

## АНАЛИЗ ДАННЫХ В ПРОГРАММЕ MINISTER

**Галина Смирнова,  
Алексей Смирнов**

Славянский-на-Кубани государственный педагогический институт  
smirnova\_g\_i@mail.ru

**Рассматривается первоначальный анализ данных, проведённый программой Minister. Приводятся примеры оформления контрольных файлов, расшифровываются значения таблицы сходимости данных, описывается процесс вычисления значений и сопутствующие данному процессу термины (критерий сходимости, сходимость анализа, анализ закреплённых данных, проверка связанности и т.д.).**

*Ключевые слова:* программа Minister, контрольный файл, таблица сходимости данных, критерий сходимости, сходимость анализа, анализ закреплённых данных, проверка связанности.

В данной работе дано описание процесса анализа результатов тестирования посредством программы Minister. Эта программа позволяет провести исследование данных по многим параметрам, о которых речь и пойдёт как в текущей статье, так и в последующих. Стоит отметить, что мы работаем с обучающей версией программы, которая носит название Minister; версия ограничена в размере файла данных. Minister позволяет обрабатывать результаты тестирования не более 75 испытуемых, ответивших не более, чем на 25 тестовых заданий. В то время как Winsteps не имеет ограничений.

После того, как вы создали свой контрольный файл и сохранили под своим именем, запустите программу. Кратко напомним, как это быстро сделать<sup>1</sup>:

1) дважды щёлкните по ярлыку программы Ministeps, загрузится рабочий лист программы;

2) программа спросит вас «Имя Вашего файла данных? (например, КСТ.txt) Нажмите «Enter» («Control file name? (e.g., КСТ.txt). Press Enter for Dialog Box:»). Нажмите клавишу «Enter»;

3) появится окно, в котором нужно выбрать файл данных, который вы создали и сохранили (с примерами организации файлов

Теория

Методика

1

Подробно см.: Linacre J. M. WINSTEPS Rasch measurement computer program. Chicago: Winsteps.com, 2005, а также: Г.И. Смирнова, А.В. Смирнов. Начало работы с программой Winsteps//Педагогические измерения. 2006. № 3. С. 106–114.

ПЕД	
	измерения

данных можно познакомиться, обратившись к пункту меню «Помощь» («Help») опции «Содержание» («Contents») и в появившемся перечне выбрать «Примеры контрольных файлов и файлов данных» («Examples of control and data files») (рис. 1));

4) программа просит вас: «Задайте имя для файла вывода данных (или нажмите "Enter" для создания временного файла)» (Report output file name (or press Enter for temporary file). Нажмите «Enter»;

5) снова нужно нажать «Enter» при появлении сообщения «Дополнительная спецификация (или нажмите "Enter")» («Extra specifications (or press Enter)»).

На данном этапе начинается анализ ваших данных (рис. 2). Если появится сообщение об ошибке, то выберите в пункте меню «Правка» («Edit») опцию «Исправить контрольный файл» («Edit Control File»), внесите исправления, на которые вам укажет про-

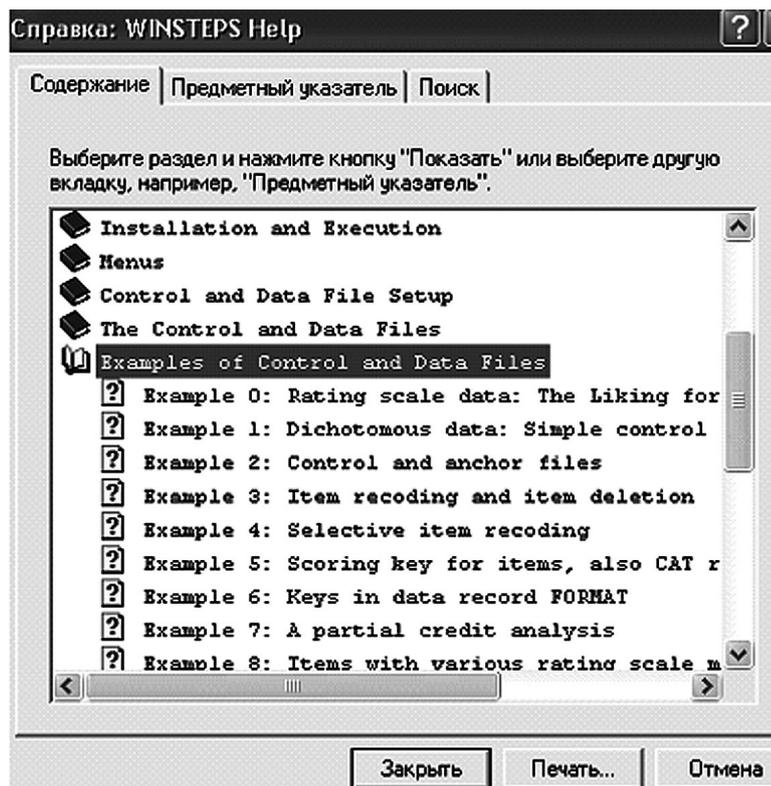


Рис. 1. Примеры организации контрольного файла и файла данных

```

MINISTEP - [C:\WINSTEP\examples\example0.txt]
File Edit Diagnosis Output Tables Output Files Batch Help Specification Plots SAS/SPSS Graphs Data Setup
75 KID Records Input.
CONVERGENCE TABLE
+Control: \examples\example0.txt Output: \examples\ZOU608WS.TXT
| PROX ACTIVE COUNT EXTREME 5 RANGE MAX LOGIT CHANGE |
| ITERATION KIDS ACTS CATS KIDS ACTS MEASURES STRUCTURE |
|-----|-----|-----|-----|-----|-----|-----|
| 1 75 25 3 3.78 3.20 3.8918 .0740 |
| 2 74 25 3 4.59 3.71 .8258 -.6158 |
| 3 74 25 3 4.83 3.92 .2511 -.1074 |
Checking connectivity
+Control: \examples\example0.txt Output: \examples\ZOU608WS.TXT
| JMLE MAX SCORE MAX LOGIT LEAST CONVERGED CATEGORY STRUCTURE |
| ITERATION RESIDUAL* CHANGE KID ACT CAT RESIDUAL CHANGE |
|-----|-----|-----|-----|-----|-----|-----|
| 1 2.84 -.1955 60 22* 2 21.44 .0076 |
| 2 .71 -.0335 53 15* 0 -5.89 -.0144 |
| 3 -.43 .0293 53 5* 1 3.48 .0101 |
| 4 .32 .0235 18 11* 1 2.71 .0079 |
| 5 .24 .0184 18 11* 0 -2.09 .0060 |
| 6 .19 .0141 18 11* 0 -1.63 .0045 |
| 7 .14 .0108 18 11* 0 -1.25 .0035 |
| 8 .11 .0082 18 11* 0 -.96 .0026 |
| 9 .08 .0062 18 11* 0 -.73 .0020 |
| 10 .06 .0048 18 11* 0 -.56 .0015 |
Calculating Fit Statistics
Standardized Residuals N(0,1) Mean: .06 S.D.: 1.04
Output written to C:\WINSTEPS\examples\ZOU608WS.TXT
CODES= 012
Measures constructed: use "Diagnosis" and "Output Tables" menus
    
```

**Теория**

Рис. 2. Начало анализа в программе Ministeps. Таблицы сходимости данных

грамма, сохраните их и выберите в пункте меню «Файл» («File») опцию «Выход, затем запуск Winsteps» («Exit, then Restart Winsteps»).

После описания данных следуют таблицы сходимости данных (Convergence table) (рис. 2)<sup>2</sup>. Визуально таблица разбита на две части. Колонки

верхней части таблицы имеют следующие значения.

PROX ITERATION — алгоритм нормального приближения для первоначального вычисления затраченного времени для подсчёта оценок. Кратко суть алгоритма можно описать так: PROX — это вычислительный процесс, достаточно про-

**2**  
 В данной статье мы будем опираться на файл данных example0.txt, который приведён в папке examples каталога Winsteps корневого каталога C.

ПЕД	
	измерения

стой для ручного вычисления<sup>3</sup>. Всё, что необходимо PROX для оценивания испытуемых и тестовых заданий, — это колонка результатов испытуемых и строка результатов для заданий. Предполагается, что уровень знаний в выборке и трудность задания приближённо распределены по нормальному закону.

Продолжаем описание таблицы.

ACTIVE COUNT — число «экстремальных» элементов, которые были исключены из дальнейшего анализа данных из-за их абсолютного/нулевого счёта.

PERSONS (может быть использовано KIDS или PUPLES в зависимости от того, как вы описали свои данные в контрольном файле) — значение ACTIVE COUNT для испытуемых. Из рисунка 2 видно, что на первом шаге алгоритма PROX число испытуемых было 75, на втором — 74, следовательно, один испытуемый был исключён из дальнейшего анализа.

ITEMS (может быть использовано ACTS, в зависимости от того, как вы описали свои данные в контрольном файле) — значение ACTIVE COUNT для тестовых заданий. 25 тестовых заданий.

CATS — количество категорий в оценочной шкале<sup>4</sup>. Шкала имеет три категории оценивания 0-1-2 (dislike — neutral — like).

EXTREME 5 RANGE — размах данных между средними арифметическими пяти лучших и пяти худших испытуемых.

PERSONS — размах между средним значением подготовленности пяти испытуемых с высоким уровнем подготовки и пяти испытуемых с низким уровнем подготовки.

ITEMS — размах между средним значением трудности пяти самых трудных заданий и пяти самых лёгких заданий.

MAX LOGIT CHANGE — изменение значения логитов трудности заданий в результате итераций.

MEASURES (единица измерения) — максимальное значение логита для испытуемых. Это значение обычно постепенно уменьшается до определённого уровня сходимости, но не меньше значения критерия LCONV=. Это значение может включать в себя не более двух знаков после запятой, поэтому изменение в 0,005 логитов может не отразиться на итоговом значении данного критерия. Установлено, что данное значение в логитах для испытуемых должно уменьшаться с каждым шагом.

Для более понятной интерпретации данного значения обратимся к термину «convergence considerations» (сходимость анализа) — это вопросы определения качества результа-

3

Wright B.D.,  
Masters G.N.  
Rating Scale Analysis.  
Chicago: MESA Press,  
1982. 206 p

4

Для дихотомической  
шкалы это значение  
равно 2.



тов уточняющего шкалирования параметров испытуемых и заданий методом максимального правдоподобия. Чтобы завершить процесс вычисления критерия совместимости, используйте клавиши Ctrl+F. Если вы сомневаетесь в принятии подобного решения, установите более узкое значение критерия сходимости для своего итогового отчёта, например:

CONVERGE=B ; both  
LCONV= and RCONV= apply

LCONV= .0001 ; largest logit  
change .0001 logits

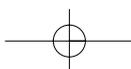
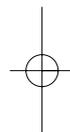
RCONV= .01 ; largest score  
residual .01 score points

MJMLE= 0 ; unlimited  
JMLE iterations

И будьте готовы остановить итерации вручную, если это понадобится, используя опцию «Finish Iterating» или Ctrl+F в пункте меню File. Это связано с тем, что модель Раша является нелинейной. Следовательно, измерения не могут быть сделаны незамедлительно и очень точно, как это предполагает решение линейного уравнения. И наоборот, оценка может быть результатом угадывания правильного ответа. Работа программы Ministeps строится на исходном предположении, что все испытуемые имеют равный уровень подготовленности и все тестовые задания имеют одинаковый уровень сложности. Ожидаемые ответы, построенные на данном предположении, срав-

ниваются с реальными данными. Некоторые испытуемые имеют более высокий уровень подготовленности, чем ожидалось, некоторые меньший. Некоторые тестовые задания оказались более сложными, некоторые более лёгкими. Новое предположение, то есть оценка, понимается как реальный уровень подготовленности испытуемого и уровень сложности задания, а также как категории оценивания оценочной шкалы, как наиболее значимые для процесса измерения по модели Раша.

Данные снова сравниваются. Это и есть «цикл», «шаг» («iteration») проверки. Ожидаемые результаты сравниваются с реальными и вычисляется исправленная оценка. Этот процесс продолжается до тех пор, пока оценка не становится меньше, чем указано в значении критерия LCONV=, или пока наибольшая разница между общим баллом и ожидаемым его значением не станет меньше значения, указанного в критерии RCONV=. Чёткие правила остановки цикла проверки обозначены в критерии CONVERGE=. Когда оценки становятся «достаточно хорошими», считается, что итерационный процесс «сошёлся» (отсюда пошло название — критерий сходимости). Итерации прекращаются. Вычисляются статистические значения и начинается формирование отчётов.



ПЕД	
	измерения

Существуют стандартные критерии сходимости, подходящие как для небольших, так и для средних объёмов данных.  $LCONV=$  труднее сопоставить с малыми объёмами данных и с данными, где оценки имеют большой разброс,  $RCONV=$  — с большими объёмами данных.

«Наименьшими элементами сходимости» (испытываемые, тестовые задания) являются те из них, которые Ministeps трудно оценить. Это происходит не потому, что их ответы не соответствуют модели, а потому что они являются экстремальными (в случае, если в векторах-строках испытываемых все нули или единицы). Согласитесь, наиболее сложно вычислить уровень подготовленности того испытываемого, который ответил правильно только на 1 из 20 заданий, нежели того, кто ответил на 10 из 20 заданий. Это потому, что модель Раша сравнивает реальные данные с моделью.

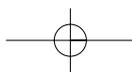
Далее, после произведённого сравнения, данные оцениваются для установления соответствия модели. Этот процесс аналогичен линейной регрессии. Параметры оцениваются на соответствие линейной регрессионной модели данных. Затем определяется место, которые занимают данные. Это является последним этапом определения экстремального значения экспериментальных величин.

Для дальнейшего понимания процесса проводимого программой Ministeps анализа поясним значение термина «Anchored analysis» (анализ закреплённых данных). Закреплённое значение (anchor) всегда различается с текущими данными то тех пор, пока не будет с ними соотнесено. Поэтому остатки никогда не будут равны 0. Сходимость осуществляется до того, как значение логита становится ничтожно мало. Таким образом,  $RCONV=$  становится бесполезным и применяется  $LCONV=$ . Предлагается примерное значение параметров:

$CONVERGE = L$  ; only  $LCONV$  is operative

$LCONV = .005$  ; smaller than visible in standard, two decimal, output.

Некоторые шаблоны представления данных требуют крайне плотного расположения чисел друг от друга. И эти критерии приобретают важное значение для тех данных, где есть пропуски. Например, для вертикально расположенного шаблона для 5 тестовых заданий сходимость достигается при 85 итерациях. Но если затянуть процесс, то есть использовать меньшие значения критериев  $LCONV=$  и  $RCONV=$ , то сходимости можно достигнуть после 400 итераций. Дальнейшие итерации не имеют смысла, так как математическая точность для этих данных уже была достигнута.



та. На рис. 3 показано сравнение уровня сложности тестовых заданий для 5 тестовых заданий, оцененных с помощью стандартного и строго заданного критерия сходимости.

CONVERGE = B

LCONV = .001 ; 10 time stricter than usual

RCONV = .01 ; 10 times stricter than usual

Помните, что сходимость может потребовать большого количества итераций и может потребоваться ваше вмешательство для завершения процесса или Ctrl+F.

Пример:

CONVERGE = B

LCONV = .001 ; 10 time stricter than usual

RCONV = .01 ; 10 times stricter than usual

Продолжаем описание таблицы сходимости.

STRUCTURE — максимальное значение логита, которое изменяется при каждой оценке измерений, не может быть столь мало как MEASURES.

Вторая часть таблицы сходимости отделена строкой «Проверка связности» («Checking connectivity»). Однако данные могут иметь несколько вариантов связи, как это показано в табл. 1. Ministeps оценивает каждого испытуемого и его уровень подготовленности, каждое тестовое задание и его уровень сложности внутри

**Теория**

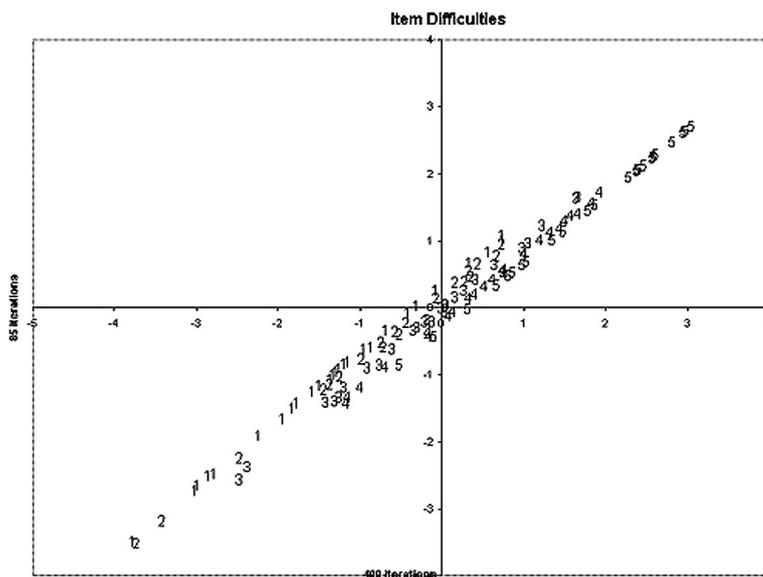
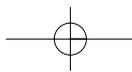


Рис. 3. Сравнение критериев сходимости



**ПЕД**  
**измерения**

одной взаимосвязанной совокупности данных. Обычно так должно происходить, однако существуют исключения. Данные могут не удовлетворять определённым условиям («well-conditioned»)⁵.

Наличие экстремальных значений (нулевых и абсолютных) в измерениях означает, что наши данные вышли за рамки стандартных связей. Ministeps пользуется байесовской логикой (Bayesian logic), чтобы проверить измерения на соответствие стандартным связям. Наиболее затруднительные ситуации описаны в файле Examsubs.txt папки «examples» каталога «Winsteps»:

Title = «Example of subset reporting»

```
Name1 = 1
Item1 = 9
NI = 10
&End
Item 1 dropped as extreme perfect score
Item 2 in subset 2
Item 3 in subset 2
Item 4 in subset 3
Item 5 in subset 3
Item 6 has a Guttman pattern:
item subset 1
Item 7 in subset 5
Item 8 in subset 5
Item 9 in subset 6
Item 10 in subset 6
END LABELS
Alf 100000 ; drops with extreme score
Ben 101001 ; subset 2
Carl 110001 ; subset 2
David 111011 ; subset 3
```

5

Fischer G.H.,  
Molenaar I.W. (eds.)  
Rasch models: foundations, recent developments, and applications.  
New York: Springer-Verlag. 1995. p. 41–43.

Таблица 1

CONVERGENCE TABLE

ITERATION	PROX			ACTIVE COUNT		EXTREME 5 RANGE		MAX LOGIT CHANGE	
	PUPILS	ACTS	CATS	PUPILS	ACTS	PUPILS	ACTS	MEASURES	STRUCTURE
1	76	25	3	3.59	1.62	3.1355	-1.229		
2	74	12	3	4.03	1.90	.3862	-.5328		
3	74	12	3	4.19	1.96	.1355	-.0783		

WARNING: DATA MAY BE AMBIGUOUSLY CONNECTED INTO 6 SUBSETS, see Connection Ambiguities

ITERATION	JMLE	MAX SCORE	MAX LOGIT	LEAST CONVERGED			CATEGORY STRUCTURE	
	RESIDUAL*	CHANGE	CHANGE	PUPIL	ACT	CAT	RESIDUAL	CHANGE
1	-2.04	.2562		7	5*	2	-.72	.0003

Standardized Residuals N(0,1) Mean: .03 S.D.: 1.24

Указание на наличие нестандартных связей, т.е. экстремальных элементов в файле данных

Edward 111101 ; subset 3  
 Frank 011 ; subset 4  
 George 001 ; subset 5  
 Henry 010 ; subset 5  
 Ivan 01 ; subset 6  
 Jack 10 ; subset 6

Тест состоит из 10 тестовых заданий. На первое тестовое задание все испытуемые дали правильный ответ, поэтому оно было расценено как экстремальное и не учитывалось при дальнейшем анализе данных. Первый испытуемый Alf дал неверные ответы на все неэкстремальные задания и так же был исключён из дальнейшего анализа.

Таким образом, удалено задание № 1 и испытуемый Alf.

*Подмножество № 1:* Тестовое задание № 6 подходит под модель Гутмана (Guttman pattern), которая проводит различия между теми, кто правильно ответил на данное тестовое задание, и теми, кто ответил неправильно, без соотнесения с различиями в данных. Таким образом, получается дистанция, равная неизвестной величине логита между теми, кто ответил правильно на тестовое задание № 6, и теми, кто ответил неправильно. Следовательно, уровень сложности тестового задания № 6 не определён.

Оставшиеся подмножества включают в себя измерения, которые могут быть измерены непосредственно в самом подмножестве, однако будут обладать

дистанцией неизвестной величины от испытуемых до тестовых заданий в других подмножествах.

*Подмножество 2:* Тестовые задания № 2, 3 и испытуемые Ben, Carl.

*Подмножество 3:* Тестовые задания № 4, 5 и испытуемые David, Edward.

*Подмножество 4:* Испытуемый Frank.

*Подмножество 5:* Тестовые задания № 7, 8 и испытуемые George, Henry.

*Подмножество 6:* Тестовые задания № 9, 10 и испытуемый Jack.

При данных обстоятельствах Ministeps выводит для отчёта один из множества вариантов возможных решений. Значение статистических данных и стандартной ошибки обычно верно. Коэффициент надёжности случаен. Сравнения при измерениях внутри подмножества являются верными. Перекрёстные сравнения во множестве являются случайными. Выбор метода решения, вероятно, зависит от закреплённых данных (anchor), равнозначных по уровню сложности тестовым заданиям (или двум равнозначным по уровню подготовленности испытуемым). В различных подмножествах для получения равных значений или подгонки закреплённого значения (anchor) дела-

ПЕД	
	измерения

ются равными значениями каждого подмножества (или любого из подмножеств). Или же необходимо проводить раздельный анализ данных. Или

нужно составить реальную или же учебную матрицу ответов, состоящую из 0 и 1.

В отчёте эти данные размещаются посередине экрана:

WARNING: DATA MAY BE AMBIGUOUSLY CONNECTED INTO 6 SUBSETS

SUBSET 1 OF 1 ITEMS includes ITEM 6: Item 6

SUBSET 2 OF 2 PERSONS includes ITEM 2 and PERSON 2: Ben

SUBSET 3 OF 2 PERSONS includes ITEM 4 and PERSON 4: David

SUBSET 4 OF 1 PERSONS includes PERSON 6: Frank

SUBSET 5 OF 2 PERSONS includes ITEM 7 and PERSON 7: George

SUBSET 6 OF 2 PERSONS includes ITEM 9 and PERSON 9: Ivan

В отчёте программа Ministeps формирует список номеров и имен испытуемых для каждого подмножества, благодаря чему вы можете сопоставить их ответы.

Обратимся к следующему элементу таблицы.

JMLE (joint maximum likelihood estimation) — данный алгоритм используется для более точного измерения. Параметры оценки для тестовых заданий прописаны в IAFILE=, для испытуемых — в PAFILE= и для структуры ответа — в SAFILE=.

Тестовые задания, удовлетворяющие критериям алгоритма JMLE, могут иметь как закреплённые значения (anchored), так и незакреплённые (unanchored), которые имеют абсолютно одинаковую структуру. Единственное отличие состоит в том, что закреплённые значения (anchored) не изменяются после каждой ите-

рации, в то время как незакреплённые (unanchored) изменяются. Алгоритм JMLE сходится, когда «подсчитанные баллы измерений = ожидаемым баллам, основанным на данных оценках». Для закреплённого значения (anchored) этот критерий сходимости никогда не встречается, однако значение статистических данных подсчитывается и включается в отчёт программой Ministeps. Сходимость основана на незакреплённых (unanchored) оценках.

Для того, чтобы зафиксировать измерение тестовых заданий (или испытуемых) в порядке, полученном в текущем анализе, нужно использовать предварительно закреплённое значение (anchor). Данный порядок можно применять при любом другом анализе, пользуясь формой «общее тестовое задание» («common item») или «общий испытуемый» («common respondent»).

son»). В отличие от «общих» методов, в которых набор данных способствует определению уровня сложности тестового задания, набор данных с закреплённым значением (anchored) не влияет на данные значения. Обычно применение тестовых заданий с закреплённым значением (anchored) (или испытуемых) не требует вычисления каких-либо формул или использования констант. В процессе закреплённого анализа уровень подготовленности испытуемого вычисляется на основе закреплённого значения (anchored) уровня сложности тестового задания. Полученные значения уровня подготовленности испытуемого используются для вычисления уровня сложности всех незакреплённых (non-anchored) тестовых заданий. Далее все незакреплённые (non-anchored) тестовые задания и испытуемые пересчитываются до тех пор, пока не будет достигнут наилучший суммарный тестовый балл. Разница между закреплённой величиной и величиной, полученной при оценке текущих данных, может быть записана в отчёте как смещение (displacements). Стандартные ошибки, ассоциирующиеся со смещением, могут использоваться для вычисления приблизительной t-статистики и проверки гипотезы, основанной на предположении, что смещение обусловлено только ошибкой вычисления.

ITERATION (шаг, цикл) — время, через которое ваши данные пересчитываются. Обычно достаточно 100 шагов, редко когда эта цифра превышает.

MAX SCORE RESIDUAL — максимальное значение остатка (разницы между общим баллом и ожидаемым) для любого испытуемого или тестового задания — используется для сравнения со значением RCONV=. Это число уменьшается до тех пор, пока это допустимо сходимостью. Знак «\*» указывает, кто из испытуемых или какое из тестовых заданий имеют остаток.

MAX LOGIT CHANGE — максимальное значение сдвига логита для любого испытуемого или тестового задания — используется для сравнения со значением LCONV=. Это число уменьшается до тех пор, пока это допустимо сходимостью.

LEAST CONVERGED (наименьшая сходимость) — входит в состав цифр, которые встречаются в отчёте для испытуемых, тестовых заданий и категорий при дальнейшем использовании критерия сходимости. Знак «\*» указывает, какое из двух значений испытуемого или тестового задания сходится в наибольшей степени. Столбец CAT (category — категория) может быть не связан со столбцом ITEM, находящимся слева. Подробнее можно узнать, рассмотрев табл. 3.2, описывающую эти особенности. Чтобы

Теория

12/0000

ПЕД  
измерения

найти описание табл. 3.2., воспользуйтесь пунктом меню «Помощь» («Help») опцией «Содержание» («Contents»), в появившемся перечне выберите «Таблицы вывода» («Output tables») и далее найдите номер нужной вам таблицы (рис. 4).

Следующий столбец таблицы имеет название CATEGORY RESIDUAL (остаток категории), что означает максимальное значение остатка (разница между общим балом, полученным при измерении, и десятичным ожидаемым значением)

для любой структуры категории. Это значение мало, однако показатель меньше, чем 0,5, не важен для исследования.

STRUCTURE CHANGE — максимальное значение логита для любой градации. Не используется для определения сходимости, приводится для вашей информации. Данное значение крайне мало.

После таблицы приводится строка «Standardized Residuals», в которой данные проверяются на соответствие нормальному распределению.

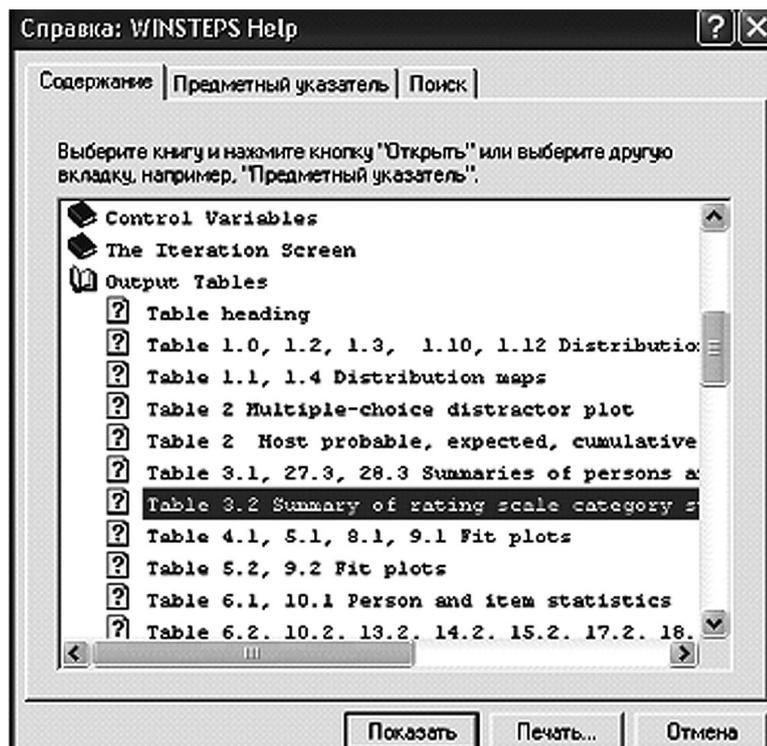
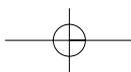


Рис. 4. Поиск таблицы 3.2



Значительное отступление от значения 0,0 и стандартное отклонение со значение 1,0 показывают, что ваши данные не удовлетворяют условиям модели Раша, по которым данные должны иметь нормальное распределение.

После формирования таблицы сходимости программа Ministeps может описать процесс создания отчётов и таб-

лиц вывода, графиков и рисунков. Количество записей зависит от объёма контрольного файла. В конце обработки ваших данных появится строка «Measures constructed — use «Diagnosis» and «Output Tables menu», согласно которой для визуализации отчёта нужно воспользоваться указанными пунктами меню программы.

**Теория**