***Categorical into Contrast***

SPSS macros by Kirill Orlov

kior@akado.ru, ttnphns@gmail.com

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

All rights reserved

*Категориальные переменные в контрастные.* Создает из категориальных переменных контраст-переменные нескольких типов и их взаимодейственные переменные. Контраст-переменные нужны прежде всего тогда, когда нужно анализировать влияние качественных факторов методами, рассчитанными на количественные входящие (напр. линейная регрессия).

*Прочтите «*[*О SPSS макросах*](https://www.spsstools.net/ru/KO-aboutmacros)*» что они такое и как их запускать.*

*Ошибка “Protected directory”.* Некоторые из макросов, описанных в текущем документе, пишут временные файлы на жесткий диск. Если вы не обладаете полными правами Администратора вашего компьютера, это может вызвать ошибку, сообщающую среди прочего: *“SPSS Statistics cannot access a file... specifies a protected directory...”* и значащую, что дефолтная директория, какую макрос хочет использовать, защищена на вашем ПК. Чтобы решить эту проблему, в окне синтаксиса скомандуйте: CD 'myfolder'., где 'myfolder' есть путь/имя некоторой папки, куда вам разрешено сохранять файлы.

**МАКРОС !KO\_CATCONT: ПЕРЕКОДИРОВКА КАТЕГОРИАЛЬНЫХ ПЕРЕМЕННЫХ В КОНТРАСТ-ПЕРЕМЕННЫЕ**

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

!KO\_catcont factors= v1 v2 /\*Одна или несколько категориальных переменных (факторов) поименно

/types= DEV SIM /\*Для каждого фактора тип контраста: IND, SIM, DEV, WDEV, HEL, WHEL,

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

/spacing= /\*Для POL WPOL, интервализация: EQUAL (тж п/у) или ASIS

/inter= ALL /\*Если нужны взаимодействия: целое число или UP целое число или ALL

/\*напр. 3 значит "все 3-way", UP 3 значит "все до 3-way"

/seq= EFFECT /\*Порядок среди переменных-взаимодействий: комбинаторный (COMBIN, тж п/у)

/\*или по эффекту (EFFECT)

/id= id /\*Опционально: переменная-индикатор наблюдений

/print= /\*Распечатка: полная (LONG, тж п/у), краткая (SHORT), минимальная (NONE)

/savelmx= 'd:\exercise\lmx.sav' /\*Вывести составленную L-матрицу для факторов и

/\*(если заказано) взаимодействий: путь/имя файла.

Минимум надо задать FACTORS, TYPES.

Что нового июль 2024.

Добавлены взвешенные типы WDEV, WHEL, WDIF.

Макрос создает из категориальных или количественных дискретных переменных контраст-переменные нужного типа и опционально их взаимодействия. Он также выдает матрицы контраст-коэффициентов (**L**-матрицы). Если фактор состоит из *k* градаций, то из него образуется *k*-1 контраст-переменных. Выходящий массив, содержащий контраст-переменные, безымянен. Используйте Data > Merge Files > Add Variables для сшития его с входящим массивом.

Контраст-переменные это числовые, количественные переменные, в которые становится необходимо перекодировать межиндивидные факторы (категориальные переменные), чтобы факторы математически смогли сыграть роль предикторов в разных линейных и нелинейных предсказательных моделях. Из фактора создается набор контрастных переменных, совокупно замещающих собой межиндивидный фактор. Одним из типов контраст-переменных являются знаменитые фиктивные переменные.

Каждая контраст-переменная это воплощение определенного *контраста*. Контраст это взвешенная линейная комбинация градаций фактора. Благодаря тому, что веса суммируются в 0, контраст есть инструмент сравнения между выбранными отдельными градациями или смешениями таковых. А полиномиальные контрасты, относящиеся к фактору как к дискретному количественному, есть инструмент исследования линейных и нелинейных зависимостей от такого фактора. Упомянутые веса в контрасте называются контраст-коэффициентами, и они прямо связаны со значениями контраст-переменной, воплощающей данный контраст.

SPSS-процедуры, такие как General Linear Model, Logistic Regression и многие иные, берущие межиндивидные факторы, предлагают для них разные типы контрастов на выбор (эти типы называются еще categorical encoding schemes). От выбора зависит назначение программой тех или иных контраст-коэффициентов и через это – зависит *смысл* параметров в предсказательной модели. Однако упомянутые гибкие процедуры в SPSS не презентуют пользователю соответствующие контраст-переменные как данные: всё происходит за сценой. Макрос !KO\_CATCONT имеет целью именно дать пользователю сами контраст-переменные как данные, физически. Имея контраст-переменные в руках, можно использовать их в качестве континуальных регрессоров (ковариат) в таких процедурах, которые не рассчитаны на категориальные предикторы (факторы), а признают только количественные предикторы. К примеру, этим путем можно в процедуре Linear Regression получить некоторые результаты, связанные с факторами, которые получаемы в One-way Analysis of Variance или в General Linear Model. Вы можете использовать одни и те же контраст-переменные в моделях с интерцептом и без интерцепта, в моделях с ковариатами и без.

Ограничения. До 26 факторов во входящих. До 36 градаций в факторе. До 4-сторонних взаимодействий. Если задана п/к SAVELMX, то – до 4-х факторов во входящих и порядок взаимодействий не выше 3.

Именование выходящих переменных

Имя контраст-переменной состоит из символа и индекса. Символ обозначает фактор, из которого образована переменная, а индекс это порядковый номер контраст-переменной от данного фактора. Символы для факторов (в порядке их задания в FACTORS): *a*, *b*, ..., *z*. Индексы: *1*, *2*, ..., *9*, *0*, *A*, *B*, ..., *Z*. Так, первая контраст-переменная для первого фактора будет названа *a1*, вторая для этого же фактора будет *a2*, и так далее. Те же самые индексы, что используются для имен контраст-переменных, используются для нумерации градаций фактора (по возрастанию кодов в факторе). Это удобно, поскольку контраст-переменные мультиномиальных типов (см. п/к TYPE) отвечают градациям. У полиномиальных типов контраст-переменные отвечают не градациям, а степенным компонентам, для чего индексы *1*, *2*, ..., опять же, удобны.

Контраст-переменные взаимодействий наследуют эту систему поименования, являясь сцеплением. *a1b3*, например, есть 2-сторонее взаимодействие “a1 by b3”, или “a1\*b3”.

Факторные переменные. Помимо контраст-переменных выходящий массив содержит сами входящие факторы, только как текстовые переменные и в вышеописанной нотации. Факторы носят имена *a*, *b*, ..., а их градации (значения) есть индексы *1*, *2*, ..., и т.д. Обратите внимание, что такие значения будут всегда, как бы ни были кодированы входящие факторные переменные.

ДАННЫЕ, использованные в примерах.

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 это факторы (несбалансированные; если вместе с F4, то и с пустыми клетками в дизайне). G1, G2, G3, G4 это еще набор факторов (сбалансированные, если без G4).
* Факторы F2 и G2 кодированы количественными дискретными значениями, а не условными кодами. Это может играть роль только при SPACING=ASIS.
* Y это количественная зависимая переменная. Y2 – дихотомическая зависимая переменная. COVAR1 это количественная ковариата. Зависимые переменные и ковариаты не нужны для работы макроса; они используются в некоторых примерах этого документа.

ПРИМЕР 1.

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

* Фактор F1 заказано раскодировать в набор контраст-переменных типа DEV, фактор F2 – тоже в тип DEV. Фактор F3 заказано раскодировать в контраст-переменные, отвечающие пользовательской контраст-матрице, последняя указана как внешний файл.
* F1 (2 градации) даст контраст-переменную a1. F2 (3 градации) даст контраст-переменные b1, b2. F3 (4 градации) даст контраст-переменные c1, c2, c3. Контраст-переменных на 1 меньше, чем есть градаций в факторе. В каждом факторе последняя градация принимается за reference. Смысл контраст-переменных типа DEV описан в п/к TYPES.
* Контраст-переменные выведены как новый безымянный массив данных. Три исходных фактора скопированы туда же, но под именами a, b, c и в текстовом формате; коды им макрос назначает сам (это всегда индексы 1, 2, ..., 9, 0, A, B, ..., Z).

***Подкоманды***

**FACTORS**

Поименно числовые переменные (факторы), минимум одна, из которых надо получить контраст-переменные. Каждое уникальное валидное значение в переменной считается факторной градацией. Значения дискретны, хотя не обязаны быть целыми. Если SPACING=ASIS, они должны быть положительными. Имена могут повторяться в списке. Порядок градаций в факторе отвечает возрастающему сортировочному порядку значений.

Пропущенные значения макрос отсеивает списочно (listwise), т.е. из всех факторов (и ID-переменной) заодно.

**TYPES**

Список в соответствие FACTORS: укажите для каждого фактора тип контраста, в который хотите перекодировать данный фактор. Вы можете также написать один тип со словом ALL после него – тогда ко всем факторам будет применен данный тип. Слово ALL нельзя употреблять, если тип пользовательский.

Фактору из *k* градаций соответствует *k*-1 контрастов, и каждому контрасту отвечает своя выдаваемая контраст-переменная. Таким образом, набор из *k*-1 числовых переменных будет создан из каждого фактора. Таблица значений контраст-переменных (*k* градаций × *k*-1 контрастов) называется **C-матрицей** (= кодировочной матрицей, basis matrix, summarized design matrix). Ей соответствует **L-матрица** (контраст-матрица, матрица контраст-коэффициентов) размерностью *k*-1 контрастов × *k* градаций. Макрос показывает обе матрицы при PRINT=LONG; причем **L**-матрицу показывает транспонированной.

Математическое отношение между матрицами: **L+ =** **C+-1** и **C+ = L+-1**, где **C+** это **C** с добавленным 1-м столбцом из единиц и **L+** это **L** с добавленным 1-м рядом из 1/*k*. Одну матрицу из другой можно получить также псевдообращением, однако это верно не для всех контраст-типов.

Число контрастов на 1 меньше числа значений в факторе по причине мультиколлинеарности: определитель SSCP-матрицы (т.е. матрицы **X´X**) набора контраст-переменных **X** стал бы равен 0, если бы число создаваемых контраст-переменных было *k*, а не *k*-1[[1]](#footnote-1).

Примеры **C** и **L** матриц. Пусть есть фактор *A* с тремя градациями (*1*, *2*, *3*).

Кодировочная матрица **C** и контраст-матрица **L** для типа Indicator (dummy):

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Контраст-переменные | |  |  | Градации | | |
| Градации | a1 | a2 |  | Контрасты | *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 |  |  |  |  |  |

Кодировочная матрица **C** и контраст-матрица **L** для типа Deviation:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Контраст-переменные | |  |  | Градации | | |
| Градации | a1 | a2 |  | Контрасты | *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 |  |  |  |  |  |

В **L**-матрице виден смысл каждого контраста (контраст-переменной), а именно, что будет *означать* регрессионный параметр при данной контраст-переменной, если ввести весь набор контраст-переменных, полученных из фактора, в качестве множественных количественных предикторов (ковариат). В каждом ряду **L**-матрицы коэффициенты суммируются в 0, и «контрастируемые» положительные vs отрицательные коэффициенты показывают, какие градации с какими градациями сравниваются данным контрастом (и следовательно, соответствующей контраст-переменной). Контраст-коэффициенты это веса в линейной комбинации градаций или значений фактора: ненулевые веса показывают, что с чем сопоставляется. Осмотр контраст-коэффициентов *не* позволяет предсказать результат – величину сравнения (= значение регрессионного параметра), но позволяет понять *содержание* его.

Типы контрастов и соответствующий каждому типу смысл регрессионных параметров при контраст-переменных приведены ниже.

a) **Мультиномиальные типы**. Фактор принимается за категориальную переменную. Каждый из *k*-1 контрастов соответствует одной градации фактора, сравнивая ее с определенной другой градацией или с микстом градаций.

IND - тип **индикаторный** (indicator). Это так называемые фиктивные (dummy) переменные; они двоичные (1 vs 0). Регрессионный параметр при *i*-й контраст-переменной будет сравнением функции предсказания между: *i*-й градацией и *k*-й градацией. Интерцепт регрессии будет функцией предсказания в *k*-й градации. *k*-я градация reference, для нее нет контраст-переменной и ее параметр логически есть 0. Контрасты этого типа неортогональны; при сбалансированном факторе контраст-переменные коррелируют с r = -1/(*k*-1). Тип IND следует толковать несколько иначе, чем тут сказано, когда в модели есть межфакторные взаимодействия (см. ПРИМЕР 8).

SIM - тип **простой** (simple). **C**-матрица данного типа контраста совпадает с **C**-матрицей типа IND после центрации столбцов последней. **L**-матрицы их одинаковы. SIM эквивалентен IND и осуществляет те же самые сравнения, отличаясь от IND лишь интерцептом: интерцепт регрессии будет функцией предсказания, невзвешенно усредненной между всеми градациями. Контрасты неортогональны; при сбалансированном факторе контраст-переменные коррелируют с r = -1/(*k*-1). SIM эквивалентен IND, однако, только в отсутствие в модели межфакторных взаимодействий. См. ПРИМЕР 8, поясняющий различия между SIM и IND в присутствии таких взаимодействий. Когда есть взаимодействия, в обычном случае надо использовать SIM, не IND.

DEV - тип **отклонений** (deviation aka effect). Это троичные переменные (1, 0, -1; а при *k*=2 только 1, -1). Регрессионный параметр при *i*-й контраст-переменной будет сравнением функции предсказания между: *i*-й градацией и невзвешенным усреднением между всеми градациями. Интерцепт регрессии будет этой функцией предсказания, невзвешенно усредненной между всеми градациями (как в SIM). *k*-я градация reference, для нее нет контраст-переменной и ее параметр равен взятой с обратным знаком сумме остальных *k*-1 параметров. Контрасты этого типа неортогональны; при сбалансированном факторе контраст-переменные коррелируют с r = 0.5. Данный тип часто выступает дефолтным для межиндивидных факторов в ANOVA.

WDEV - **взвешенный** тип **отклонений**. Эта версия предыдущего учитывает факт возможной несбалансированности. Регрессионный параметр при *i*-й контраст-переменной будет сравнением функции предсказания между: *i*-й градацией и взвешенным (размером градации) усреднением между всеми градациями. Интерцепт – как при DEV. В условиях сбалансированного фактора этот тип совпадает с DEV.

HEL - тип **Хелмерта** (Helmert). Регрессионный параметр при *i*-й контраст-переменной будет сравнением функции предсказания между: *i*-й градацией и невзвешенным усреднением между последующими градациями. Интерцепт регрессии будет функцией предсказания, невзвешенно усредненной между всеми градациями. Для *k*-й градации нет контраст-переменной. Контрасты этого типа ортогональны, поэтому если фактор сбалансированный, контраст-переменные получаются некоррелированны.

WHEL - **взвешенный** тип **Хелмерта**. Эта версия предыдущего учитывает факт возможной несбалансированности. Регрессионный параметр при *i*-й контраст-переменной будет сравнением функции предсказания между: *i*-й градацией и взвешенным (размером градации) усреднением между последующими градациями. Интерцепт – как при HEL. Контраст-переменные получаются всегда некоррелированны. Однако контрасты в **L**-матрице ортогональны только в условиях сбалансированного фактора (тогда этот тип совпадает с HEL).

DIF - тип **разниц** (difference), или **обратный контраст Хелмерта** (reverse Helmert). Это перевернутый предыдущий тип. Для 1-й градации нет контраст-переменной. Регрессионный параметр при *i*-й контраст-переменной будет сравнением функции предсказания между: *i*+1-й градацией и невзвешенным усреднением между предшествующими градациями. Интерцепт регрессии будет функцией предсказания, невзвешенно усредненной между всеми градациями. Контрасты этого типа ортогональны, поэтому если фактор сбалансированный, контраст-переменные получаются некоррелированны.

WDIF - **взвешенный** тип **разниц**. Эта версия предыдущего учитывает факт возможной несбалансированности. Регрессионный параметр при *i*-й контраст-переменной будет сравнением функции предсказания между: *i*+1-й градацией и взвешенным (размером градации) усреднением между предшествующими градациями. Интерцепт – как при DIF. Контраст-переменные получаются всегда некоррелированны. Однако контрасты в **L**-матрице ортогональны только в условиях сбалансированного фактора (тогда этот тип совпадает с DIF).

REP - тип **повторительный** (repeated). Регрессионный параметр при *i*-й контраст-переменной будет сравнением функции предсказания между: *i*-й градацией и *i*+1-й градацией. Интерцепт регрессии будет функцией предсказания, невзвешенно усредненной между всеми градациями. Для *k*-й градации нет контраст-переменной. Контрасты этого типа неортогональны; при сбалансированном факторе контраст-переменные коррелируют положительно с разной силой.

*k*-я градация в типах IND, SIM и DEV/WDEV именуется reference, и ее выбор произволен (на пользователе). Если вы хотите назначить reference-градацией не *k*-ю градацию, то просто перекодируйте входящий фактор: припишите самый высокий код той градации, которую желаете сделать reference. Она станет *k*-й и reference. Типы HEL/WHEL, DIF/WDIF и REP упорядочивают градации, поэтому reference-градация при этих типах «фиксирована» и должна планироваться заранее при кодировке фактора.

Мультиномиальные типы, кроме WDEV, WHEL, WDIF ­– невзвешенные: **L** и **C** матрицы вычисляются одинаково как при сбалансированном, так и несбалансированном факторе.

Мультиномиальные типы контрастов существуют в SPSS Statistics в GLM/UNIANOVA, MANOVA, LOGISTIC REGRESSION и некоторых других процедурах. Тип IND присутствует не везде (по ряду причин); почти всегда вместо него годится тип SIM.

ПРИМЕР 2. Тип Indicator.

dataset name data.

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

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

execute.

dataset name merged.

* Макрос раскодирует факторы G1 (2 градации), G3 (3 градации) в тип IND каждый.
* Выходящий массив сшивается с входящим; дальнейшее делается на сшитом массиве MERGED:

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

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

* Эти две команды эквивалентны, они делают регрессию Y по контраст-переменным как количественным предикторам.
* Регрессионный параметр при b1, например, равен разнице : , поскольку соответствующие коэффициенты в L-матрице фактора B (т.е. G3) есть 1 0 -1. Это есть сравнение градации B=1 с reference-градацией B=3 (контраст-тип IND). Разумеется, поскольку в нашей модели есть еще второй фактор, и он с 2-мя градациями, то каждый член вышеприведенной разницы – это число, которое есть средняя из 2-х чисел (одно на каждую градацию второго фактора). Т.е. : составное: это средняя двух чисел : ; и аналогично – в отношении : .
* Интерцепт при типе IND равен reference градации. Поскольку факторов в модели два, то интерцепт есть на пересечении их reference градаций; т.е. равен .

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

* ANOVA Y по факторам A и B (т.е. = G1 и G3), модель главных эффектов. Обратите внимание, что параметры, выдаваемые UNIANOVA, совпадают с регрессионными параметрами, полученными выше на контраст-переменных. Это оттого, что процедура UNIANOVA (т.е. GLM) внутренне основана на раскодировке факторов в Индикаторный тип.

ПРИМЕР 3. Типы Deviation и Repeated.

dataset name data.

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

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

execute.

dataset name merged.

* Макрос раскодирует фактор F2 (3 градации) в тип DEV и фактор F3 (4 градации) в тип REP.
* Выходящий массив сшивается с входящим; дальнейшее делается на сшитом массиве 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.

* Эти две команды эквивалентны, они делают регрессию Y по контраст-переменным как количественным предикторам.
* Регрессионный параметр при a1, например, равен разнице : , поскольку соответствующие коэффициенты в L-матрице фактора A (т.е. F2) есть .6667 -.3333 -.3333. И эта разница равна: , что есть сравнение градации A=1 с невзвешенной средней всех трех градаций (контраст-тип DEV). Разумеется, поскольку в модели есть еще второй фактор (как было и в предыдущем примере), и он с 4-мя градациями, то каждый член вышеприведенной разницы – это число, которое есть средняя из 4-х чисел (одно на каждую градацию второго фактора).
* Регрессионный параметр при b2, например, равен разнице : , поскольку соответствующие коэффициенты в L-матрице фактора B (т.е. F3) есть 0 1 -1 0. Это сравнение градации B=2 с градацией B=3 (контраст-тип REP). Разумеется, поскольку в модели есть еще второй фактор, и он с 3-мя градациями, то каждый член вышеприведенной разницы – это число, которое есть средняя из 3-х чисел (одно на каждую градацию второго фактора).
* Интерцепт при типах DEV и REP это средняя всех уникальных значений .

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

* ANOVA Y по факторам A и B (т.е. = F2 и F3), модель главных эффектов. Для фактора A заказан тип Deviation, а для фактора B тип Repeated. Результаты – оценки контрастов в таблице “Contrast Results (K Matrix)” – совпадают с регрессионными параметрами, полученными выше на контраст-переменных.
* Обратите внимание, параметры в таблице “Parameter Estimates” не совпадают с этими результатами. Это потому, что, как сказано в ПРИМЕРЕ 2, в GLM эта таблица всегда отвечает кодировке Indicator. Однако в процедуре MANOVA, где алгоритм более прямой, чем в GLM, параметры всегда отвечают тому типу контраста, который вы заказали:

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#.

* Тот же анализ посредством процедуры MANOVA (A# и B# это просто числовые копии A и B; вы могли бы задать вместо них входящие F2 F3 с тем же успехом). Параметры (таблица “Estimates”) совпадают с оценками контрастов в UNIANOVA и регрессионными параметрами, полученными на контраст-переменных.
* Подробный отчет PRINT= DESIGN(ALL) показывает в числе прочего кодировочные матрицы (Basis matrices), которые показывает и !KO\_CATCONT. Матрица “Between-Subjects Design Matrix” это резюме значений всех контраст-переменных вместе, т.е. является этим:

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).

* (кроме 1-го столбца, который в выдаче MANOVA соответствует интерцепту).

ПРИМЕР 3a. Тип Deviation невзвешенный и взвешенный.

dataset name data.

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

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

execute.

dataset name merged.

* Макрос раскодирует фактор F3 (4 градации) один раз в тип DEV (переменные a1 a2 a3) и второй раз в тип WDEV (переменные b1 b2 b3). Выходящий массив сшивается с входящим; дальнейшее делается на сшитом массиве MERGED:

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

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

* Первая регрессия использует a1 a2 a3 в качестве предикторов. Регрессионный параметр при типе DEV – это разница средних между градацией и невзвешенной общей средней всех градаций. Параметр при a1, например, есть разница : .
* Вторая регрессия использует b1 b2 b3 в качестве предикторов. Регрессионный параметр при типе WDEV – это разница средних между градацией и взвешенной общей средней всех градаций, т.е. total средней выборки. Параметр при b1, например, есть разница : .

b) **Полиномиальные типы**. Фактор принимается не за категориальный фактор, а за переменную количественную (см. подкоманду SPACING). Порождаются линейный, квадратный, и более высокой степени контрасты для изучения линейной и криволинейной зависимостей от такого мерного предиктора. 1-й контраст отвечает линейной зависимости от него, 2-й контраст – квадратной зависимости сверх линейной, 3-й контраст – кубической зависимости сверх[[2]](#footnote-2) линейной и квадратной, и так далее. *i*-я контраст-переменная соответствует, таким образом, не *i*-й градации (как у мультиномиальных типов), а *i*-й степени, степени нелинейности. Порождается *k*-1 контрастов, которые все вместе вполне заменяют собой фактор. Макрос предлагает два типа – невзвешенный и взвешенный. Эти два типа тождественны в случае, если фактор сбалансированный (равноразмерные группы), они различаются в ситуации несбалансированного фактора.

POL - **невзвешенный полиномиальный** (unweighted polynomial). Этот тип игнорирует факт несбалансированности: уровням фактора придается равная значимость. Невзвешенные контрасты всегда ортогональны в **L**-матрице (а **C** совпадает с **L**´). Контраст-переменные получаются некоррелированны только в условиях сбалансированного фактора.

WPOL - **взвешенный полиномиальный** (weighted polynomial). Этот тип учитывает факт несбалансированности, и контраст-переменные получаются всегда некоррелированны. Однако контрасты в **L**-матрице ортогональны только в условиях сбалансированного фактора.

Взвешенный полиномиальный контраст существует в SPSS Statistics в One-way Analysis of Variance (ONEWAY). Невзвешенный тип существует также в GLM/UNIANOVA, MANOVA, LOGISTIC REGRESSION и ряде других процедур.

ПРИМЕР 4. Полиномиальные контрасты.

dataset name data.

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

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

execute.

dataset name merged.

* В этом примере один и тот же фактор F3 (4 уровня, фактор несбалансированный) раскодируется один раз в невзвешенный полиномиальный тип (получится набор переменных a1 a2 a3), второй раз в взвешенный полиномиальный тип (получится набор переменных b1 b2 b3). Переменные последнего типа некоррелированны, несмотря на несбалансированность.
* Выходящий массив сшивается с входящим; дальнейшее делается на сшитом массиве MERGED:

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

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

* Регрессия на контраст-переменных (невзвешенный полиномиальный тип) дает те же результаты, что ANOVA с заказом полиномиального типа для фактора F3.
* В модели с предикторами a1, a2, a3 регрессионный коэффициент при a1 выражает линейный тренд, коэффициент при a2 выражает квадратный тренд (сверх линейного), коэффициент при a3 выражает кубический тренд (сверх линейного и квадратного).

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

oneway y by f3 /polynomial= 3.

* Регрессия на контраст-переменных (взвешенный полиномиальный тип) дает те же результаты, что ANOVA с заказом взвешенного полиномиального типа для фактора F3. Взвешенный тип в SPSS Statistics можно получить в процедуре ONEWAY. Эта процедура не показывает в Anova-таблице значения контрастов (т.е. оценки параметров), но показывает их *p*-значения. И они те же, что *p*-значения параметров в регрессии. В Anova-таблице вы видите как взвешенный полиномиальный тип, так и невзвешенный полиномиальный тип (результаты для последнего мы уже получили выше).

c) **Пользовательский**, или специальный контраст. Инлайн-ввод или имя файла.

*матрица/файл* - **пользовательский контраст**. Введите свою матрицу контраст-коэффициентов (**L**-матрицу) для фактора, размером *k*-1 контрастов × *k* градаций. В норме сумма коэффициентов в каждом контрасте должна быть нулевой.

Если ввод инлайн, то поместите матрицу в фигурные скобки {}; элементы ряда разделяются запятой, ряды разделяются точкой-с-запятой. Матрица пишется по-рядно. Перенос на новую строку – произволен. Пример матрицы с *k*=4:

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

0, 2, -1, -1;

0, 0, 1, -1 }

Она же одной строкой: {3, -1, -1, -1; 0, 2, -1, -1; 0, 0, 1, -1}

Элементы вводимой матрицы не обязаны быть целыми, но должны быть числами, не выражением (*не* разрешается, в частности, вводить натуральные дроби, к примеру: -1/3).

Если же указываете имя .SAV файла (в кавычках или апострофах), то его данные должны быть подобным *k*-1 × *k* массивом числовых коэффициентов; имена столбцов роли не играют.

ПРИМЕР 5.

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

* Фактору G2 (3 градации) заказана пользовательская контраст-матрица (2×3), введенная инлайн. Фактору G3 (3 градации) заказана контраст-матрица (2×3), внешний файл; эта матрица была: {-1,0,1;-0.333333,-0.333333,0.666666}. Фактору G4 заказан тип Helmert.
* Затребовано создать и переменные-взаимодействия.

Ортогональность контрастов. Выше было отмечено, что контрасты типов HEL, DIF, POL – ортогональные[[3]](#footnote-3). *Ортогональными* контрастами называют такие, для которых сумма произведений соответствующих элементов у каждой пары рядов в **L**-матрице равна 0; другими словами, **LL´** есть диагональная матрица. Когда контрасты ортогональны, то в условиях сбалансированного фактора контраст-переменные не коррелируют между собой, поэтому их индивидуальные эффекты не будут перекрываться и станут суммироваться точно в общее предсказание всем их набором. Если же фактор несбалансированный или если контрасты не ортогональные, то контраст-переменные коррелируют, поэтому их индивидуальные эффекты будут перекрываются и суммарно окажутся меньше общего предсказания их набором. Так имеет место при одновременном вводе всех переменных набора как предикторов (что соответствует Type III SS aka *unique* method). Вы можете вводить контраст-предикторы последовательно (Type I SS aka *sequential* method), но следует как правило использовать весь набор: контраст переменные подразумевают друг друга, т.к. вместе представляют единый фактор и исчерпывают его.

Контрасты при наличии взаимодействий. Когда в модели есть взаимодейственные статьи, смысл контраста, принадлежащего фактору (т.е. главному эффекту), уточняется с учетом присутствия взаимодействий. Смысл контраста не меняется фундаментально, а конкретизируется с поправкой на взаимодействия. См. ПРИМЕР 8, 9. Особо отметим, что в условиях присутствия взаимодействия смысл контрастов типов IND и SIM, базово похожих, конкретизируется неодинаково.

**SPACING**

Эта подкоманда действует только для полиномиальных контрастов: TYPES упоминает кл. слово POL или WPOL. По умолчанию и при SPACING=EQUAL градации фактора принимаются разделенными равными интервалами, реальная кодировка входящих факторов, которым назначен полиномиальный тип, не играет роли. Т.е. фактор берется как порядковый и затем уровням приписываются количественные значения («метрики») 1, 2, 3...

При SPACING=ASIS кодировка входящих факторов, которым назначен полиномиальный тип, берется как есть, т.е. фактор используется как готовая мерная переменная с существующими в ней соотношениями интервалов между уровнями. Только соотношение интервалов, не величина значений, играет роль. Значения должны быть положительными.

ПРИМЕР 6.

dataset name data.

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

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

execute.

dataset name merged.

* Макрос раскодирует факторы G3 и G2 в тип POL. SPACING=ASIS требует относиться к значениям входящих G3, G2 буквально, как к интервальным.
* Выходящий массив сшивается с входящим; дальнейшее делается на сшитом массиве 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.

* Эти две команды эквивалентны. Первая делает регрессию Y по контраст-переменным, полученным из факторов. Вторая делает ANOVA, раскодирующий факторы в полиномиальный тип. Регрессионные параметры в первой команде совпадают с оценками контрастов в таблице “Contrast Results (K Matrix)” второй команды.
* Заметьте, что в задании второй команды, в подкомандах CONTRAST, затребованы пользовательские интервалы между градациями в факторе. В данном случае указаны сами значения входящих факторов как они есть. Такое задание соответствует тому, что мы сделали при создании контраст-переменных, указав SPACING=ASIS в !KO\_CATCONT.

**INTER**

По умолчанию макрос создает только контраст-переменные факторов (т.е. главных эффектов). Если вам нужны и контраст-переменные, являющие взаимодействия факторов, когда факторов больше одного, то затребуйте их здесь. Эти переменные создаются перемножением факторных контраст-переменных (сами факторные контраст-переменные не зависят от подкоманды INTER). Высший порядок, который вы вправе заказать – 4-сторонние взаимодействия.

Укажите порядок целым числом от 2 до 4. Ключевое слово UP перед числом означает «все взаимодействия вплоть до» этого порядка. Например, UP 3 есть заказ на 2-сторонние и 3-сторонние взаимодействия. Вы также можете написать INTER=ALL, что значит «все возможные взаимодействия» (но не выше порядка 4). INTER=2 и INTER= UP 2 эквивалентны, т.к. 2 это самый малый порядок взаимодействий.

Если в факторах много градаций, заказ взаимодействий существенно замедляет работу и нагружает ресурс оперативной памяти. !KO\_CATCONT делает почти всю работу в RAM-памяти.

ПРИМЕР 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.

* Макрос раскодирует факторы G1, G2, G3, G4 соответственно в заказанные типы SIM, IND, SIM, DEV. Из контраст-переменных затребовано создать контраст-переменные взаимодействий факторов вплоть до 3-сторонних.
* Выходящий массив сшивается с входящим; дальнейшее делается на сшитом массиве 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.

* Логистическая регрессия Y2 по контраст-переменным, включая взаимодейственные, как количественным предикторам. Ковариата COVAR1 также включена в модель.

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.

* Логистическая регрессия Y2 по факторам A, B, C, D (т.е. = G1, G2, G3, G4), модель с главными и эффектами и взаимодействиями до 3-сторонних; ковариата COVAR1 также присутствует. Заказанные типы контрастов, т.е. способы раскодировки факторов: Simple, Indicator, Simple, Deviation. Результаты – регрессионные параметры – совпадают с таковыми, полученными выше на контраст-переменных.

**SEQ**

Эта подкоманда сказывается только на последовательности, в какой идут в выдаче контраст-переменные взаимодействий. По умолчанию и при SEQ=COMBIN это комбинаторный порядок. При SEQ=EFFECT все переменные, составляющие эффект, собраны вместе. Например, если есть факторы A (3 градации), B (3 градации) и C (2 градации), то комбинаторный порядок контраст-переменных 2-сторонних взаимодействий такой:

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

А порядок тех же переменных при SEQ=EFFECT:

a1b1 a1b2 a2b1 a2b2 *a1c1 a2c1* b1c1 b2c1

Переменные, составляющие эффекты A\*B, A\*C, B\*C, не разлучены.

**ID**

Опциональная числовая переменная-идентификатор наблюдений, укажите имя. Т.к. макрос отсеивает невалидные наблюдения списочно, в выходящем массиве может случиться меньше наблюдений, чем во входящем. ID-переменная позволит правильно сшить входящий массив с выданными макросом переменными.

**PRINT**

Отчет в окно результатов:

LONG - (тж. по умолчанию/незаданию) полный отчет, включая **C**- и **L**-матрицы для каждого фактора, проверка ортогональности контрастов и т.д.

SHORT - краткий отчет: информация о факторах: их символы, градации, частоты; определитель корреляционной матрицы контраст-переменных[[4]](#footnote-4).

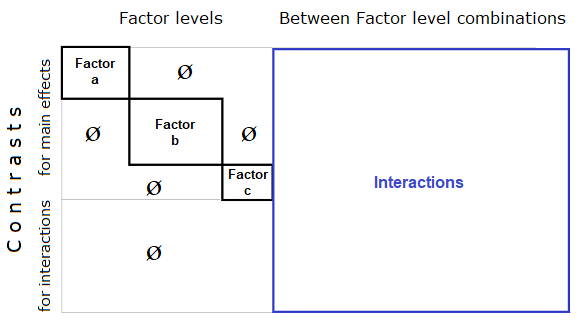
NONE - отменить отчет, кроме информации о символах факторов.

**SAVELMX**

Опциональная подкоманда составить и сохранить **L**-матрицу (матрицу контраст-коэффициентов) для всех факторов вместе и их взаимодействий (если последние были заказаны в INTER). Вы можете впоследствии ввести эту **L**-матрицу (целиком или избранные ее ряды, контрасты) в такие процедуры как GLM/UNIANOVA, ONEWAY, MANOVA, LOGISTIC REGRESSION, COXREG, MIXED и другие. В этих процедурах существуют подкоманды LMATRIX или CONTRAST или TEST или EMMEANS с опцией указания пользовательских контрастов через синтаксис. См. ПРИМЕР 10, где **L**-матрица вводится в подкоманду LMATRIX процедуры GLM.

Если были заказаны взаимодействия, дизайн-таблица которых содержала пустые клетки, подкоманда SAVELMX не изготовит матрицу и выдаст предупреждение.

Укажите путь/имя .SAV файла в кавычках или апострофах, для сохранения. П/к SAVELMX разрешена только если факторов не больше 4-х и взаимодействия, если заказаны, не выше порядка 3. Структура **L**-матрицы такова:



Ряды – это контрасты и соответствуют контраст-переменным (сумма коэффициентов в каждом ряду в норме 0), а столбцы соответствуют всем факторным уровням и (если заказаны взаимодействия) межфакторным комбинациям уровней. Другими словами, столбцы это подробный перечень всех статей дизайна в “overparameterized” стиле (т.е. без исключения “redundant” градаций). Левую-верхнюю часть занимают по-факторные **L**-матрицы, сшитые блок-диагонально. Если взаимодействия не заказывались, матрица есть только эта часть. Правую часть занимают коэффициенты, появившиеся из-за вовлечения взаимодействий в модель. Заметьте, что эти коэффициенты для контрастов-главных эффектов (“for main effects”) в этой правой части матрицы не нулевые; они представляют собой «расписание» коэффициентов, существующих в левой части, по ячейкам межфакторного дизайна.

ПРИМЕР 8. Индикаторный и простой типы в присутствии взаимодействия.

!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.

* Макрос раскодирует факторы G2 (3 градации) и G3 (3 градации) в тип SIM. Заказаны взаимодействия между факторами. Объединенная L-матрица сохранена. В ней столбцы 1–3 отвечают главному эффекту A (= G2), столбцы 4–6 отвечают главному эффекту B (= G3), столбцы 7–15 отвечают эффекту A\*B ( = G2\*G3).
* Рассмотрим для примера 1-й ряд L-матрицы (этот контраст отвечает контраст-переменной a1 и выражает смысл регрессионного параметра при этой переменной в условиях, если все 8 контраст-переменных выступают регрессорами, т.е. в условиях модели A B A\*B).

Этот контраст есть: 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, то есть:

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

Как видим, 1-й контраст типа Simple сравнивает 1-ю градацию фактора A с 3-й его градацией, и в условиях присутствия взаимодействия каждый из двух ненулевых коэффициентов поровну «расписан» между уровнями второго фактора (B), с которым A взаимодействует. Если в отсутствие взаимодейственного эффекта в модели сравнение было бы разницей : , игнорируя фактор B, то в условиях присутствия взаимодействия эта разница скорректирована: . Разница равномерно «размазана» по градациям B.

* Рассмотрим 5-й ряд L-матрицы (этот контраст отвечает контраст-переменной a1b1 и выражает смысл регрессионного параметра при этой переменной, являющейся статьей эффекта взаимодействия A\*B).

Этот контраст есть: A 0 0 0 B 0 0 0 A\*B 1 0 -1 0 0 0 -1 0 1, то есть:

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

Как видим, 1-й взаимодейственный контраст являет разницу : (; регрессионный коэффициент при a1b1 будет этим сравнением.

* Интерцепт при типе SIM это средняя всех уникальных значений .

!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.

* Этот пуск макроса полностью как предыдущий, но раскодирует факторы в тип IND.
* Рассмотрим для примера 1-й ряд L-матрицы (этот контраст отвечает контраст-переменной a1 и выражает смысл регрессионного параметра при этой переменной в условиях, если все 8 контраст-переменных выступают регрессорами, т.е. в условиях модели A B A\*B).

Этот контраст есть: A 1 0 -1 B 0 0 0 A\*B 0 0 1 0 0 0 0 0 -1, то есть:

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

Как видим, 1-й контраст типа Indicator сравнивает 1-ю градацию фактора A с 3-й его градацией, и в условиях присутствия взаимодействия каждый из двух ненулевых коэффициентов «расписан» между уровнями второго фактора (B) неравномерно, так, что всё записывается на счет B=3 (reference градация в B). То есть сравнение : делается на самом деле только в избранной градации B: . Понятно, что это не то же сравнение, что было выше при типе SIM. Это простой эффект, а не главный эффект. Таким образом, в условиях наличия в модели взаимодействий результаты для главных эффектов различатся у SIM и IND, и именно SIM следует предпочесть, если коэффициент при a1 должен выражать главный эффект.

* Рассмотрим 5-й ряд L-матрицы (этот контраст отвечает контраст-переменной a1b1 и выражает смысл регрессионного параметра при этой переменной, являющейся статьей эффекта взаимодействия A\*B).

Этот контраст есть: A 0 0 0 B 0 0 0 A\*B 1 0 -1 0 0 0 -1 0 1. Заметим, что это то же самое, что было в этом месте у типа SIM. И действительно: результаты (регрессионные параметры = оценки контрастов) у типов SIM и IND совпадут – для контраст-переменных, являющихся статьями взаимодействия. Скажем *правило* в общем виде: результаты SIM и IND всегда одинаковы для эффекта высшего порядка, присутствующего в модели, и только для него. Как частный случай: если в модели нет взаимодействий, то результаты SIM и IND одинаковы для главных эффектов – что и было написано в характеристике этих двух типов в п/к TYPES.

* Интерцепт при типе IND равен в клетке “все факторы = reference-градация”.

ПРИМЕР 9. Тип отклонений в присутствии взаимодействия.

!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.

* Макрос раскодирует факторы G2 (3 градации) и G3 (3 градации) в тип DEV. Заказаны взаимодействия между факторами. Объединенная L-матрица сохранена. В ней столбцы 1–3 отвечают главному эффекту A (= G2), столбцы 4–6 отвечают главному эффекту B (= G3), столбцы 7–15 отвечают эффекту A\*B ( = G2\*G3).
* Рассмотрим для примера 1-й ряд L-матрицы (этот контраст отвечает контраст-переменной a1 и выражает смысл регрессионного параметра при этой переменной в условиях, если все 8 контраст-переменных выступают регрессорами, т.е. в условиях модели A B A\*B).

Этот контраст есть: 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, то есть:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | 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 |

Как видим, 1-й контраст типа Deviation сравнивает1-ю градацию фактора A с невзвешенным усреднением всех трех его градаций: , и в условиях присутствия взаимодействия эти контраст-коэффициенты (2/3, -1/3, -1/3) поровну «расписаны» между уровнями второго фактора (B), с которым A взаимодействует. Разница равномерно «размазана» по градациям B, и эта равномерность является залогом того, что a1 корректно (разумно) выражает главный эффект в присутствии взаимодействия. Отметим, что другие типы контрастов, кроме IND, поступают сходным образом.

* Рассмотрим 5-й ряд L-матрицы (этот контраст отвечает контраст-переменной a1b1 и выражает смысл регрессионного параметра при этой переменной, являющейся статьей эффекта взаимодействия A\*B).

Этот контраст есть: 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, то есть:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | 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 |

Как видим, 1-й взаимодейственный контраст являет разницу : (; регрессионный коэффициент при a1b1 будет этим сравнением.

* Интерцепт при типе DEV это средняя всех уникальных значений .

ПРИМЕР 10. Применение L-матрицы.

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.

* Макрос раскодирует факторы F1, F2, F3, F4 в разные контраст типы. Затребовано создать и взаимодействия до 3-сторонних. L-матрица (главные факторы + взаимодействия) сохранена.
* Выходящий массив сшивается с входящим; дальнейшее делается на сшитом массиве 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.

* Регрессия Y по всем созданным контраст-переменным; т.е. это эквивалент 4-факторного ANOVA с главными эффектами и всеми взаимодействиями до 3-сторонних.

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#.

* Действительно, MANOVA с исходными факторами дает те же результаты, что регрессия на контраст-переменных. Параметры – в таблице “Estimates”. (AUTORECODE понадобилось потому, что MANOVA требует числовых входящих, кодированных последовательными целыми).

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.

* А это те же результаты, получаемые процедурой UNIANOVA (= GLM). Оценки контрастов (совпадающие с регрессионными параметрами выше) – в разделе “Custom Hypothesis Tests...”. Однако UNIANOVA не показывает – среди результатов контрастов – результатов, относящихся к взаимодейственным статьям. Мы видим только результаты, совпадающие с параметрами контраст-переменных a1, b1, b2, c1, c2, c3, d1, но не видим остальных параметров: a1b1, ..., b2c3d1. Чтобы получить их в UNIANOVA, необходимо прямо ввести полную L-матрицу, которую сохранил !KO\_CATCONT. В UNIANOVA такая L-матрица вводится через подкоманду 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.

* Открываем сохраненную L-матрицу. Чтобы матрица была надежно estimable в UNIANOVA, выставляем формат с многими десятичными цифрами. Создаем текстовые константы из букв (имен факторов): главные и взаимодейственные эффекты. Они будут помечать в L-матрице блоки столбцов. Создаем также столбец END = точка-с-запятой (ею будет закрываться строка в LMATRIX).

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

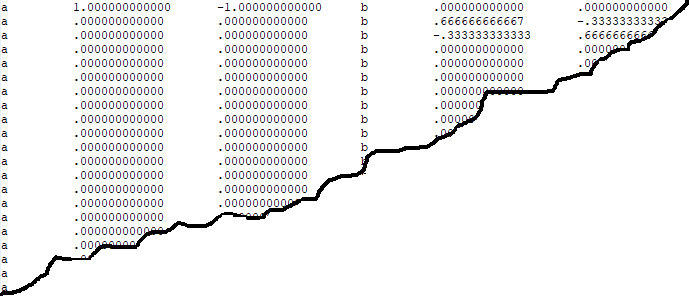
COL119 COL120 COL121 COL122 COL123 COL124 COL125 COL126 COL127 COL128 COL129

COL130 COL131

end

/format= list nocasenum /cells= none.

* Ставим созданные текстовые переменные-разделители блоков столбцов на правильные места. Эти правильные места макрос !KO\_CATCONT подсказывает вам в окне результатов всегда, когда задана п/к SAVELMX. Теперь пускаем SUMMARIZE и получаем в окне результатов готовую для ввода L-матрицу. Открыв таблицу, выделяем все ярлыки рядов и затем правым кликом: Select > Data Cells. Copy. Paste в подкоманду LMATRIX.
* Вместо SUMMARIZE можно было употребить LIST, но это не так удобно.



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.

* Результаты – оценки контрастов в таблице “Contrast Results (K Matrix)” – совпадают с регрессионными параметрами, полученными ранее на контраст-переменных, причем мы видим в окне результатов статьи не только главных эффектов, но и взаимодействий.
* Удобство L-матрицы контрастов в том, что ее ряды (контрасты) можно вводить в подкоманды, подобные LMATRIX, выборочно. Оценки контрастов будут те же, что при вводе полной L-матрицы, покрывающей всю предсказательную модель. Например, в данном примере мы, положим, могли бы ввести в LMATRIX только 2-й и 3-й ряды (они отвечают контрастам b1 и b2). Их результаты оказались бы по-прежнему равны регрессионным параметрам у b1 и b2 в условиях множественной регрессии со всеми переменным от a1 до b2c3d1.

***Особые режимы***

Макрос не рассчитан на расщепленное состояние массива. Взвешенность наблюдений игнорируется. Макрос слушается фильтрации (FILTER, SELECT IF, USE и др.) и временных (под TEMPORARY) операций.

1. Исключением является тип IND, где det(SSCP) не нулевой. По этой причине возможно вводить в качестве предикторов набор из *k* фиктивных (dummy) переменных в том случае, если модель *без* интерцепта. [↑](#footnote-ref-1)
2. «сверх» означает последовательный инкремент. Пусть X есть фактор как мерная переменная. Тогда «линейная» контраст-переменная линейно эквивалентна X. «Квадратная» контраст-переменная линейно эквивалентна остатку от линейного предсказания X2 от X. «Кубическая» контраст-переменная линейно эквивалентна остатку от линейного предсказания X3 от X2 и X, и так далее. [↑](#footnote-ref-2)
3. Тип WPOL = POL при сбалансированном факторе. [↑](#footnote-ref-3)
4. SPSS Statistics может иногда сообщить: “The determinant has a magnitude too small to be represented”. Это не ошибка макроса и в порядке вещей. [↑](#footnote-ref-4)