***Fit covariates***

SPSS macros by Kirill Orlov

kior@comtv.ru, ttnphns@gmail.com

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

All rights reserved

*Fitting variables to a matrix of coefficients.* The macros modify variables’ values so that the variables have strength of relations according to a user-specified matrix (correlation, covariance, or cross-product). Option of insurance against heteroscedasticity allows to achieve homoscedastic relationships.

To shape variables to have a specific covariation with each other use [!KO\_FITCOV](#_MACRO_!FITCOV:_FITTING).

To shape variables to have a specific covariation with third variables use [!KO\_FITVAR](#_MACRO_!FITVAR:_FITTING).

Both macros are capable to correct heteroscedasticity. Both macros treat data as continuous as they are free to change values to fractional ones.

To shape random data first generate those random variables. For example, in this manner:

do repeat var= var1 to var10.

compute var= rv.normal(5,2).

end repeat.

execute.

# MACRO !KO\_FITCOV: FITTING VARIABLES TO A SQUARE SYMMETRIC MATRIX OF COEFFICIENTS

Version 1 (June 2003). Tested on SPSS 11, 11.5, 13, 14.

!KO\_fitcov vars= *v1 to v10* /\*Variables to get fit: list of names (may use “to”) that should be also among the matrix column names;

/\*or ALL – take all variables of the file and all columns of the matrix, - they should be of the same number while names need not be the same

/matrix= *'d:\exercise\corrmx.sav'* /\*Target matrix or matrix that contains it as submatrix; external file

/method= IPCOMP /\*Mode of fitting the covariates: QUICK – noniteratively via principal components, IREGR – iteratively via regression, IPCOMP - iteratively via principal components

/adjust= NONE /\*For iterative modes, additional request: NONE, CENTER (center the immediately being fitted variable, default),

/\*HSCED1 (insure against fan-like heterosced-ty), HSCED2 (insure against symmetric heterosced-ty)

/iter= *20 2* /\*For iterative modes: Number of iterations and, optionally, increment for iterations

/msave= *'d:\exercise\newcorrmx.sav'* /\*Optional: Path/name of external file to save the reproduced matrix.

Minimal specification VARS, MATRIX, METHOD, ITER

The macro modifies the input variables’ values so that the variables show the strength of relations with each other close or equal to coefficients of a user-specified matrix, correlation, covariance, or SSCP[[1]](#footnote-1). With the help of this macro one can “carry” existant or random variables to the needed amounts of covariation between them. In addition, there is an option to correct heteroscedasticity, i.e. carry to homoscedasticity[[2]](#footnote-2), in the relations between variables (that might be occasionally the only aim to use the macro).

EXAMPLE 1.

!KO\_fitcov vars= v1 v8 to v15 /matrix= 'd:\exercise\corrmx.sav' /method= IREGR /iter= 20 2.

* Variables in the file, v1, v8 and through v15, need to be brought to intercorrelations according to matrix “corrmx.sav”, using iterative regression method with starting number of iterations 20 and increment 2.

# Procedure

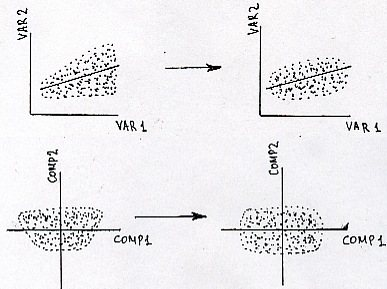
Fitting of variables (covariates) flows in iterative algorithm (invented by the macro’s author) consisting of recurrent adding, to values of a variable, the correction which is the weighted mean of individual corrections other variables demand to make – to provide the association coefficients “they need” to have with our covariate. The covariates values serve as weights, due to what the “demand” of every covariate can be gradually satisfied. In so doing, it is attained that the given covariate, as a result, shows associations with other covariates as much as one wants close to the target (specified) coefficients. Introducing new terms and another power for weights in the formula of correction allows to iteratively satisfy some extra requests, such as homoscedasticity (formulas find below).

The iterative algorithm is implemented in two “methods”. There is also a third method which is not iterative.

* *Регрессионный метод*. Вторая в списке входящая переменная берется как зависимая (предиктант), а 1-я как предсказывающая (предиктор, или регрессор). Должный между ними линейно-регрессионный коэффициент узнается из целевой матрицы и вычисляются предсказанные этой моделью значения, которые тут же вычитанием из предиктанта дают исходные регрессионные остатки. Эти остатки подгоняются итеративно к тому, чтобы иметь нулевое коварьирование с предиктантом (таким образом, целевыми *непосредственно* выступают нулевые коэффициенты). После подгонки и обеспечения должной их дисперсии остатки обратно складываются с предсказанными значениями в исправленную 2-ю переменную. После того, 3-я переменная берется за предиктант, а 1-я и 2-я за предикторы, и процедура повторяется теперь над 3-й переменной. И так далее: каждая переменная выступает зависимой от предшествующих, подгоняется (своими регрессионными остатками) и переходит в разряд предшествующих.
* *Главнокомпонентный метод.* Из целевой матрицы узнаем собственные числа и собственные векторы. Последние, перемножаясь с входящими переменными, дают заготовки главных компонент. Эти, начиная со 2-й, последовательно подгоняются итеративно (так же, как остатки подгоняются в регрессионном методе) к тому чтобы иметь нулевое коварьирование с предшествующими компонентами, – поскольку они должны стать взаимоортогональны. Заодно пересчитывается масштаб компонент на собственные числа целевой матрицы. По окончании подгонки компонент они обратно вращаются в переменные.
* *Неитеративный (быстрый) главнокомпонентный метод.* Входящие переменные вращаются в свои главные компоненты. Из целевой матрицы узнаем собственные числа и собственные векторы. Компоненты пересчитываются своим масштабом на целевые собственные числа, после чего компоненты обратно вращаются в переменные. Этот метод, заключающийся в пропорциональном сжатии/распускании облака данных вдоль компонент, скор и дает всегда полное достижение целевых коэффициентов ковариатами, но не обладает дополнительными возможностями.

Дополнительные возможности

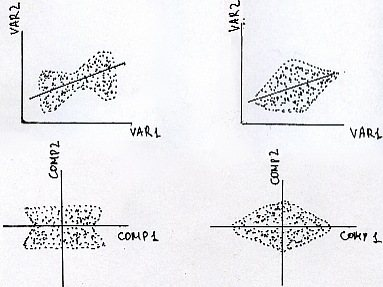
* *Центрация*. Во время подгонки переменной регрессионных остатков или компоненты накладывается требование достижения нулевой (близкой к 0) средней у этой, непосредственно подгоняемой переменной. Такое условие обычно весьма желательно: регрессионные остатки в норме должны иметь нулевую среднюю; а главные компоненты данных, за исключением 1-й компоненты, почти всегда имеют близкие или равные нулю средние. Опция центрации важна не как самоцель: при ней обыкновенно подгонка под коэффициенты успешнее.
* *Страхование от веерной гетероскедастичности*. Веерная гетероскедастичность довольно часто встречается в данных и делает ненадежной оценку регрессионных моделей. На рисунке, в верхнем ряду показана веерная гетероскедастичность зависимой переменной VAR2 вдоль предиктора VAR1, и то как она может быть исправлена в гомоскедастичность макросом: регрессионным методом с подключением опции «страховать от веерной гетероскедастичности». В этом случае макрос добивается (при достаточном числе итераций) гомоскедастичности текущей подгоняемой переменной в ее отношениях со всеми своими предикторами, т.е. предшествующими (уже подогнанными) переменными. Именно, *если подгоняемая сейчас своими остатками переменная – i-я, то ее изменчивость, «толщина» облака, - становится симметрична относительно центра всех i-1 предшествующих переменных.* Заметьте, что гомоскедастичность переменной *не гарантируется* в ее отношениях с последующими переменными, т.к. подгонка переменных при регрессионном методе идет последовательно.



Во втором ряду рисунка показано исправление веерной гетероскедастичности главнокомпонентным методом с опцией «страховать от веерной гетероскедастичности». В отличие от регрессионного метода, тут макрос добивается гомоскедастичности не самой подгоняемой сейчас компоненты, а всех предшествующих (уже подогнанных) компонент вдоль нее. Именно, *если подгоняемая сейчас компонента – i-я, то облако становится симметрично относительно ее центра в пространстве первых i компонент*. Так, на рисунке неравномерность изменчивости 1-й компоненты COMP1 вдоль 2-й компоненты COMP2, подгоняемой сейчас, заменена на равномерность. Гомоскедастичность последующих компонент (COMP3 и т.д.) не гарантируется вдоль подгоняемой COMP2.

На рисунке выше исправленные облака показаны «правильными». Однако в действительности страхование от веерной гетероскедастичности не гарантирует именно такой правильности, поскольку не препятствует случайному возникновению либо сохранению в данных Х-образной либо ромбической гетероскедастичности. Поэтому в макросе есть и более радикальная опция борьбы с гетероскедастичностью – страхование от симметричной гетероскедастичности.

* *Страхование от симметричной гетероскедастичности.* На рисунке ниже изображены симметричные виды гетероскедастичности – Х-образная (слева) и ромбовидная (справа). Опция «страхование от симметричной гетероскедастичности» подгоняет такие облака данных к гомоскедастичным, показанным на предыдущем рисунке справа. Веерная гетероскедастичность также исправляется этой опцией[[3]](#footnote-3). В сочетании с регрессионным методом опция делает так, что *если подгоняемая сейчас своими остатками переменная – i-я, то ее изменчивость, «толщина» облака, - становится равномерна вдоль всех i-1 предшествующих переменных.*



В сочетании с главнокомпонентным методом данная опция обладает способностью ликвидировать не только гетероскедастичность вдоль подгоняемой компоненты, но и гетероскедастичность самой подгоняемой компоненты вдоль предшествующих (т.е. исправляет двойную X-образность). Именно, *если подгоняемая сейчас компонента – i-я, то облако становится «равноутолщено» вдоль нее и вдоль всех предшествующих компонент.* Это означает, что после подгонки последней компоненты облако по всем направлениям многомерного пространства будет иметь «правильную», приближенную к эллипсоидной, форму.

Свойства методов

* *Регрессионный метод.* Единственный из 3-х методов, который не изменяет нелинейным образом первую переменную из списка входящих переменных. Он единственный позволяет вводить прямоугольные матрицы коэффициентов (см. макрос !KO\_FITVAR ниже). Достигнутая гомоскедастичность, если эта опция заказана, носит последовательный характер: каждая переменная гомоскедастична в отношениях с предшествующими в списке, но не обязательно гомоскедастична в отношениях с последующими в списке. С увеличением числа переменных подгонка последних в списке может быть заметно хуже, чем первых; это можно компенсировать инкрементом количества итераций. Если последующие планы исследователя связаны с регрессионным анализом, этот метод может оказаться предпочтительнее других (зависимую переменную лучше поместить в конце списка).
* *Главнокомпонентный метод.* При том же числе итераций как правило дает несколько лучшее достижение целевых коэффициентов, чем регрессионный метод (эта разница практически малозначима). Значения входящих переменных изменяет обычно меньше, чем два других метода, т.е. он наиболее «щадящий» по отношению к данным (что практически может быть важно). Ухудшение подгонки каждой следующей ковариаты малозаметно, т.к. подгоняются не ковариаты, а компоненты, которые потом поворачиваются в ковариаты, «размазывая» тем самым по ним недоподогнанность. Если целью является гомоскедастичность в парных отношениях между всеми переменными, то этот метод предпочтительнее регрессионного. Он также, очевидно, лучше, если данные предназначаются в дальнейшем для факторного анализа.
* *Неитеративный (быстрый) главнокомпонентный метод.* Этот метод, в отличие от итеративных, всегда дает точное достижение целевых коэффициентов. Входящие и итоговые ковариаты пропорционально связаны: каждая входящая переменная есть линейный многочлен всех итоговых, и каждая итоговая есть линейный многочлен всех входящих. Из трех методов, этот изменяет значения переменных обычно сильнее всех и, кроме того, наиболее неравномерно: большие по абсолютной величине значения изменяет относительно сильнее, чем это делают два других метода[[4]](#footnote-4).

***Результаты***

Макрос выводит в новый рабочий файл подогнанные ковариаты. Если в целевой матрице CORR или COV были ряды дополнительных статистик, ковариаты обладают ими. Если таких рядов не было, ковариаты выходят стандартизованными (при CORR матрице) или центрованными (при COV матрице).

При итеративных методах распечатывается информация о степени их подогнанности под целевую матрицу коэффициентов, т.е. насколько мало различаются целевая матрица и матрица, воспроизведенная итоговыми переменными. Этот показатель – «общая недостигнутость» - есть средняя по матрице абсолютная разница между соответствующими внедиагональными элементами целевой и воспроизведенной матриц. Для того чтобы показатель был стандартен независимо от типа целевой матрицы (корреляционная, ковариационная и т.д.), обе сравниваемые матрицы перед вычислением показателя переводятся в корреляционные матрицы.

Кроме недостигнутости, усредненной по всей матрице, распечатываются два постолбцовых ее показателя: усредненная недостигнутость в отношениях данной переменной со всеми предшествующими (эта информация имеет значение при регрессионном методе, показывая падение качества подгонки и тем самым помогая подобрать инкремент) и максимальные в столбцах недостигнутости – наибольшая абсолютная разница между двумя матрицами в данном их столбце. Диагональные элементы целевой и воспроизведенной матриц всегда полностью совпадают, и они не участвуют в вычислениях недостигнутости.

Макрос не выдает сведений о степени достижения гомоскедастичности. Интересующийся исследователь должен сам изучить ковариаты на этот предмет. Гетеро/гомоскедастичность бывает видна на графиках, сведения о ней можно получить из регрессионного анализа.

# Requirements for the matrix

The matrix of target (needed to be achieved) coefficients must be square symmetrical. Presence of columns ROWTYPE\_ (1st) and VARNAME\_ (2nd) is necessary. There should be no other columns besides these and the columns of the matrix body. Variable names as rows of VARNAME\_ must be written in *capital letters* (in SPSS prior version 12) or must *coincide in case* with the names of the variables in the working data file (SPSS versions 12 or higher). The names should not exceed 8 characters.

The matrix type is specified by values of column ROWTYPE\_ which must be written in capital letters. The following matrix types are allowed:

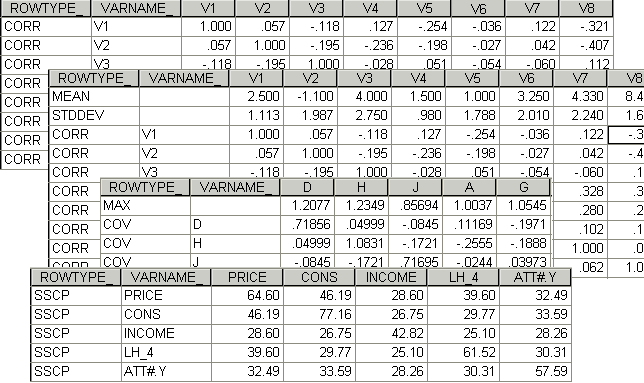
CORR correlation (on the diagonal – ones)

COV covariance (on the diagonal – variances)

SSCP sum of cross-products (on the diagonal – sums of squares)

MSCP mean cross-products (on the diagonal – mean squares), this is the SSCP matrix devided by a sample size N

Matrix CORR or COV may, to a wish of a user, have rows of additional statistics above them; these are statistics to which values the output variables should be brought to. The type of statistic should be notified in capital letters in ROWTYPE\_. Above COV there may be *one* row, any of the following statistics: MIN (minimal value), MAX (maximal value), MEAN (arithmetic mean). Above CORR there may be *two* (not one) rows: any combination of the following 4 stats: MIN, MAX, MEAN, STDDEV (or SD; standard deviation), *the earlier in this list among the two must occupy the 1st row*. Examples of matrices:



From up to down:

* Correlation matrix
* Correlation matrix with rows of additional statistics: mean and st. deviation
* Covariance matrix with row of additional statistic: maximal value
* Sum-of-cross-products matrix.

# Подкоманды

## MATRIX

Целевая матрица, внешний sav-файл – путь/имя в апострофах или кавычках. Пропущенные значения недопустимы в матрице.

## VARS

Переменные рабочего файла данных, которые требуется подогнать под коэффициенты целевой матрицы. Это могут быть собранные в исследовании или созданные случайно переменные. Их можно задать двояко:

*Список имен* - подгонять эти переменные (список можно писать поименно и/или ч-з to). Все эти переменные должны фигурировать среди имен столбцов/рядов целевой матрицы (в ней могут быть и другие имена). Порядок переменных в списке переменных или в файле не обязан совпадать с их порядком в матрице. В SPSS 12 и выше версий имена переменных в файле должны совпадать регистром букв с одноименными в целевой матрице.

ALL - подгонять все переменные рабочего файла. Их число должно равняться числу столбцов/рядов целевой матрицы, но имена там и там могут различаться: соответствующими будут считаться переменные и столбцы/ряды с одинаковым порядковым номером.

Если во входящих переменных есть пропущенные значения, макрос сразу заменит их на валидное значение 0. В ходе подгонки эти значения могут стать ненулевыми. Если замена на нуль исследователя не устраивает, пусть избавится от пропущенных данных заранее.

## METHOD

Specify method of fitting the covariates:

IREGR - Iterative regressional. The immediately being fitted variables are regression residuals.

IPCOMP - Iterative principal component. The immediately being fitted variables are principal components.

QUICK - Quick principal components. No iterative fitting.

## ADJUST

Specify additional requirements (this subcommand is inactive with METHOD=QUICK):

NONE - None. Just attain the target coefficients.

CENTER - (default if omit or unspecify the subcommand) Attain zero mean for the the immediately being fitted variable (with METHOD=IREGR it is residuals; with METHOD=IPCOMP it is 2nd and subsequent components).

HSCED1 - Besides CENTER[[5]](#footnote-5), attain symmetric (as measured by squared deviations) scatter cloud. This is insurance against fan-like heteroscedasticity.

HSCED2 - Besides CENTER, attain equally thickened (as measured by squared deviations) scatter cloud. This is insurance against symmetric heteroscedasticity.

## ITER

В этой подкоманде, которая не действует при METHOD=QUICK, задайте одно или два целых числа: начальное число итераций и, опционально, после него инкремент, на который должно увеличиваться число итераций при переходе к следующей подгоняемой ковариаты/компоненты. Инкремент позволяет в определенной мере компенсировать ухудшение подгонки каждой очередной переменной/компоненты, связанное с «накоплением» недоподогнанности. Если переменных много, рекомендуется выше и инкремент. Инкремент имеет значение в основном при регрессионном методе.

Как начальное число итераций, так и инкремент имеет смысл делать больше при опциях достижения гомоскедастичности, т.к. такое достижение является дополнительным бременем, и если задать слишком мало итераций, ни коэффициенты, ни гомоскедастичность не будут достигнуты в удовлетворяющей пользователя мере. Удовлетворительное число итераций или инкремент нетрудно подобрать, глядя на распечатываемую макросом статистику «недостигнутости». Разумеется, от ITER и от количества переменных прямо зависит время исполнения макроса.

## MSAVE

Опционально: вы можете сохранить воспроизводимую ковариатами матрицу как внешний sav-файл. Для этого укажите путь/имя файла.

### Особые режимы

Макрос не рассчитан на расщепленное состояние файла. Взвешенность игнорируется макросом.

# Некоторые вопросы

*Как исправить гетероскедастичность, не меняя наблюдаемых сил связи.* Надо просто ввести целевую матрицу, в которой коэффициенты равны тем, что наблюдаются между переменными. Задейстровать HSCED1 или HSCED2.

*Если подгонять случайные переменные, то какого распределения они должны быть?* Если нужно, чтобы выходящие ковариаты имели приблизительно нормальное распределение, нужно и входящие переменные брать из нормального или хотя бы симметричного распределения. Это особенно важно для переменной, которую вы хотите в дальнейшем рассматривать как зависимую от остальных ковариат в регрессионном анализе.

*Выдается сообщение, что целевая матрица не грамова.* Любая реальная (полученная из данных) ковариационная или кросс-продуктная матрица всегда является «положительно-(полу)определенной», или «матрицей Грама». Если ваша целевая матрица взята из головы или набита с ошибками, она может не быть такой. Это означает, что при данных диагональных величинах некоторые внедиагональные величины нереалистичны. Вы должны исправить матрицу.

*Выдается сообщение, что целевая матрица вырождена.* Это значит что некоторые ряды/столбцы матрицы однозначно предсказываются другими рядами/столбцами. Макрос не может работать с такой матрицей.

***Формулы подгонки***

Непосредственно подгоняемая переменная X (в регрессионном методе это регрессионные остатки, а в главнокомпонентном методе это заготовки главных компонент) на каждой итерации подправляется по формуле.

i наблюдение в файле (всего n наблюдений).

Vj каждая уже подогнанная раньше чем X переменная (их m).

Сj разница между наблюдаемой в начале итерации суммой кросс-произведений (мерой корреляции) X с Vj и таковой целевой. Т.к. согласно алгоритму все непосредственно целевые коэффициенты = 0, Cj это сама наблюдаемая сейчас сумма кросс-произведений X с Vj.

При ADJUST=NONE:



При ADJUST=CENTER. Перед итерацией X центруется (средняя приводится к 0).:



При ADJUST=HSCED1 или HSCED2. Перед итерацией X центруется.:



Si знак наблюдаемого значения Xi (+ или –). Этот множитель опускается при сочетании IPCOMP + HSCED1.

Wj центрованная переменная Vj – при сочетании IREGR + HSCED1; abs(Vj) – при сочетании IPCOMP + HSCED1; центрованная abs(центрованная Vj) – при сочетании IREGR + HSCED2; центрованная abs(Vj) – при сочетании IPCOMP + HSCED2.

Kj наблюдаемая в начале итерации сумма кросс-произведений X с Wj – при сочетании IPCOMP + HSCED1; abs(X) с Wj – при остальных сочетаниях.

На выходе из каждой итерации сумма квадратов X-переменной приводится к должной: Xi = Xi \* Sqrt( SSнужная / SSнаблюдаемая).

# MACRO !KO\_FITVAR: FITTING VARIABLES TO A RECTANGULAR MATRIX OF COEFFICIENTS

Version 1 (Nov 2005). Tested on SPSS 11, 11.5, 13, 14.

!KO\_fitvar cvars= *v1 to v5* /\*Variables in the working file to get fit; these should be also among the matrix column names

/rvars= *w1 w4 w10* /\*Variables in the working file which to fit to; these should be also among the matrix row names

/matrix= *'d:\exercise\covmx.sav'* /\*Rectangular target matrix or matrix that contains it as submatrix; external file

/adjust= HSCED1 /\*Additional request: NONE, CENTER (center the immediately being fitted variable, default),

/\*HSCED1 (insure against fan-like heterosced-ty), HSCED2 (insure against symmetric heterosced-ty)

/iter= *20*  /\*Number of iterations

/msave= /\*Optional: Path/name of external file to save the reproduced matrix.

Minimal specification CVARS, RVARS, MATRIX, ITER

This macro, a modification of !KO\_FITCOV, likewise transforms variables to achieve the needed values of association coefficients (correlation, covariance, or sums of cross-products). Unlike !KO\_FITCOV, it fits variables to covarying not with each other but with other variables that will not be altered in any way by the macro. The input is (1) one or more variables-to-fit, (2) one or more variables to fit those to, (3) a rectangular matrix of coefficients. The same iterative algorithm as in !KO\_FITCOV is applied, but only regression method is possible here. As in !KO\_FITCOV, there is the option to achieve homoscedasticity (which as well might be the only aim of the macro usage).

EXAMPLE 1.

!KO\_fitvar cvars= v1 v8 to v15 /rvars= w11 w19 /matrix= 'd:\exercise\corrmx.sav' /iter= 20.

* Variables in the file, v1, v8 and through v15, need to be brought to correlations with variables w11 and w19 according to coefficients stored in matrix “corrmx.sav”. Number of iterations 20.

***Результаты***

Макрос выводит в новый рабочий файл подогнанные переменные. Если в целевой матрице CORR или COV были ряды дополнительных статистик, переменные обладают ими. Если таких рядов не было, переменные выходят стандартизованными (при CORR матрице) или центрованными (при COV матрице).

В окно результатов распечатывается информация о степени их подогнанности под целевую матрицу коэффициентов, т.е. насколько мало различаются целевая матрица и матрица, воспроизведенная итоговыми переменными. Этот показатель – «общая недостигнутость» - есть средняя по матрице абсолютная разница между соответствующими элементами целевой и воспроизведенной матриц. Для того чтобы показатель был стандартен, независимо от типа целевой матрицы (корреляционная, ковариационная и т.д.), обе сравниваемые матрицы перед вычислением показателя переводятся в корреляционные матрицы. Кроме общей недостигнутости, распечатываются максимальные в столбцах недостигнутости – наибольшая абсолютная разница между двумя матрицами в данном их столбце.

Макрос не выдает сведений о степени достижения гомоскедастичности. Интересующийся исследователь должен сам изучить связи на этот предмет. Гетеро/гомоскедастичность бывает видна на графиках, сведения о ней можно получить из регрессионного анализа.

# Requirements for the matrix

The matrix must be “assymmetric rectangular”, that is, with rows and columns defined by different lists, not one list, of variables. The variables that need to get fit must correspond to *columns* of matrix, and the variables to associations with which they must be fitted must correspond to *rows* of matrix. Variable names should be no longer than 8 characters. Columns ROWTYPE\_ (1st) and VARNAME\_ (2nd) are required. There should be no other columns besides these and the columns of the matrix body. Variable names in VARNAME\_ must be written in *capital letters* (in SPSS prior version 12) or must *coincide in case* with the names of the variables in the working data file (SPSS versions 12 or higher). Names of variables-columns of matrix must *coincide in case* with the names of the variables in the working data file (SPSS versions 12 or higher).

The matrix type is specified by values of column ROWTYPE\_ which must be written in capital letters. The following matrix types are allowed:

CORR correlation

COV covariance

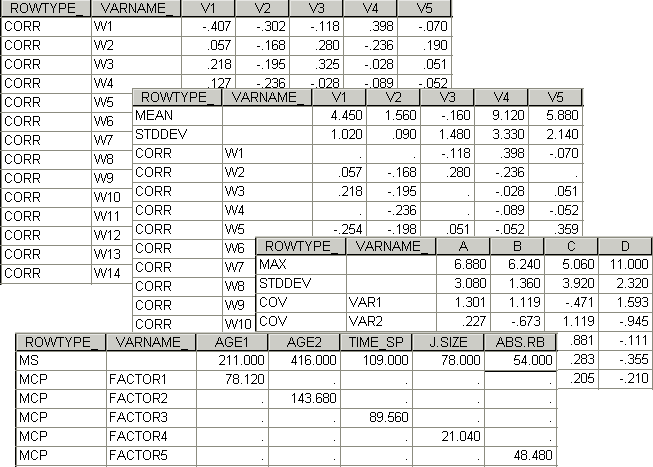
SCP sum of cross-products

MCP mean cross-products; this is the SCP devided by a sample size N

Above the matrix body, there must be a *required* row, also notified in ROWTYPE\_:

* Above SCP matrix must be row SS (sum of squares)
* Above MCP matrix must be row MS (mean square)
* Above COV matrix must be row STDDEV (or SD; standard deviation)
* Above CORR matrix there is no required row

Still above it, for type CORR or COV, there *may* be additional rows of statistics to which values output variables should be brought to. Above COV there may be *one* row, any of the following statistics: MIN (minimal value), MAX (maximal value), MEAN (arithmetic mean). Above CORR there may be *two* (not one) rows: any combination of the following 4 stats: MIN, MAX, MEAN, STDDEV (or SD; standard deviation), *the earlier in this list among the two must occupy the 1st row*. Examples of matrices:



From up to down:

* Correlation matrix
* Correlation matrix with rows of additional statistics: mean and st. deviation
* Covariance matrix with row of additional statistic: maximal value; row STDDEV required
* Mean cross-products matrix; row MS required

В теле целевой матрицы разрешены *пропущенные значения* (system- или user-missing). Пользователь, иначе говоря, может оставлять в теле матрицы пустые ячейки на месте коэффициентов, к которым не требуется подгонять переменные. Так, на рисунке выше 2-я сверху матрица требует подогнать переменную V1 к целевым корреляциям с переменными W2, W3, W5, но не требует подогнать ее к определенным корреляциям с W1 и W4. Корреляции V1 с этими последними исследователя (и макрос) не интересуют, - и они получатся такие, какие получатся. Другой пример: на нижней матрице указаны коэффициенты только на диагонали, только они и составляют целевую матрицу. Задача поэтому чисто двупеременная: создать нужную силу связи AGE1 с FACTOR1, AGE2 с FACTOR2, и т.д. Пропущенные значения *не разрешены* в обязательном ряду и в дополнительном ряду, которые лежат над матрицей. В матрице должно не быть значения: ровно *–999*, т.к. оно зарезервировано макросом для своих нужд.

# Подкоманды

## MATRIX

Целевая матрица, внешний sav-файл – путь/имя в апострофах или кавычках.

## CVARS, RVARS

CVARS это поименный или ч-з to список переменных рабочего файла данных, которые требуется подогнать под целевые коэффициенты. Все эти имена должны присутствовать среди имен столбцов целевой матрицы (которых может быть больше); порядок их в списке или в файле не обязан совпадать с порядком в матрице и не сказывается на результатах. Аналогично, RVARS это список переменных файла, к коварьированию с которыми нужно подгонять CVARS. Их имена должны присутствовать среди рядов целевой матрицы. В SPSS 12 и выше версий имена переменных в файле должны совпадать регистром букв с одноименными в целевой матрице. Переменные RVARS не изменяются макросом и не выдаются обратно на выход. Любой из двух списков может состоять даже из одной переменной.

Пропущенные значения в RVARS недопустимы, а в CVARS допустимы – они будут сразу заменены на валидный нуль. В ходе подгонки эти значения могут стать ненулевыми. Если замена на нуль исследователя не устраивает, пусть он избавится от пропущенных данных заранее.

## ADJUST

Specify additional requirements of fitting:

NONE - None. Just attain the target coefficients.

CENTER - (default if omit or unspecify the subcommand) Attain zero mean for the the immediately being fitted variable (the residuals).

HSCED1 - Besides CENTER, attain symmetric scatter cloud. This is insurance against fan-like heteroscedasticity.

HSCED2 - Besides CENTER, attain equally thickened scatter cloud. This is insurance against symmetric heteroscedasticity.

## ITER

Задайте число подгоночных итераций каждой переменной.

## MSAVE

Опционально: вы можете сохранить воспроизводимую переменными матрицу как внешний sav-файл. Для этого укажите путь/имя файла. В матрице нет пустых ячеек. Коэффициенты, соответствующие пустым ячейкам целевой матрицы, это те, к которым переменные не подгонялись. Эти коэффициенты не участвуют в вычислении показателя недостигнутости.

### Особые режимы

Макрос не рассчитан на расщепленное состояние файла. Взвешенность игнорируется макросом.

1. Общее название класса коэффициентов, о которых идет речь – «скалярные произведения векторов-переменных» или «суммарные кросс-произведения». [↑](#footnote-ref-1)
2. Гетероскедастичность – нежелательное для статистика явление, когда вариативность признака непостоянна вдоль диапазона значений другого признака, с которым он связан. Гомоскедастичность – противоположное, обычно желательное явление – когда вариативность постоянна. [↑](#footnote-ref-2)
3. Если вы не ожидаете от данных симметричной гетероскедастичности или не заботитесь о ней, а преследуете только веерную, то лучше воспользоваться все же опцией «страхование от веерной гетероскедастичности» как более подходящей к этому случаю. [↑](#footnote-ref-3)
4. Известен еще неитеративный, тоже точный метод с использованием корня Холецкого целевой матрицы, однако, по моим наблюдениям, он еще сильнее изменяет значения входящих переменных, чем неитеративный главнокомпонентный способ, т.е. уступает ему, если стоит задача – *подправить* входящие переменные под целевую матрицу. [↑](#footnote-ref-4)
5. В условиях HSCED1 и HSCED2, в методе IPCOMP 1-я компонента тоже центруется, как и остальные [↑](#footnote-ref-5)