).

This contrast is: A 1 0 -1 B 0 0 0 A\*B 0 0 1 0 0 0 0 0 -1, that is:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | B | | |
|  |  | 0 | 0 | 0 |
| A | 1 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 |
| -1 | 0 | 0 | -1 |

As we see, the 1-st contrast of type Indicator compares the 1-st level of factor A with the 3-rd its level, and in the condition of present interactions each of the two nonzero coefficients is “scheduled” between levels of the second factor (B) not uniformly, in a way that everything is assigned on account of B=3 (reference level in B). That is, comparison : is being made actually only in the selected level of B: . Clearly, this is not the same comparison as the one seen above with type SIM. It is a simple effect, not main effect. So, under the presence of interactions in a model results for main effects will differ for SIM and IND, and specifically SIM should be preferred if coefficient at a1 is to express main effect.

* Let’s consider the 5-th row of L matrix (this contrast corresponds to contrast variable a1b1 and expresses the meaning of regression parameter at this variable, which is a term in the interaction effect A\*B).

This contrast is: A 0 0 0 B 0 0 0 A\*B 1 0 -1 0 0 0 -1 0 1. Note that this is the same what was in this place with type SIM. And in fact: results (regression parameters = contrast estimates) of types SIM and IND will coincide – for the contrast variables that are the interaction terms. Let’s say the *rule* in general form: results after SIM and IND are always identical for the effect of a highest order present in the model, and only for that effect. As a specific case: if no interactions in the model, then results after SIM and IND are identical for main effects – what has been written while characterizing these two types in s/c TYPES.

* Intercept with type IND equals in the cell “all factors = reference level”.

EXAMPLE 9. Deviation type in the presence of interactions.

!KO\_catcont factors= g2 g3 /types= DEV ALL /inter= ALL /savelmx= 'd:\exercise\dev\_lmx.sav'.

get file 'd:\exercise\dev\_lmx.sav'.

format all (f6.4).

list.

* The macro encodes factors G2 (3 levels) and G3 (3 levels) into type DEV. Interactions between the factors are requested. Combined L matrix is saved. Columns 1–3 in it correspond to main effect of A (= G2 columns 4–6 correspond to main effect of B (= G3), columns 7–15 correspond to effect A\*B ( = G2\*G3).
* Let’s consider as example the 1-st row of L matrix (this contrast corresponds to contrast variable a1 and expresses the meaning of regression parameter at that variable in the situation when all 8 contrast variables serve to be regressors, i.e., in case of the model A B A\*B).

This contrast is: A 2/3 -1/3 -1/3 B 0 0 0 A\*B 2/9 2/9 2/9 -1/9 -1/9 -1/9 -1/9 -1/9 -1/9, that is:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | B | | |
|  |  | 0 | 0 | 0 |
| A | 2/3 | 2/9 | 2/9 | 2/9 |
| -1/3 | -1/9 | -1/9 | -1/9 |
| -1/3 | -1/9 | -1/9 | -1/9 |

As we see, the 1-st contrast of type Deviation compares the 1-st level of factor A with the unweighted averaged of all its levels: , and in the condition of present interactions these contrast coefficients (2/3, -1/3, -1/3) are uniformly “scheduled” between levels of the second factor (B), with which A interacts. The difference is evenly “spread” across levels of B, and the uniformity is the pledge of that a1 correctly (reasonably) expresses the main effect under the interaction present. Let’s comment, that other contrast types, except IND, act the similar way.

* Let’s consider the 5-th row of L matrix (this contrast corresponds to contrast variable a1b1 and expresses the meaning of regression parameter at this variable, which is a term in the interaction effect A\*B).

This contrast is: A 0 0 0 B 0 0 0 A\*B 4/9 -2/9 -2/9 -2/9 1/9 1/9 -2/9 1/9 1/9, that is:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | B | | |
|  |  | 0 | 0 | 0 |
| A | 0 | 4/9 | -2/9 | -2/9 |
| 0 | -2/9 | 1/9 | 1/9 |
| 0 | -2/9 | 1/9 | 1/9 |

As we see, the 1-st interaction contrast gives the difference : (; and regression coefficient at a1b1 will be just this comparison.

* Intercept with type DEV is the average of all the unique values of .

EXAMPLE 10. Using of L matrix.

dataset name data.

!KO\_catcont factors= f1 f2 f3 f4 /types= SIM DEV POL DIF /inter= UP 3 /seq= EFFECT

/savelmx= 'D:\exercise\lmx.sav'.

match files /file= data /file= \*.

execute.

dataset name merged.

* The macro encodes factors F1, F2, F3, F4 into various contrast types. Also, interactions up to 3-way are requested to create. L matrix (main effects + interactions) is also saved.
* The output dataset is merged with the input one; further things are done on the merged dataset MERGED:

regression /dependent y

/method= enter a1 b1 b2 c1 c2 c3 d1

a1b1 a1b2 a1c1 a1c2 a1c3 a1d1 b1c1 b1c2 b1c3 b2c1 b2c2 b2c3 b1d1 b2d1 c1d1 c2d1 c3d1

a1b1c1 a1b1c2 a1b1c3 a1b2c1 a1b2c2 a1b2c3 a1b1d1 a1b2d1 a1c1d1 a1c2d1 a1c3d1

b1c1d1 b1c2d1 b1c3d1 b2c1d1 b2c2d1 b2c3d1.

* Regression of Y on all the created contrast variables; i.e., it is the equivalent of the 4-factor ANOVA with main effects and all interactions up to 3-way.

autorecode a b c d /into a# b# c# d#.

manova y by a#(1 2) b#(1 3) c#(1 4) d#(1 2)

/contrast(a#)= simple /contrast(b#)= deviation /contrast(c#)= polynomial /contrast(d#)= difference

/print= parameters(estim) design(overall)

/design= a# b# c# d# a#\*b# a#\*c# a#\*d# b#\*c# b#\*d# c#\*d# a#\*b#\*c# a#\*b#\*d# a#\*c#\*d# b#\*c#\*d#.

* Indeed, MANOVA with the factors yields the same results as the regression on the contrast variables. Parameters are in table “Estimates”. (AUTORECODE needed because MANOVA demands numeric input variables coded by sequential integers).

unianova y by a b c d

/contrast(a)= simple /contrast(b)= deviation /contrast(c)= polynomial /contrast(d)= difference

/print test(lmatrix)

/design= a b c d a\*b a\*c a\*d b\*c b\*d c\*d a\*b\*c a\*b\*d a\*c\*d b\*c\*d.

* And this is the same results obtainable by UNIANOVA (= GLM) procedure. Contrast estimates (coinciding with the regression parameters above) are in section “Custom Hypothesis Tests...”. However, UNIANOVA does not show – among the results for contrasts – the results pertaining to the interaction terms. We can see only results matching the parameters of contrast variables a1, b1, b2, c1, c2, c3, d1, and we don’t see other parameters: a1b1, ..., b2c3d1. In order to receive them in UNIANOVA, it is necessary to input the complete L matrix, which !KO\_CATCONT has saved. In UNIANOVA, such an L matrix is introduced via subcommand LMATRIX:

get file 'd:\exercise\lmx.sav'.

format all (f14.12).

string a b c d ab ac ad bc bd cd abc abd acd bcd end (a5).

compute a= 'a'.

compute b= 'b'.

compute c= 'c'.

compute d= 'd'.

compute ab= 'a\*b'.

compute ac= 'a\*c'.

compute ad= 'a\*d'.

compute bc= 'b\*c'.

compute bd= 'b\*d'.

compute cd= 'c\*d'.

compute abc= 'a\*b\*c'.

compute abd= 'a\*b\*d'.

compute acd= 'a\*c\*d'.

compute bcd= 'b\*c\*d'.

compute end= ';'.

execute.

* Open the saved L matrix. For the matrix to be reliably estimable in UNIANOVA, set format with a lot of decimal digits. Then create string constants of letters (factor names): main and interaction effects. They will mark blocks of columns in the L matrix. Create also column END = semicolon (a row in LMATRIX will be closed by it).

summarize /tables=

a COL1 COL2 b COL3 COL4 COL5 c COL6 COL7 COL8 COL9 d COL10 COL11

ab COL12 COL13 COL14 COL15 COL16 COL17

ac COL18 COL19 COL20 COL21 COL22 COL23 COL24 COL25

ad COL26 COL27 COL28 COL29

bc COL30 COL31 COL32 COL33 COL34 COL35 COL36 COL37 COL38 COL39 COL40 COL41

bd COL42 COL43 COL44 COL45 COL46 COL47

cd COL48 COL49 COL50 COL51 COL52 COL53 COL54 COL55

abc COL56 COL57 COL58 COL59 COL60 COL61 COL62 COL63 COL64 COL65 COL66 COL67

COL68 COL69 COL70 COL71 COL72 COL73 COL74 COL75 COL76 COL77 COL78 COL79

abd COL80 COL81 COL82 COL83 COL84 COL85 COL86 COL87 COL88 COL89 COL90 COL91

acd COL92 COL93 COL94 COL95 COL96 COL97 COL98 COL99 COL100 COL101 COL102 COL103

COL104 COL105 COL106 COL107

bcd COL108 COL109 COL110 COL111 COL112 COL113 COL114 COL115 COL116 COL117 COL118

COL119 COL120 COL121 COL122 COL123 COL124 COL125 COL126 COL127 COL128 COL129

COL130 COL131

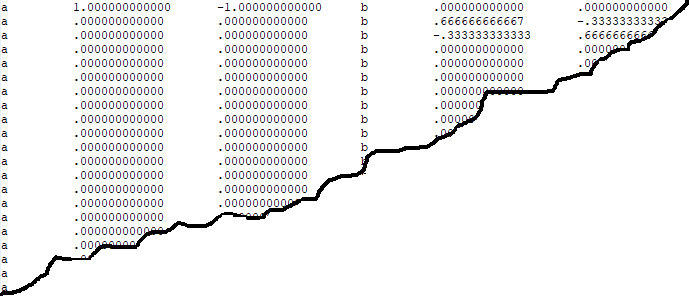
end

/format= list nocasenum /cells= none.

* Insert the just created variables – column-blocks-separators – on their right places. These right places !KO\_CATCONT

always suggests to you, in the Output window, when s/c SAVELMX is specified. Then run SUMMARIZE and receive in the Output window the ready-for-input L matrix. Having opened the table, select all row labels and then, by right-click: Select > Data Cells. Copy. Paste in LMATRIX subcommand where indicated below.

* Instead of SUMMARIZE, you might use LIST, though it is less convenient.



dataset activate merged.

unianova y by a b c d

/print test(lmatrix)

/lmatrix=

[PASTE THE MATRIX HERE]

/design= a b c d a\*b a\*c a\*d b\*c b\*d c\*d a\*b\*c a\*b\*d a\*c\*d b\*c\*d.

* Results – contrast estimates in table “Contrast Results (K Matrix)” – coincide with the regression parameters obtained earlier on the contrast variables, and there seen in the results window are the terms for interactions too, not only for the main effects.
* Convenience of an L matrix of contrasts is that its rows (contrasts) can be input in subcommands, like LMATRIX, selectively. Contrast estimates will be identical to when you input the complete L matrix covering the entire prediction model. For instance, in the current example we might have input in LMATRIX only, say, the 2-nd and the 3-rd rows (they correspond to contrasts b1 and b2). Their results would have been still equal to the regression parameters at b1 and b2 of the multiple regression containing all the variables from a1 to b2c3d1 as regressors.

***Special regimes***

The macro is not suited for split state of the dataset. Case weights are ignored. The macro obeys filtering (FILTER, SELECT IF, USE etc.) and temporary (under TEMPORARY) operations.

1. Type IND is a special case, where det(SSCP) is not zero. Due to this, it is possible to enter a set of *k* dummies as predictors, but only in case the model is *without* intercept. [↑](#footnote-ref-1)
2. “above” means sequential increment. Let X be factor as scale variable. Then “linear” contrast variable is linearly equivalent to X. “Quadratic” contrast variable is linearly equivalent to the residual of linear prediction of X2 on X. “Cubic” contrast variable is linearly equivalent to the residual of linear prediction of X3 on X2 and X, and so forth. [↑](#footnote-ref-2)
3. Type WPOL = POL under balanced factor. [↑](#footnote-ref-3)
4. SPSS Statistics may inform sometimes: “The determinant has a magnitude too small to be represented”. It isn’t the macro’s error and is normal. [↑](#footnote-ref-4)