

# Многоуровневый и многофакторный анализ неметрических данных в СПСС: новые процедуры

## **Новые процедуры в СПСС:**

- Обобщенные модели
- Смешанные модели:

- Линейная

- Смешанные линейные

## **•Новые книги для психологов:**

- Ronald H. Heck, Scott L. Thomas, Lynn N. Tabata.

1. “Multilevel Modeling of Categorical Outcomes Using IBM SPSS”. 2012.
2. “Multilevel and Longitudinal Modeling with IBM SPSS”. 2010.

# Новые процедуры – новые ВОЗМОЖНОСТИ

- Дефицитарность процедур ДА в рамках общей линейной модели.
- Анализ многоуровневых данных.
- Анализ номинальных (бинарных), мультиноминальных и порядковых данных.

# Новые подходы и термины

**MLM** - многоуровневое моделирование: подход к статистической обработке данных, имеющих вложенную или иерархическую структуру. Т.е. это построение моделей, в которых наблюдения представлены **в разных контекстах**. Этот подход описывается разными терминами: случайные коэффициенты, смешанные эффекты, иерархические линейные и многоуровневые регрессионные модели.

# Новые подходы и термины

В СПСС «Смешанными моделями» называются модели, включающие фиксированные и случайные эффекты. Фактически, там, где есть повторные измерения – скоррелированные данные, и необходимость оценки отдельных групповых вариаций, там и есть смешанные модели. Процедура **MIXED** обеспечивает обработку 2-х и более уровневых данных.

**GENLIN** – процедура обобщенного линейного моделирования.

**GENLIN MIXED** – в СПСС это обобщенные модели с категориальными данными (начиная с 19 версии, 2010 г.).

# НОВЫЕ ПОДХОДЫ И ТЕРМИНЫ

The generalized linear model expands the general linear model so that the dependent variable is linearly related to the factors and covariates via a **specified link function**. The GenLin-model allows for the dependent variable to have a non-normal distribution. It covers widely used statistical models, such as:

1. linear regression for normally distributed responses;
2. logistic models for binary data;
3. loglinear models for count data,
4. log-log models for interval data.

**Функция связи** – это способ преобразования ЗП, позволяющей ее оценить с помощью линейной модели. Например:

**Identity.**  $f(x)=x$ .

**Complementary log-log.**  $f(x)=\log(-\log(1-x))$ . *This is appropriate only with the binomial distribution.*

**Cumulative Cauchit.**  $f(x) = \tan(\pi (x - 0.5))$ , This is appropriate only with the multinomial distribution.

**Log.**  $f(x)=\log(x)$ . *This link can be used with any distribution.*

**Negative binomial.**  $f(x)=\log(x / (x+k-1))$ , where  $k$  is the ancillary parameter of the negative binomial distribution. This is appropriate only with the negative binomial distribution.

**Negative log-log.**  $f(x)=-\log(-\log(x))$ . *This is appropriate only with the binomial distribution.*

**Odds power.**  $f(x)=[(x/(1-x))^\alpha-1]/\alpha$ , if  $\alpha \neq 0$ .  $f(x)=\log(x)$ , if  $\alpha=0$ .  $\alpha$  is the required number specification and must be a real number. This is appropriate only with the binomial distribution.

**Probit.**  $f(x)=\Phi^{-1}(x)$ , where  $\Phi^{-1}$  is the inverse standard normal cumulative distribution function. This is appropriate only with the binomial distribution.

# Обработка одноуровневых моделей

Процедура **GENLIN** – **Обобщенные  
линейные модели.**

Предполагается, что испытуемые случайно выбраны из одной совокупности, переменные несвязанные и нет их какой-либо группировки более высокого уровня.

Установки и опции процедуры **GENLIN**:

1. Тип модели: указать модель или **САМОСТОЯТЕЛЬНО** выбрать тип распределения и соответствующую функцию связи.
2. Указать ЗП и ее опорную категорию –

# Обработка одноуровневых моделей

3. Предикторы и Модель: 1) включение предикторов и эффектов их взаимодействия; 2) **смещение** - важен при Пуассоновском распределении, отражает вес или значимость ЗП (по умолчанию=1).
4. Оценивание: 1) метод оценки параметров модели – **гибридный** (в нем сначала реализуется метод Фишера, а затем другой – в зависимости от сходимости и достижения максимума итераций); 2) расчет ковариационной матрицы – на основе модели или **робастная** процедура; 3)



# Результаты (файл: ch3data.sav)

- **Статистики согласия<sup>a</sup>**
  - Значение ст.св.    Значение/ст.св.
  - Уклонение    5,191    1    5,191
  - Масштабированное уклонение    5,191    1
  - Хи-квадрат Пирсона    5,175    1    5,175
  - Масштабированное значение хи-квадрат Пирсона    5,175    1
  - Log-правдоподобие<sup>b</sup>    -3833,346
  - Информационный критерий Акаике (AIC)    7672,693
  - Скорректированный информационный критерий Акаике (AICc) для выборки конечного объема    7672,696
  - Информационный критерий Байеса (BIC)    7693,044
  - Состоятельный информационный критерий Акаике (CAIC)    7696,044
- Информационные критерии представлены в виде "меньший-лучший". При расчете информационных критериев используется функции логарифма отношения правдоподобия

# Результаты

## 1. Общий критерий<sup>a</sup>

Chi-квadrat отношения правдоподобия ст.св. Знч.  
408,781 2 ,000

Зависимая переменная: Умение читать

Модель: (Константа), female, minor

□ Сравнивает подогнанную модель с моделью, содержащей только константу.

## 2. Проверка эффектов модели

Источник Тип III

	Chi-квadrat	Вальда	ст.св.	Знч.
(Константа)	930,494	1	,000	
female	56,921	1	,000	
minor	337,743	1	,000	

Зависимая переменная: Умение читать

Модель: (Константа), female, minor

# Результаты

## Оценки параметров

Параметр В Стд. Ошибка Проверка гипотезы Exp(B)  
Exp(B)

(Константа) 1 <sup>b</sup>	,0466	,000	1,738
[female=0]	<b>-,418</b>	<b>,0555</b>	<b>,658</b>
[minor=0]	<b>1,034</b>	<b>,0562</b>	<b>2,811</b>

Мальчики читают лучше, чем девочки в 0,658  
раза

Дети из обеспеченных семей читают лучше в 2,811  
раза

# Обобщенные уравнения оценки

Это расширение Обобщенной линейной модели на данные, включающие повторные измерения на одних и тех же испытуемых. ОУОц – описывают и оценивают ковариацию повторных наблюдений.



**Можно включать в модель межгрупповые и внутригрупповые факторы.**

# Установки и опции

Вкладка **Повтор**: указываем: а) переменную, соответствующую испытуемым (subjects variables, групповые переменные), т.е. ту переменную, в которую **включены** внутригрупповые факторы – фактически это все повторные данные; б) переменную, указывающую, какому временному периоду (повторному измерению) соответствует каждое измерение – 1,2, 3 ...

# Установки и опции

**Тип рабочей корреляционной матрицы**, описывающей структуры повторных измерений ЗП. Заранее сложно предугадать оптимальную структуру корреляций, поэтому опытным путем выбираем оптимальную.

**Независимая** допускает, что повторные измерения некоррелированы, что не соответствует большинству случаев.

**Взаимозаменяемая** (или составная симметрия – однородные дисперсии и ковариации) допускает однородные корреляции между ее элементами, т.е. допускаются одинаковые корреляции.

**Авторегрессионная** или AP-1 матрица предполагает, что повторные измерения имеют авторегрессионную структуру первого порядка.

**M-зависимая матрица** предполагает, что последовательные измерения имеют общий к-т корреляции, пары измерений, разделенные третьим измерением, с которым имеют общий к-т корреляции.

**Неструктурированная** корр. матрица предполагает, что корреляции между разными элементами различны.

- Как и для смешанных многоуровневых моделей выбор структуры матрицы не сильно влияет на оценки модели.
- Если процедура сообщает о том, что модель не сходится, то следует изменить структуру корр. матрицы.

# Пример: ch5data1.sav

Нужно оценить, как изменялась в течение года:

- 1) Динамика успешности чтения в целом (время - ковариата),
- 2) По четвертям (время – внутригрупповой фактор),
- 3) С учетом пола ученика (пол – межгрупповой фактор),
- 4) С учетом межфакторного взаимодействия  $\text{Время} * \text{Пол}$  (смешанный 2-х факторный дизайн).

# Результаты

## Оценки параметров

Параметр	B	Стд. Ошибка	Проверка гипотезы	Exp(B)	Верхняя
(Константа)	1	,0647	,000	1,871	
time	-,068	,0229	<b>,003</b>	<b>,934</b>	
[female=1]	,458	,0968	<b>,000</b>	<b>1,580</b>	
[female=1] * time	,028	,0335	,405	1,028	

Зависимая переменная: Уровень чтения

Модель: (Константа), time, female, female \* time