

VII. Отбор данных

- [Выбор наблюдений](#)
 - [Классификация операторов](#)
 - [Операторы отношения](#)
 - [Логические операторы](#)
 - [Булева алгебра](#)
 - [Функции](#)
 - [Ввод условного выражения](#)
 - [Примеры отбора данных](#)
- [Извлечение случайной выборки](#)
- [Сортировка наблюдений](#)
- [Разделение наблюдений на группы](#)

В этой главе мы на примере файлов wahl.sav и studium.sav покажем разнообразные возможности, предоставляемые в SPSS для отбора данных. Отбор данных — это выбор наблюдений по определенным критериям; так, например, при опросе избирателей (файл wahl.sav) можно отобрать только мужчин, голосующих за ХДС/ХСС, а при опросе студентов (файл studium.sav) — только студенток, изучающих психологию и медицину. После этого все вычисления будут проводиться только с этими отобранными наблюдениями.

Для этого в SPSS существует три принципиальные возможности:

- Выбор наблюдений по определенному условию (логическому выражению),
- Извлечение случайной выборки наблюдений из файла данных,
- Разделение наблюдений на группы в соответствии со значениями одной или нескольких переменных.

Данная глава разбита на три раздела, посвященные каждой из этих возможностей. Еще в одном разделе рассматривается вопрос сортировки наблюдений, содержащихся в файле данных, по значениям выбранных переменных.

7.1. Выбор наблюдений

Проведем частотный анализ переменной *partei* (партия). При этом мы будем учитывать только респондентов-женщин. Поступите следующим образом:

- Загрузите файл wahl.sav в редактор данных.
- Выберите в меню команды Data (Данные) Select Cases... (Выбрать наблюдения)

Откроется диалоговое окно Select Cases (см. рис. 7.1). По умолчанию в этом диалоге выбран пункт All cases (Все наблюдения).

- Выберите пункт If condition is satisfied (Если выполняется условие) и щелкните на кнопке If... (Если). Откроется диалоговое окно Select Cases: If (см. рис. 7.2).

Это диалоговое окно разделено на следующие части:

- Список исходных переменных: Содержит переменные, содержащиеся в открытом файле данных. В нашем случае это переменные *fragebnr*, *sex*, *alter* и *partei*.
- Редактор условий: Здесь записывается логическое выражение, по которому должны быть отобраны наблюдения. В данный момент редактор условий пока пуст.

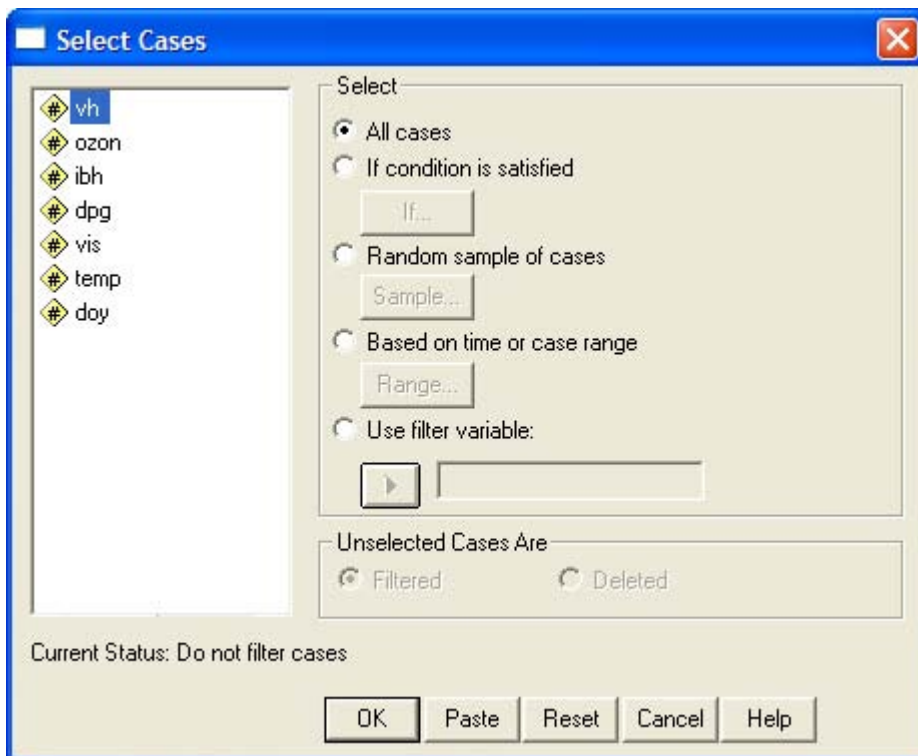


Рис. 7.1: Диалоговое окно Select Cases

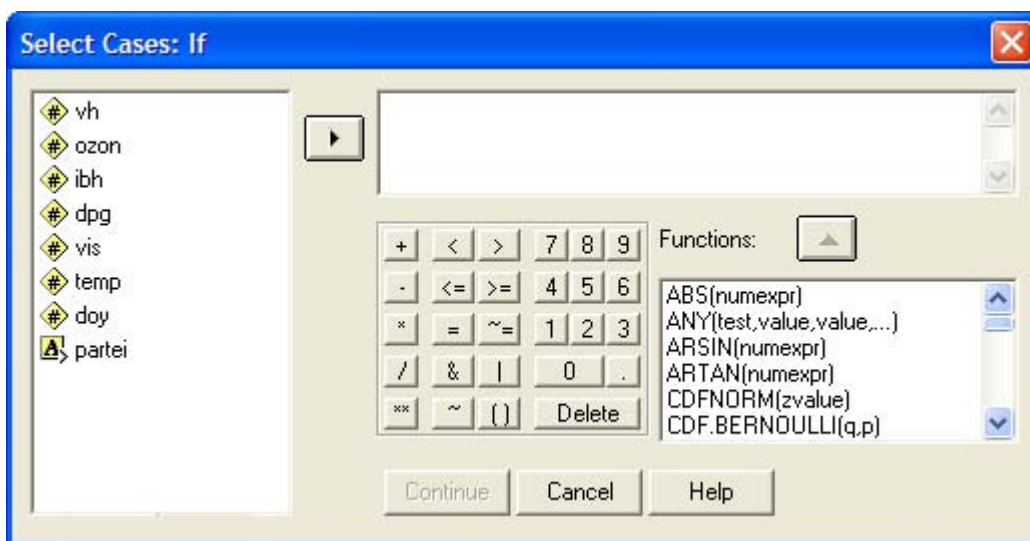


Рис. 7.2: Диалоговое окно Select Cases: If

- Кнопка с треугольником: Эта кнопка позволяет перенести переменную из списка исходных переменных в редактор условий.
- Клавиатура: Содержит цифры, а также арифметические, логические операторы и операторы отношения; с ней можно работать как с обыкновенным калькулятором. Если щелкнуть на какой-нибудь кнопке мышью, соответствующий знак, например, +, *, 7, будет скопирован в редактор условий.
- Список функций: Содержит около 140 функций. Каждую из функции можно скопировать в редактор условий двойным щелчком.

7.1.1. Классификация операторов

Операторы делятся на арифметические, логические и операторы отношения. Арифметические операторы применяются в так называемых арифметических выражениях (математических

формулах), которые при отборе данных имеют лишь второстепенное значение. Арифметические операторы всегда можно использовать в логических выражениях, однако это встречается нечасто. Решающую роль эти операторы играют при модификации данных; поэтому они и описаны в разделе 8.1, посвященном модификации данных.

Логические операторы и операторы отношения применяются исключительно в логических выражениях, которые рассматриваются в настоящей главе.

7.1.2. Операторы отношения

Отношение — это логическое выражение, в котором два значения сравниваются друг с другом посредством оператора отношения. В областях, где применяется SPSS в операторах отношения значения переменной сравниваются с каким-либо численным значением (константой), например

`sex = 2 parti ~ = 3 alter > 30`

Для построения логических выражений могут применяться следующие операторы отношения:

Знак на кнопке	Альтернативный текст	Значение (рус./англ.)
<	LT	меньше (less than)
>	GT	больше (greater than)
<=	LE	меньше или равно (less than or equal to)
>=	GE	больше или равно (greater than or equal to)
=	EQ	равно (equal to)
~ =	NE или <>	не равно (not equal to)

Операторы можно ввести в редактор условий либо щелкнув в диалоговом окне на кнопке с соответствующим знаком, либо введя с клавиатуры альтернативный текст. Например, вместо ~ = можно ввести NE или <>.

7.1.3. Логические операторы

Для построения условных выражений могут применяться следующие логические операторы:

Знак на кнопке	Альтернативный текст	Значение
&	AND	Логическое И
	OR	Логическое ИЛИ
~	NOT	Логическое НЕ

Логические операторы AND и OR связывают два отношения, а логический оператор NOT меняет значение истинности условного выражения на противоположное. Между логическими операторами устанавливаются следующие приоритеты:

Приоритет	Оператор
1	NOT
2	AND
3	OR

7.1.4. Булева алгебра

Логические операторы основаны на принципах булевой алгебры (логики высказываний), краткий обзор которых приводится в данном разделе.

Оператор И (конъюнкция)

Выражение 1	Выражение 2	Результат
и	и	и
и	л	л
л	и	л
л	л	л

Легенда: и = истина (true); л = ложь (false)

При конъюнкции все участвующие выражения (отношения) должны быть истинными, чтобы общий результат также являлся истинным. Примеры:

Выражение	Истинность
$(3 < 7) \text{ AND } (8 > 5)$	и
$(12 = 8) \text{ AND } (4 = 4)$	л
$(3 \leq 5) \text{ AND } (4 >= 1)$	и
$(8 = 4) \text{ AND } (7 = 3)$	л

Оператор ИЛИ (дизъюнкция)

Выражение 1	Выражение 2	Результат
и	и	и
и	л	и
л	и	и
л	л	л

При дизъюнкции хотя бы одно из участвующих отношений должно быть истинным, чтобы общий результат также был истинным. Примеры:

Выражение	Истинность
$(3 < 5) \text{ OR } (47 + 1 < 10)$	и
$(3 = 8) \text{ OR } (7 > 5)$	и
$(4 : 7 = 2) \text{ OR } (8 * 4 = 21)$	л
$(42 = 16) \text{ OR } (23 = 3)$	и

Логическое НЕ (отрицание)

Выражение	Результат
и	л
л	и

Отрицание меняет истинность выражения на противоположную. Примеры:

Выражение	Истинность
NOT [(3<5) AND (4>5)]	и
NOT [(4<5) AND (8<12)]	л

При отрицании следует учитывать эквивалентность операторов:

отрицаемый оператор	эквивалентный оператор
<	>
>	< =
< =	>
> =	<

В заключение приведем пример более сложного логического выражения: [(NOT A) AND (NOT B)] OR C

Согласно правилам приоритета скобки здесь не нужны. Мы поместили их только для повышения наглядности. Истинность выражения можно определить при помощи следующей таблицы:

A	И	C	NOT A	NOT B	(NOT A) AND (NOT B)	OR C
и	и	и	л	л	л	и
и	и	л	л	л	л	л
и	л	и	л	и	л	и
и	л	л	л	л	л	л
л	и	и	и	л	л	и
л	и	л	и	л	л	л
л	л	и	и	и	и	и
л	л	л	и	и	и	и

Для более сложных выражений также следует составлять подобные таблицы.

Если все эти элементы логики высказываний кажутся вам слишком математизированными или абстрактными, вполне можно ориентироваться по разговорному употреблению союза "и". Высказывание: "Я был в кино и видел интересный фильм", истинно тогда и только тогда, когда истинны обе его части. Если, несмотря на то, что вы ходили в кино, но на сеансе заснули от скуки, это выражение не будет истинным. Также оно не будет истинным, если вы смотрели интересный фильм по телевизору. И, конечно же, оно будет совершенно ложным (хотя здесь нас не интересует степень ложности), если вы и не были в кино, и не смотрели там интересный фильм.

Иначе обстоит дело при разговорном применении союза "или", которое в основном означает исключающее "или", когда, например, дети хотят получить на Рождество или компьютер, или велосипед.

7.1.5. Функции

Список функций, который мы сейчас рассмотрим, — следующая важная часть диалогового окна Select Cases: If.

Этот список содержит множество математических функций, большая часть из которых, однако, имеет отношение только к модификации данных (расчету новых переменных). Поэтому обзор этих функций представлен в соответствующем разделе (см. раздел 8.1.2). Здесь мы рассмотрим только логические и строковые функции.

Логические функции

В SPSS реализованы две логические функции:

- **RANGE (variable, begin, end):** Функция RANGE возвращает значение 1, или true, если значение переменной лежит в диапазоне между заданными начальным и конечным значениями. Переменная может иметь как численный, так и строковый тип. RANGE (alter, 18, 22) возвращает значение 1, то есть true, если значение переменной alter лежит между 18 и 22 включительно. Можно задавать несколько диапазонов, например, RANGE (alter, 1,17, 63, 99). В этом случае функция возвращает true, если значение переменной alter лежит между 1 или 17 или между 63 и 99 включительно. В функции RANGE можно также использовать переменные строкового типа, например, RANGE (name, A, Mzzzzzz). Тогда функция будет возвращать 1 для имен, начинающихся с букв от A до M включительно. Если имя начинается с другой буквы, функция возвратит 0.
- **ANY (variable, val1, val2, val3,...):** Функция ANY возвращает значение 1, или true, если значение переменной (значение первого аргумента) совпадает по крайней мере с одним из значений, указанных в последующем списке параметров (val1, val2, val3, ...). В противном случае возвращается значение 0 или false. Первый элемент, как правило, — переменная численного или символьного типа. Примеры: ANY (jahr, 1991, 1992, 1993, 1994) возвращает true, если значение переменной jahr равно 1991, 1992, 1993 или 1994. ANY (name, Schmidt, Meier, Raabe) возвращает значение true или 1 в тех случаях, когда переменная name содержит значения Schmidt, Meier или Raabe. Во всех остальных случаях возвращается значение 0. Не забывайте заключать строковые значения в двойные кавычки.

Строковые функции

Из общего количества 18 строковых функций мы рассмотрим три самых важных, на наш взгляд.

- **SUBSTR (variable, begin, length):** Эта функция извлекает определенную часть из строки. Она возвращает подстроку или отдельный символ. Например, если строковая переменная name содержит значение Mannheim, то следующий вызов функции

```
< class="listing"> SUBSTR (name, 1, 2)
```

возвратит значение Ma. Здесь из переменной name извлекаются два знака (третий аргумент) начиная с первой позиции (второй аргумент). Выражение

```
SUBSTR (name, 1, 2) = Ma
```

будет истинным для значений переменной Maus, Mannesmann или Mahlmann. При сравнении со строками вместо двойных кавычек (= "Ma") можно также применять простые (= 'Ma'). Однако смешение простых и двойных кавычек (= 'Ma') не допускается.

- **UPCASE (argument):** Функция UPCASE преобразует строчные буквы в прописные. В качестве аргумента можно задавать строку или переменную символьного типа. UPCASE (vorname) возвращает значение ANNA, если переменная vorname имеет значение Anna.
- **LOWER (argument):** Функция LOWER преобразует прописные буквы в строчные.

В качестве параметра можно задавать строку или переменную символьного типа.

- **LOWER (vorname)** возвращает значение anna, если переменная vorname имеет значение ANNA или Anna.

Функции переносятся в редактор условий следующим образом:

- Поместите курсор на место в условном выражении, на котором должна быть вставлена функция.
- Дважды щелкните на функции в списке функций или выделите функцию и щелкните на кнопке с треугольником около списка функций.

Функция будет вставлена в выражение. Вместо аргументов в этой функции будут стоять вопросительные знаки. Количество вопросительных знаков указывает минимальное количество аргументов, которое следует вставить. Отредактировать функцию можно следующим образом:

- Выделите вопросительные знаки во вставленной функции.
- Замените их соответствующими аргументами. Имена переменных для аргументов можно перенести из списка исходных переменных.

В заключение мы составим список приоритетов при построении логических выражений:

Приоритет	Оператор/функция	Значение
1	()	Оператор скобок
2	Функции	Различные значения
3	<	Меньше
	<=	Меньше или равно
	>	Больше
	>=	Больше или равно
	=	Равно
	"="	Не равно
4	—	Логическое НЕ
5	&,	Логическое И
6		Логическое ИЛИ

7.1.6. Ввод условного выражения

Теперь попробуем снова выполнить отбор, но в этот раз будем выбирать только респондентов-женщин. Выполните следующие действия:

- Перенесите переменную *sex* в редактор условий, дважды щелкнув на ней или выделив ее и щелкнув на кнопке с треугольником.
- Щелкните на кнопке со знаком равенства на клавиатуре. Этот знак будет скопирован в редактор условий.
- Щелкните на кнопке 1 на клавиатуре. Знак будет скопирован в редактор условий. Вид диалогового окна показан на рис. 7.3.

Условие имеет вид $sex = 1$, то есть будут выбраны все наблюдения, для которых переменная *sex* имеет значение 1 (женский).

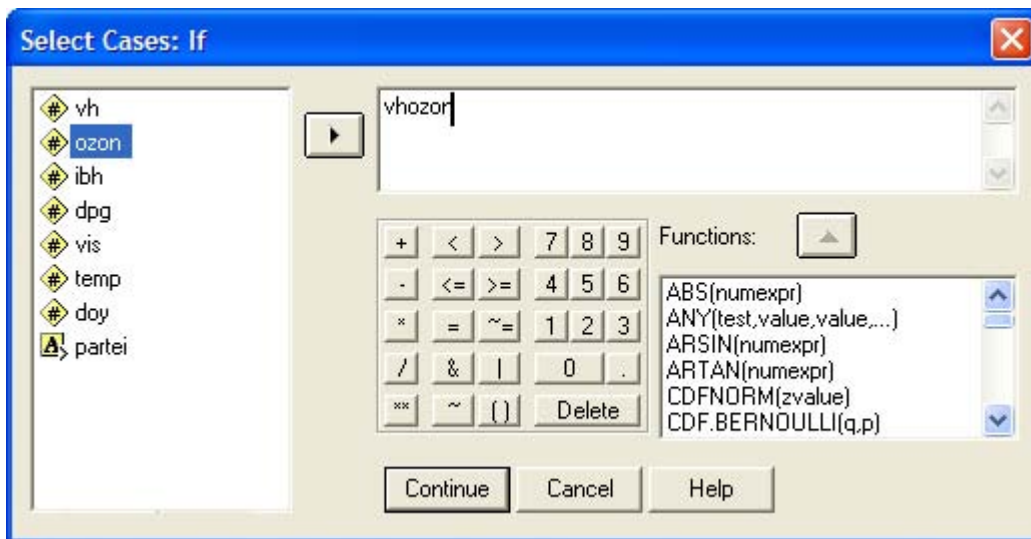


Рис. 7.3: Условие в редакторе условий

- Подтвердите выбор кнопкой Continue (Продолжить). Вы вернетесь в диалог Select Cases. Однако теперь в диалоговом окне появилось условие `sex = 1`.
- Щелкните на кнопке ОК. Вы снова окажетесь в редакторе данных.

Примечание: Выбранные опции соответствуют следующему командному синтаксису:

```
SELECT IF sex = 1.
```

```
EXECUTE
```

Теперь фильтрация наблюдений включена. О том, что отбор, заданный с помощью диалоговых окон осуществлен свидетельствует сообщение Filter on (Фильтр включен), которое появляется в строке состояния в нижней части окна SPSS. Система создает переменную `filter_S`. Это численная переменная с длиной один байт. Она имеет следующие метки значений: 0 = Not Selected (Не выбрано), 1 = Selected (Выбрано), так как нуль обозначает ложь (false), а единица — истину (true). При всех последующих операциях будут учитываться только наблюдения, для которых значение этой переменной равно 1, то есть те, для которых выполняется условие `sex = 1`. Номера неотобранных наблюдений отображаются зачеркиванием в левом крае редактора данных. Теперь проведем частотный анализ переменной `partei`. Мы получим следующий результат:

Партия

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ХДС/ХСС	5	33,3	35,7	35,7
	СДП	1	6,7	7,1	42,9
	сдпг	4	26,7	28,6	71,4
	Зеленые/Союз 90	2	13,3	14,3	85,7
	пдс	1	6,7	7,1	92,9
	Прочие	1	6,7	7,1	100,0
	Всего	14	93,3	100,0	
Missing Total	нет данных	1 15	6,7 100,0		

Из 30 наблюдений файла `wahl.sav` условие отбора выполняется в 15 наблюдениях; для них `sex = 1`. Эти 15 наблюдений и учитываются при частотном анализе переменной `partei`. Для одного из отобранных наблюдений данных о партии нет.

Обратите внимание, что фильтр действует и при остальных статистических процедурах. Команда SPSS SELECT IF или соответствующие настройки в диалоговых окнах фильтруют наблюдения постоянно, то есть до тех пор, пока фильтр не будет удален или деактивирован. Чтобы удалить фильтр, поступите следующим образом:

- Щелкните на имени переменной filter_\$. Весь столбец будет выделен.
- Нажмите клавишу <Backspace>. Переменная фильтра будет удалена.

Если требуется не удалять фильтр, а лишь временно деактивизировать его, выполните следующие действия:

- Выберите в меню команды Data (Данные) elect Cases... (Выбрать наблюдения)
- В диалоговом окне Select Cases щелкните на кнопке All cases (Все наблюдения). Условие фильтра будет деактивировано, однако переменная filter_S сохранится. В любой момент ее можно будет активизировать снова.

На уровне синтаксических команд отбор можно выполнить при помощи единственной процедуры, которая показана ниже. Для этого применяется команда TEMPORARY:

```
TEMPORARY.
```

```
SELECT IF sex = 1.
```

```
FREQUENCIES
```

```
VARIABLES = partei.
```

Временный фильтр можно ввести только вручную в редакторе синтаксиса SPSS; через диалоговые окна этого сделать невозможно. Этот пример показывает, что непосредственный ввод команд в редакторе синтаксиса имеет некоторые преимущества. Об этом мы еще расскажем в главе 26 (Программирование).

При вводе команд в редакторе синтаксиса следует обращать внимание на различие между численными и строковыми переменными.

Численная переменная:

```
SELECT IF sex = 1.
```

Строковая переменная:

```
SELECT IF fragebnr = "W-001".
```

Для строковых переменных (как fragebnr (код анкеты) в этом примере) следует применять простые или двойные кавычки. Слова SELECT IF необходимы только при непосредственном вводе команды в редакторе синтаксиса; та же самая строка в редакторе условий диалога Select Cases: If будет более компактной:

```
sex = 1
```

ИЛИ

```
fragebnr = "W-001"
```

Здесь также следует учитывать различие между численными и строковыми переменными.

7.1.7. Примеры отбора данных

Здесь мы представим некоторые примеры отбора данных. Рассмотрим следующие условия:

1. Требуется отобразить только респондентов-мужчин. В редакторе условий вводится следующая строка:

```
sex = 2
```

Эту строку можно набрать непосредственно или перенести с помощью кнопки с треугольником и кнопок клавиатуры.

2. Требуется отобразить только респондентов-женщин, которые голосовали за ХДС/ХСС. В редакторе условий вводится следующая строка:

```
sex = 1 & partei = 1
```

ИЛИ

```
sex = 1 AND partei = 1
```

Обратите внимание на значение переменной фильтра в наблюдении 22 (fragebnr = 0-007). Здесь это системное пропущенное значение. В этом случае SPSS не может сделать никакого вывода об истинности, так как переменная partei имеет значение 0 = нет данных или данные не введены. Поэтому условие `sex = 1 & partei = 1` в наблюдении 22 нельзя проверить на истинность. Оно может быть как истинным, так и ложным. Для такого неопределенного случая SPSS присваивает переменной filter_\$ системное пропущенное значение.

Следовательно, таблицу истинности можно дополнить случаем отсутствующих значений:

Конъюнкция

Логическое выражение	Результат
true AND true	true
true AND false	false
false AND true	false
false AND false	false
true AND missing	missing
false AND missing	false
missing AND missing	missing

Дизъюнкция

Логическое выражение	Результат
true OR true	true
true OR false	true
false OR true	true
false OR false	false
true OR missing	true
false OR missing	missing
missing OR missing	missing

Отрицание:

Логическое выражение	Результат
true	false
false	true
missing	missing

Если результат логического выражения равен missing (отсутствует), то данный случай, как и при результате false, не учитывается при дальнейшей обработке.

3. Требуется отобрать только респондентов, имеющих возраст от 40 до 60 лет включительно.

```
alter >= 40 & alter <= 60
```

ИЛИ

```
alter >= 40 AND alter <= 60
```

Более изящным будет применение здесь функции

```
RANGE: RANGE (alter, 40, 60).
```

4. Требуется отобрать только респондентов-женщин, которые старше 60 лет. `sex =1 & alter > 60`
ИЛИ `sex =1 AND alter > 60`.

5. Требуется отобрать только респондентов-мужчин, возраст которых не превышает 25 лет и которые голосовали за СДПГ. При формулировке условия не старше 25 лет применяется оператор NOT:

```
sex = 2 S partei = 3 & ~ alter > 25
```

ИЛИ

```
sex = 2 S partei = 3 S NOT alter > 25.
```

Оператор NOT обязательно должен стоять в начале логического выражения. Выражение `& alter ~> 25` не допускается в SPSS. В этом случае вы получите сообщение об ошибке с подсказкой, где должен находиться оператор NOT.

6. Требуется отобрать респондентов, которые голосовали за ХДС, СДП или республиканцев.

```
partei = 1 | partei = 2 | partei = 6
```

ИЛИ

```
partei = 1 OR partei = 2 OR partei = 6.
```

Здесь более изящным будет применение функции ANY:

```
ANY (partei, 1, 2, 6).
```

7. Отберем респондентов, которые опрашивались в Западной Германии:

```
fragebnr >= "W-"
```

Здесь более изящным будет применение функции SUBSTR:

```
SUBSTR (fragebnr,1,1) = "W"
```

ИЛИ

```
SUBSTR (fragebnr,1,2) = "W-"
```

Можно также применить функцию RANGE:

```
RANGE (fragebnr, W-001, W-999)
```

ИЛИ

```
RANGE (fragebnr, "W-001", "W-999").
```

8. Отберем респондентов, которые опрашивались в Восточной Германии: `fragebnr >= "0-" & fragebnr < "W-"`

Достаточно также просто ввести

```
fragebnr < "W-"
```

И в этом случае изящнее будет вариант с SUBSTRING:

```
SUBSTR(fragebnr,1,1) = "0"
```

или

```
SUBSTR(fragebnr,1,2) = "0-"
```

Можно также применить функцию RANGE:

```
RANGE (fragebnr, "0-001", "0-999")
```

Удобно использовать оператор NOT:

```
~ fragebnr >= "W"
```

Далее мы рассмотрим применение функций UPCASE и LOWER. При этом будем исходить из следующей ситуации.: При вводе номеров анкет иногда по ошибке вме-^то прописного "W" для Западной Германии было закодировано строчное "w". Эти наблюдения не будут отобраны по условию `SUBSTR(fragebnr, 1,1) = "W"`. В таком ^лучае может помочь функция UPCASE или LOWER:

```
SUBSTR (UPCASE (fragebnr,1,1) = "W").
```

Рассмотренная конструкция называется вложенной функцией. Вложенные функции вычисляются в направлении изнутри наружу. Функция UPCASE преобразует содержимое переменной `fragebnr` в прописные буквы. Преобразованное содержимое затем передается в функцию SUBSTR. Эта функция выделяет из строки первую букву. Полученная буква сравнивается с буквой W. Если они совпадают, данное наблюдение выбирается, то есть переменная фильтра `filter_S` приобретает значение 1. Если применяется функция LOWER, строка в редакторе условий будет выглядеть так:

```
SUBSTR (LOWER (fragebnr, 1,1) = "w").
```

Функция LOWER преобразует содержимое переменной `fragebnr` в строчные буквы. Преобразованное содержимое передается в функцию SUBSTR. Эта функция выделяет из строки первую букву. Полученная буква сравнивается с буквой w. Если они совпадают, данное наблюдение отбирается.

7.2. Извлечение случайной выборки

При большом количестве наблюдений для экономии времени может быть полезно использовать небольшую случайную выборку при первой предварительной проверке гипотезы. Чтобы извлечь случайную выборку из совокупности всех наблюдений, выполните следующие действия:

- Выберите в меню команды Data (Данные) Select Cases... (Выбрать наблюдения)
- Выберите пункт Random sample of cases (Случайная выборка), а затем щелкните на кнопке Sample... (Выборка). Откроется диалоговое окно Select Cases: Random Sample (Выбрать наблюдения: Случайная выборка).

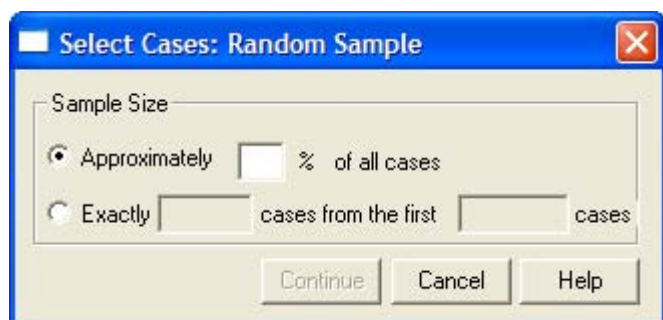


Рис. 7.4: Диалоговое окно *Select Cases: Random Sample*

В группе Sample Size (Размер выборки) можно выбрать один из следующих способов определения объема выборки:

- **Approximately (Приблизительно):** Пользователь может указать здесь процентного значение. SPSS создаст случайную выборку с объемом, приблизительно соответствующим указанному проценту наблюдений.
- **Exactly (Точно):** Пользователь должен указать здесь точное количество наблюдений в случайной выборке. Кроме того, здесь надо задать количество наблюдений, из которых будет извлечена выборка. Второе число не должно превышать общего количества наблюдений в файле данных. Для каждой случайной выборки генератор случайных чисел SPSS использует новое начальное значение. Таким образом, каждый раз при обращении к данному диалогу создается новая выборка наблюдений, отличная от прежних. Если требуется, чтобы случайная выборка повторялась, надо задать начальное значение самостоятельно.
- Для этого выберите в меню команды Transform (Преобразовать) Random Number Seed... (Установить начальное положение генератора случайных чисел)

Откроется диалоговое окно Random Number Seed.

Начальное значение может быть любым положительным целым числом. Это значение можно задать самостоятельно или предоставить сделать это SPSS (вариант Random Seed, принятый по умолчанию).



Рис. 7.5: Диалоговое окно *Random Number Seed*.

7.3. Сортировка наблюдений

Данные в SPSS можно сортировать в соответствии со значениями одной или нескольких переменных. Рассмотрим следующий пример: Требуется упорядочить данные файла wahl.sav по возрасту. Для этого поступите следующим образом:

- Выберите в меню команды Data (Данные) Sort Cases... (Сортировать наблюдения)

Откроется диалоговое окно Sort Cases. Переменные файла данных будут отображены в списке исходных переменных.

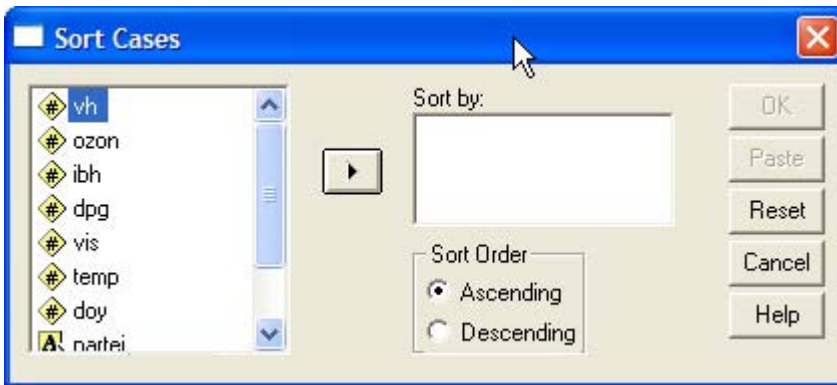


Рис. 7.6: Диалоговое окно Sort Cases

- Перенесите переменную alter в список Sort by (Сортировать по). В группе Sort order (Порядок сортировки) по умолчанию выбран вариант Ascending (По возрастанию). Эта опция сортирует наблюдения в порядке возрастания значения переменной сортировки, а следующая опция, Descending — в порядке убывания.
- Подтвердите настройки кнопкой ОК. В редакторе данных файл wahl.sav будет отсортирован по возрастанию значений переменной alter.

Примечание: Выбранные опции соответствуют следующему командному синтаксису:

```
SORT CASES BY alter (A) .
```

или, если надо сортировать по убыванию:

```
SORT CASES BY alter (D) .
```

Здесь A обозначает ascending (возрастание), а D — descending (убывание). Если выбрать несколько переменных сортировки, их последовательность в списке Sort by будет определять порядок, в котором будут отсортированы наблюдения. Рассмотрим следующий пример: Необходимо отсортировать файл wahl.sav по значениям переменных partei и alter. Переменная partei должна быть первым критерием сортировки, а переменная alter — вторым. Сортировка по переменной partei должна быть в порядке возрастания, а по переменной alter — в порядке убывания. Для этого перенесите в список переменных сортировки вначале переменную partei, а затем переменную alter. Выделите переменную alter и щелкните на опции Ascending.

Примечание: Выбранные опции соответствуют следующему командному синтаксису:

```
SORT CASES BY partei (A) alter (D) .
```

В редакторе данных файл wahl.sav будет отсортирован по возрастанию значений переменной partei. Наблюдения, относящиеся к одной и той же партии будут отсортированы по убыванию возраста.

7.4. Разделение наблюдений на группы

В SPSS можно выполнять анализ данных отдельно по группам. Группой в этом контексте называется определенное количество наблюдений с одинаковыми значениями признаков. Чтобы можно было производить обработку по группам, файл должен быть отсортирован по группирующему переменным. Такой переменной может быть, например, переменная sex. В этом случае все переменные со значением признака 1 (женский) образуют одну группу, а все переменные со значением признака 2 (мужской) — другую группу. С каждой группой можно проводить определенные операции, например, выполнять частотный анализ. При этом частотный анализ проводится отдельно для признаков мужской и женский. В SPSS такое разделение на группы можно выполнять автоматически. Рассмотрим следующий пример, основанный на опросе студентов об их психическом состоянии и социальном положении:

Проведем частотный анализ переменной psuche (психическое состояние) отдельно для всех изучаемых специальностей. В соответствии со значениями переменной fach (специальность) у нас образуются 9 групп (1 = Юриспруденция, 2 = Экономика, 3 = Гуманитарные науки, 4 = Психология и т.д.). В этом случае файл данных studium.sav должен быть сначала отсортирован по переменной fach. Поступите следующим образом:

- Загрузите файл studium.sav в редактор данных.
- Выберите в меню команды Data (Данные) Split File... (Разделить файл) Откроется диалоговое окно Split File.

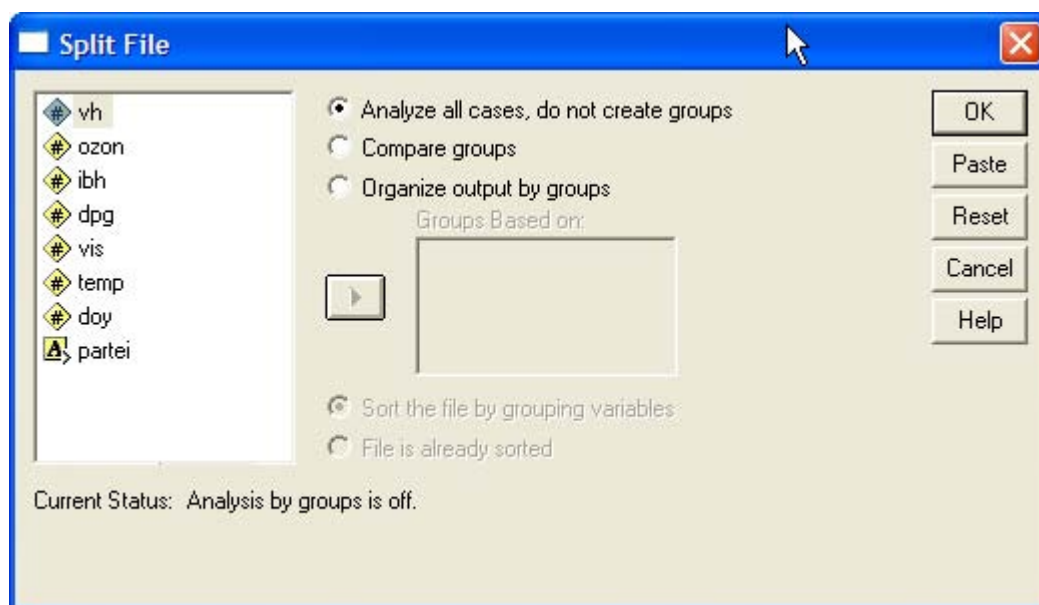


Рис. 7.7: Диалоговое окно Split File

По умолчанию разделение на группы не предполагается. Если выбрать пункт Organize output by groups (Разделить вывод на группы), мы получим вывод результатов по каждой группе отдельно. Эти группы должны быть определены в поле Groups based on (Группы, созданные на основе) на базе соответствующих переменных.

Еще одну возможность предоставляет опция Compare Groups (Сравнить группы). Она организует вывод таким образом, что можно визуально сравнить разные группы друг с другом. Но сначала мы рассмотрим отдельный вывод.

- Выберите опцию Organize output by groups. Для отдельного выполнения операций по группам необходимо, чтобы файл данных был предварительно отсортирован по этим группирующим переменным. По этой причине опция Sort the file by grouping variables (Сортировать файл по группирующим переменным) выбрана по умолчанию.

- Перенесите переменную fach в поле Groups based on. Если выбирается несколько группирующих переменных, то последовательность, в которой они стоят в списке, определяет порядок или приоритет сортировки.
- Щелкните на кнопке ОК. Файл studium.sav будет отсортирован по переменной fach, то есть разбит на группы в соответствии с ее значениями. Сообщение File split on (Разделение файла включено) в строке состояния внизу окна SPSS информирует об активации режиме разделения.
- Выполните частотный анализ переменной psuche.

Вы получите следующий результат (ниже для экономии места показаны частотные таблицы только для специальностей Юриспруденция и Естественные науки).

Специальность = Юриспруденция

Статистика^(a)

Психическое состояние	
N Valid	22
Missing	0
a. Специальность = Юриспруденция	

Психическое состояние^(a)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Крайне неустойчивое	2	9,1	9,1	9,1
	Неустойчивое	5	22,7	22,7	31,8
	Устойчивое	12	54,5	54,5	86,4
	Очень устойчивое	3	13,6	13,6	100,0
	Total	22	100,0	100,0	
Специальность = юриспруденция					

Специальность = Естественные науки

Статистика^(a)

N Valid		18
Missing		1
a. Специальность = Естественные науки		

Психическое состояние^(a)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Крайне неустойчивое	1	5,3	5,6	5,6
	Неустойчивое	4	21,1	22,2	27,8
	Устойчивое	11	57,9	61,1	88,9
	Очень устойчивое	2	10,5	11,1	100,0
	Всего	18	94,7	100,0	
Missing	нет данных	1	5,3		
Всего			19	100,0	

а. Специальность = Естественные науки

Как видно, результаты частотного анализа переменной psyche выводятся отдельно по специальностям студентов.

- Теперь выберите другой пункт — Compare groups (Сравнить группы).
- Снова выполните частотный анализ переменной psyche. Вы получите следующую результирующую таблицу:

Психическое состояние

Специальность			Frequency	Percent	Valid Percent	Cumulative Percent
Юриспруденция	Valid	Крайне неустойчивое	2	9,1	9,1	9,1
		Неустойчивое	5	22,7	22,7	31,8
		Устойчивое	12	54,5	54,5	86,4
		Очень устойчивое	3	13,6	13,6	100,0
		Всего	22	100,0	100,0	
Экономика	Valid	Крайне неустойчивое	1	5,3	5,6	5,6
		Неустойчивое	4	21,1	22,2	27,8
		Устойчивое	11	57,9	61,1	88,9
		Очень устойчивое	2	10,5	11,1	100,0
		Всего	18	94,7	100,0	
	Missing	нет данных	1	5,3		
	Total	19	100,0			
Гуманитарные науки	Valid	Крайне неустойчивое	10	40,0	40,0	40,0
		Неустойчивое	14	56,0	56,0	96,0
		Устойчивое	1	4,0	4,0	100,0
		Всего	25	100,0	100,0	
Психология	Valid	Крайне неустойчивое	3	27,3	27,3	27,3
		Неустойчивое	6	54,5	54,5	81,8
		Устойчивое	2	18,2	18,2	100,0
		Всего	11	100,0	100,0	
Теология	Valid	Крайне неустойчивое	2	22,2	22,2	22,2
		Неустойчивое	5	55,6	55,6	77,8
		Устойчивое	2	22,2	22,2	100,0
		Всего	9	100,0	100,0	
Медицина	Valid	Крайне неустойчивое	1	10,0	10,0	10,0
		Неустойчивое	3	30,0	30,0	40,0
		Устойчивое	5	50,0	50,0	90,0
		Очень устойчивое	1	10,0	10,0	100,0
		Всего	10	100,0	100,0	

Естественные науки	Valid	Неустойчивое	3	33,3	33,3	33,3
		Устойчивое	6	66,7	66,7	100,0
		Всего	9	100,0	100,0	
Техника	Valid	Крайне неустойчивое	1	50,0	50,0	50,0
		Устойчивое	1	50,0	50,0	100,0
		Всего	2	100,0	100,0	
Прочие	Valid	Устойчивое	1	100,0	100,0	100,0

Учтите, что файл данных останется разделенным на подгруппы, пока вы не деактивируете соответствующие опции. Для этого поступите следующим образом:

- Выберите в меню команды Data (Данные) Split File... (Разделить файл)
- В диалоговом окне Split File выберите опцию Analyze all cases, do not create groups (Анализировать все наблюдения, не создавать группы). Теперь разделение файла убрано.
- Если требуется дополнительно убрать сортировку по специальностям, выберите в меню следующие команды Data (Данные) Sort Cases... (Сортировать наблюдения)
- Перенесите переменную fragebnr (код анкеты) в список переменных сортировки и подтвердите операцию кнопкой ОК. Данные будут отсортированы в исходном порядке — по номерам анкет.

На этом мы заканчиваем обзор возможностей отбора данных в SPSS и переходим к изучению средств модификации данных.