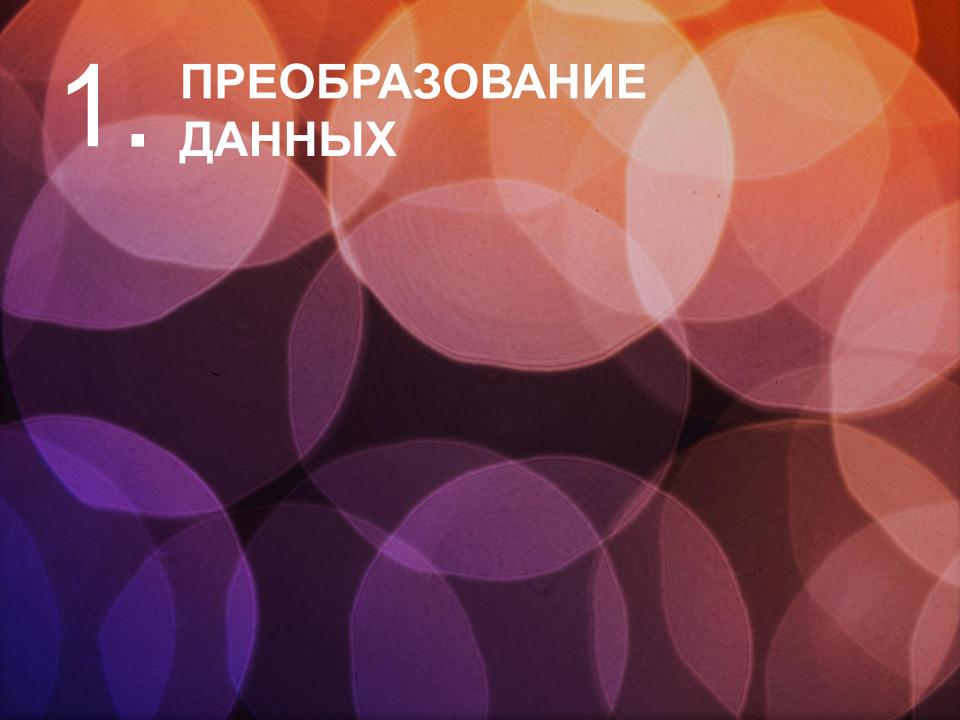


ОГЛАВЛЕНИЕ

- 1. Преобразование данных
 - 1.1. Вычисление новых переменных
 - 1.2. Перекодирование данных
- 2. Сортировка наблюдений
- 3. Отбор данных для анализа
- 4. Объединение данных
- 5. Взвешивание данных



Преобразование данных

Преобразование (модификация) данных – представляет собой создание новых переменных или изменение кодирования первоначально собранных данных.

B SPSS существует много возможностей для преобразования данных:

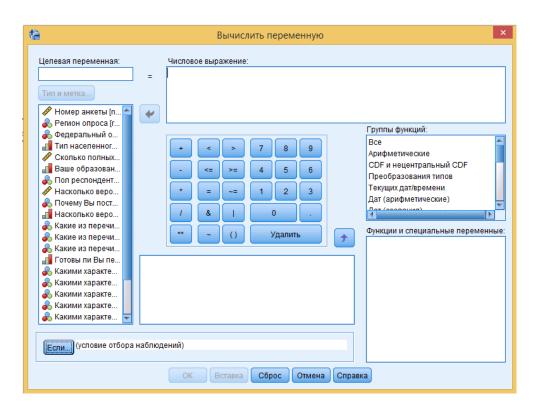
- Вычисление новых переменных путем использования различных арифметических выражений (математических формул)
- Перекодирование значений
- Вычисление новых переменных при выполнении определенного условия
- Агрегирование данных
- Ранговые преобразования
- Вычисление весов наблюдений

Рассмотрим ключевые способы модификации данных.



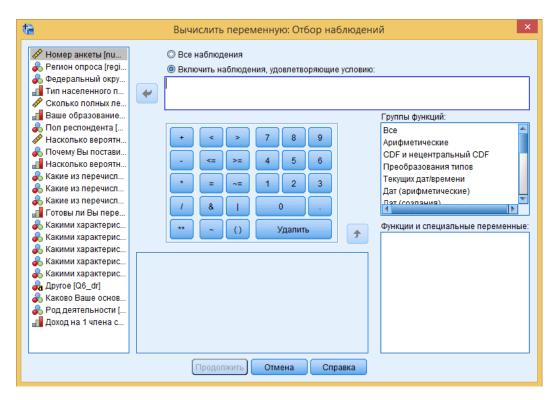
Путем вычислений в SPSS можно образовать новые переменные и добавить их в файл данных

- Загрузите файл job.sav.
- Выберите в меню «Преобразование» → «Вычислить переменную».



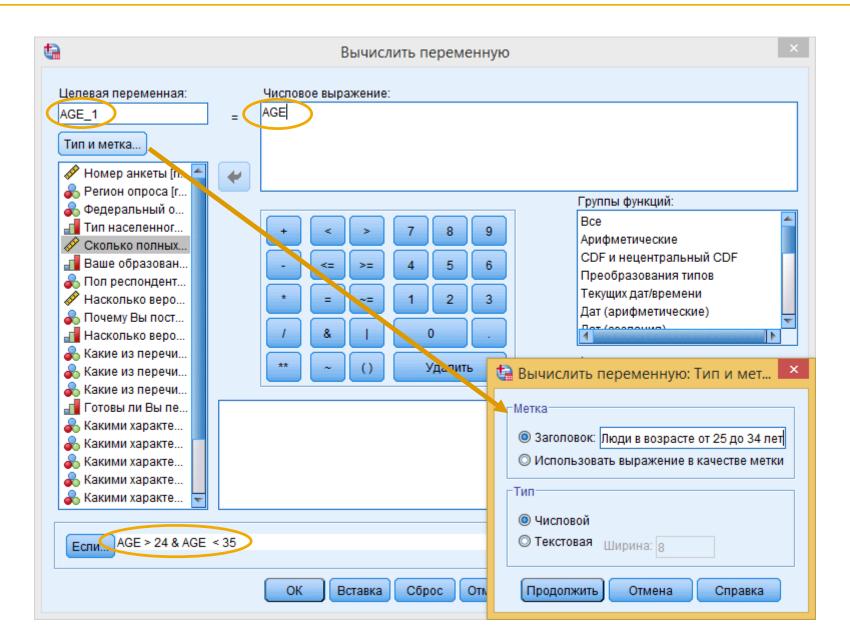
- В поле **«Целевая переменная»** указывается имя переменной (существующей или новой), которой присваивается вычисленное значение.
- В поле **«Числовое выражение»** вводится выражение, применяемое для определения значения выходной целевой переменной.
- Тип и метка можно задать тип и метку новой переменной.

- При нажатии на кнопку «Если…» всплывает новое диалоговое окно.
- В нем можно прописывать условия для новой переменной.
- Выберете «Включить наблюдения, удовлетворяющие условию» и вводите условие.



Пример: необходимо отобрать людей старше 24 и моложе 35 лет так, чтобы у них сохранились данные по возрасту.

- Выбрать «Преобразование»/ «Вычислить переменную».
- В строке «Целевая переменная» пишем название новой переменной, например, AGE_1.
- В «Если…» прописываем в строке «Включить наблюдения, удовлетворяющие условию»: AGE > 24 & AGE < 35, где переменная AGE возраст. Нажимаем «Продолжить».
- В строке «Числовое выражение» вставляем переменную AGE.
- Нажимаем на «Тип и метка». В сплывающем окне прописываем смысловое значение новой переменной: «Люди в возрасте от 25 до 34 лет».
- Нажимаем «Продолжить».
- Нажимаем ОК.



Смотрим частотное распределение по новой переменной AGE_1.

AGE_1

		Частота	Проценты	Процент допустимых	Накопленны й процент
Допустимо	25,00	22	4,4	15,9	15,9
	26,00	13	2,6	9,4	25,4
	27,00	12	2,4	8,7	34,1
	28,00	15	3,0	10,9	44,9
	29,00	9	1,8	6,5	51,4
	30,00	16	3,2	11,6	63,0
	31,00	20	4,0	14,5	77,5
	32,00	11	2,2	8,0	85,5
	33,00	7	1,4	5,1	90,6
	34,00	13	2,6	9,4	100,0
	Всего	138	27,6	100,0	
Пропущенные	Системные	362	72,4		
Bcero		500	100,0		



Перекодирование численных данных необходимо, например, когда первоначальное разнообразие исходных данных не нужно для последующего анализа. В этом случае перекодирование означает уменьшение объема обрабатываемой информации.

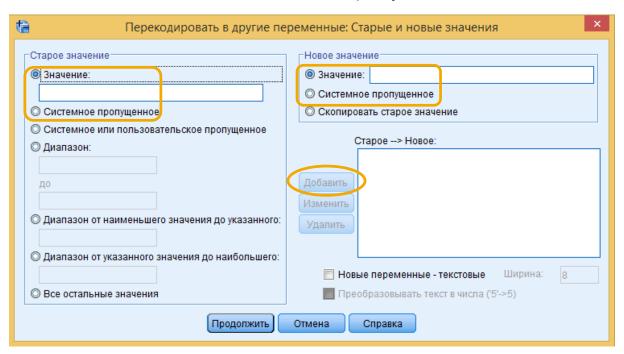
Существует два основных способа перекодирования:

- в те же переменные (старые значения заменяются новыми)
- в другие переменные (в старой переменной остаются старые значения, а в новой появляются новые, перекодированные)

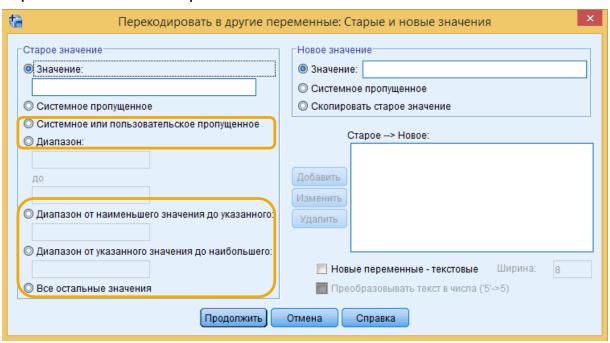
Перекодирование производится следующим образом:

- Загрузите файл job.sav в редактор данных.
- Выберите в меню команды «Преобразование» → «Перекодировать в другие переменные».
- Перенесите переменную AGE (возраст) в поле «Входная переменная» →
 «Выходная переменная». Вопросительный знак, добавленный в поле, говорит о
 том, что надо задать имя выходной переменной.
- Введите в поле **«Имя»** текст AGE_rec. Щелкните на кнопке **«Изменить».** Вопросительный знак в поле будет заменен на AGE_rec.
- Введите в поле **«Метка»** обозначение: **«Возраст (закодированный)».** Подтвердите ввод, щелкнув на **«Изменить».**
- Чтобы установить значения, которые следует перекодировать, щелкните на кнопку «Старые и новые значения». Откроется диалоговое окно «Перекодировать в другие переменные: Старые и новые значения».

- Для осуществления каждого перекодирования надо указать значение или диапазон входной переменной и соответствующее значение выходной переменной. Перекодирование завершается щелчком на кнопке «Добавить».
- После завершения во вкладке «Представление переменных» присвойте значениям метки.
- «Значение»: Вводится отдельное значение.
- «Системное пропущенное»: С помощью этой опции значение входной переменной обозначается, как системное пропущенное.



- «Системное или пользовательское пропущенное»: Эта опция служит для обозначения всех пользовательских или системных пропущенных значений. В списке значений переменных пользовательские пропущенные значения отображаются как пропущенные.
- «Диапазон»: Здесь можно задать замкнутый интервал значений.
- «Диапазон: от наименьшего до»: В этом случае будут перекодированы все значения от наименьшего наблюдаемого до указанного.
- «Диапазон: до наибольшего»: В этом случае будут перекодированы все значения от указанного до наибольшего наблюдаемого.
- «Все остальные значения»: Эта опция касается всех еще не указанных значений. В списке значений переменных они отображаются как ELSE.



20-25 = 1

26-34 = **2**

35-44 = 3

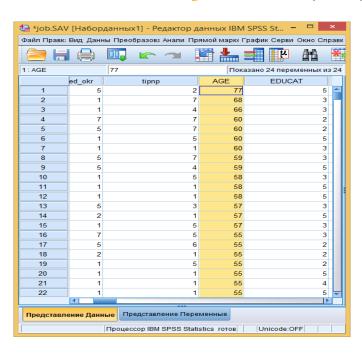
45-55 = 4

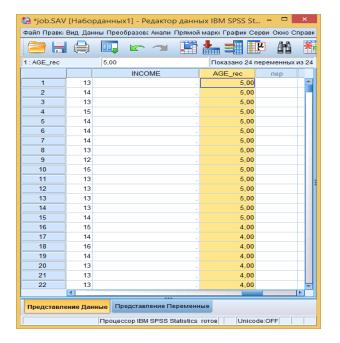
56 и выше **= 5**

Новая переменная появляется в конце массива последним столбцом.

важно:

- Всегда проверяйте перекодировку
- Следите за тем, чтобы группы не пересекались
- Перекодируйте в другие переменные для избегания потери данных
- Проверяйте наличие пустых ячеек

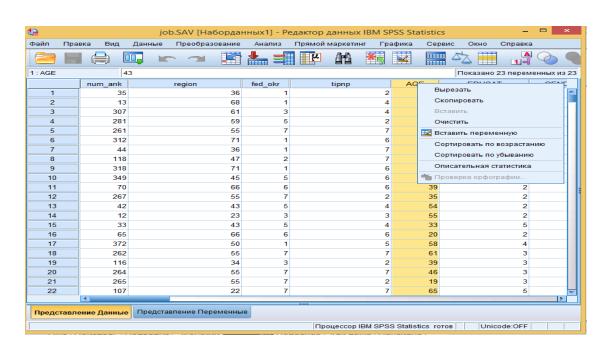






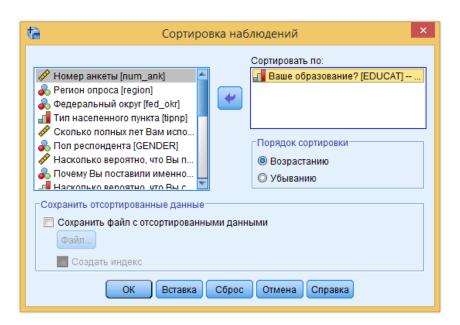
Данные в SPSS можно сортировать в соответствии со значениями одной или нескольких переменных.

«Сортировка наблюдения» - предназначена для реорганизации данных файла. Эта операция очень распространена, поскольку позволяет расположить информацию в том порядке, в котором это удобно исследователю в текущий момент.



1. Упорядочение по одной переменной:

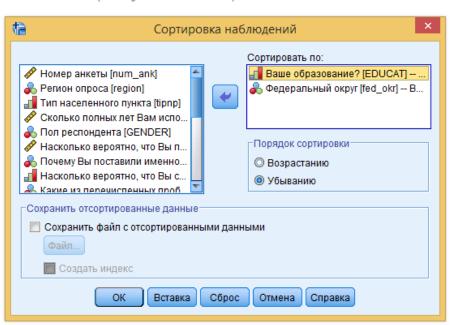
- Нажать на столбец с переменной на вкладке «Представление Данных».
- Выбрать «Сортировать по возрастанию» или «Сортировать по убыванию».
- Заметьте, что данные сортируются не внутри столбцов и строк, а целиком.



2. Упорядочение по нескольким переменным:

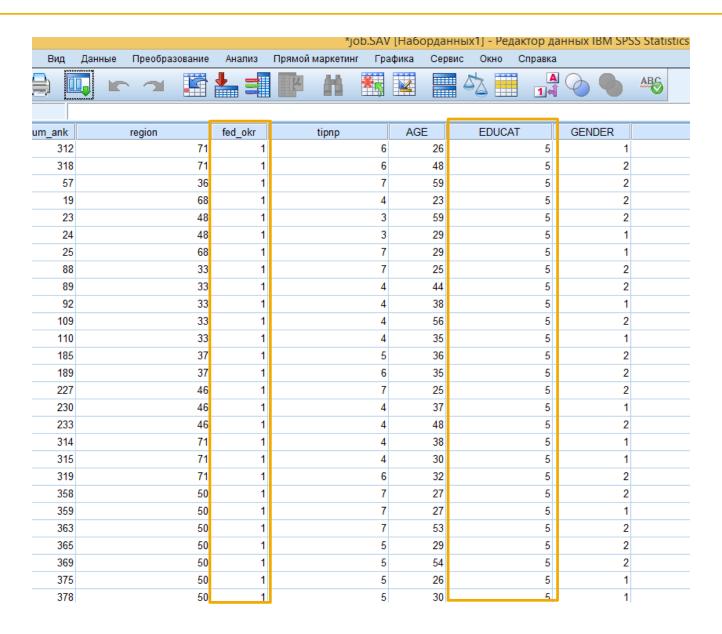
• Если выбрать несколько переменных, наблюдения сортируются по значениям каждой переменной внутри категорий, задаваемых значениями предшествующей переменной в списке сортирующих.

Пример: Требуется упорядочить данные файла job.sav по переменной fed_okr (по возрастанию) и EDUCAT (по убыванию).



Ход действий при упорядочении по нескольким переменным

- 1. Выберите в меню команду **«Данные»** → **«Сортировать наблюдения».**
- 2. Откроется диалоговое окно **«Сортировать наблюдения».** Переменные файла данных будут отображены в списке исходных переменных.
- 3. Перенесите сначала переменную fed_okr в список **«Сортировать по».** В группе «Порядок сортировки» по умолчанию выбран вариант **«По возрастанию».** Оставьте так.
- 4. Затем перенесите вправо переменную EDUCAT и установите выбор на параметре «По убыванию».
- 5. Подтвердите настройки кнопкой ОК.
- 6. Проверьте правильность сортировки во вкладке «Представление Данные».





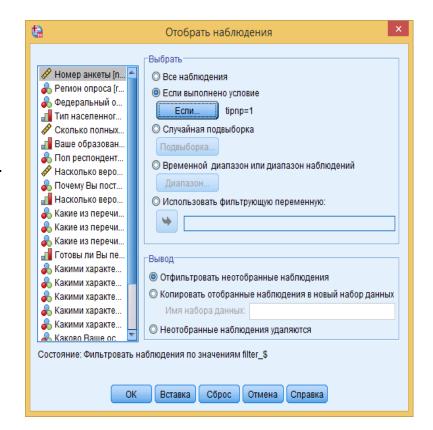
Отбор данных — это выбор наблюдений по определенным критериям. Так, например, при опросе избирателей можно отобрать только мужчин, голосующих за определенную партию, а при опросе студентов — только студенток, изучающих психологию и медицину.

После этого все вычисления будут проводиться только с этими отобранными наблюдениями.

Пример: Откройте файл "job.sav". Проведем частотный анализ переменной Q3 («Насколько вероятно, что Вы смените по собственному желанию место работы в ближайшие 6 месяцев?»). Ответы «Скорее не буду менять место работы» и «Точно не буду менять место работы» дали 53,2% опрошенных россиян.

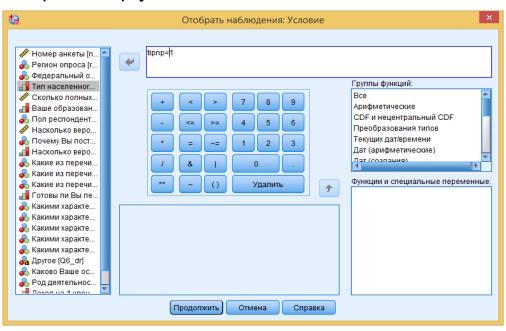
Теперь посчитаем, каковы ответы, но только среди жителей Москвы и Санкт-Петербурга.

- Выберите в меню команды «Данные»
 → «Отобрать наблюдения».
- Выберите пункт **«Если выполняется условие»** и щелкните на кнопке «Если». Откроется диалоговое окно **«Выбрать наблюдения: Если».**
- Находим слева переменную Тип населенного пункта (tipnp), находим в списке код интересующих нас городов (здесь – 1).
- Переносим переменную вправо и задаем условие tipnp=1.



Диалоговое окно разделено на следующие части:

- 1. «Список исходных переменных»: Содержит переменные, содержащиеся в открытом файле данных.
- 2. **«Редактор условий»:** Записывается логическое выражение, по которому должны быть отобраны наблюдения.
- **3. «Клавиатура»:** Содержит цифры, а также арифметические, логические операторы и операторы отношения.
- 4. «Список функций»: Содержит около 140 функций. Каждую из функции можно скопировать в редактор условий двойным щелчком.



Знак на кнопке	Значение (рус./англ.)	
<	меньше (less than)	
>	больше (greater than)	
<=	меньше или равно (less than or equal to)	
>=	больше или равно (greater than or equal to)	
=	равно (equal to)	
~=	не равно (not equal to)	

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

Фильтр в SPSS действует и при остальных статистических процедурах. Команда **«Если»** 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:

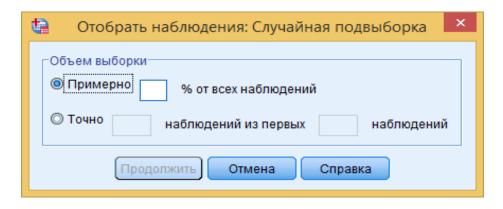
- Щелкните на имени переменной filter_\$. Весь столбец будет выделен.
- Нажмите клавишу <Backspace>. Переменная фильтра будет удалена.

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

- Выберите в меню команды «Данные» → «Выбрать наблюдения».
- В диалоговом окне щелкните на кнопке **«Все наблюдения».** Условие фильтра будет деактивировано, однако переменная filter_\$ сохранится. В любой момент ее можно будет активировать снова.

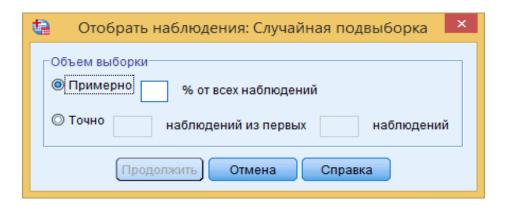
При большом количестве наблюдений для экономии времени может быть полезно использовать небольшую случайную выборку при первой предварительной проверке гипотезы. Чтобы извлечь случайную выборку из совокупности всех наблюдений:

- Выберите «Данные» → «Выбрать наблюдения».
- Выберите пункт **«Случайная подвыборка».**



Можно выбрать один из следующих способов определения объема выборки:

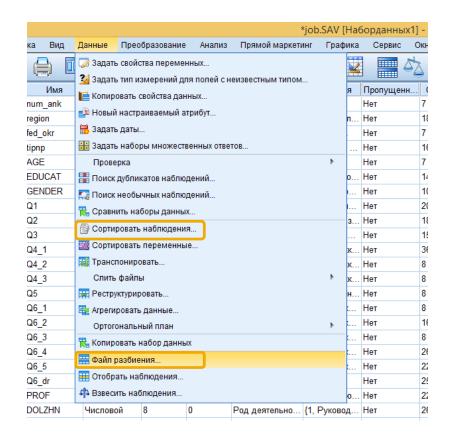
- «Примерно»: Пользователь может указать здесь процентное значение. SPSS создаст случайную выборку с объемом, приблизительно соответствующим указанному проценту наблюдений.
- «Точно»: Пользователь должен указать точное количество наблюдений в случайной выборке. Также надо задать количество наблюдений, из которых будет извлечена выборка. Второе число не должно превышать общего количества наблюдений.



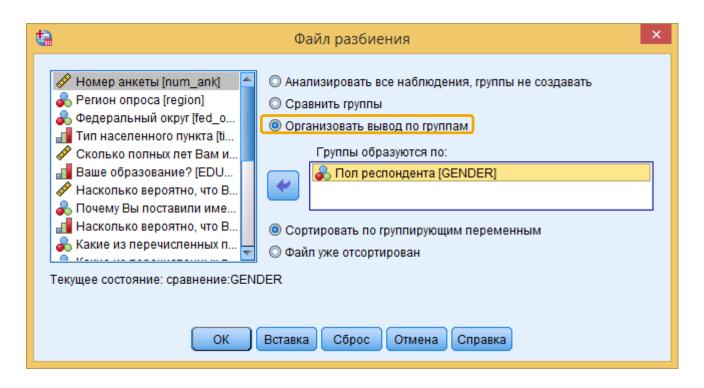
Можно выполнить анализ данных отдельно по группам (наблюдения с одинаковым значением признака). Для этого необходимо отсортировать по группирующим переменным.

Пример: Необходимо посмотреть распределение по образованию по мужчинам и женщинам.

- Отсортировать по возрастанию по полу: «Данные» → «Сортировать наблюдения». Выбрать переменную GENDER (пол). Нажать ОК.
- 2. «Данные» \rightarrow «Файл разбиения» (Split file).



- «Организовать вывод по группам». Выбрать переменную GENDER (пол).
 Нажать ОК.
- **4. «Анализ»** → **«Описательные статистики»** → **«Частоты»**. Выбрать переменную EDUCAT (образование). Нажать ОК.



Выводятся частотные данные по уровню образования в отдельных таблицах для мужчин и для женщин.

Пол респондента = Мужской

Статистика^а

Ваше образование?

N	Допустимо	234
	Пропущенные	0

а. Пол респондента = Мужской

Ваше образование?а

		Частота	Проценты	Процент допустимых	Накопленны й процент
Допустимо	Неполное среднее образование или ниже	1	,4	,4	,4
	Среднее образование (школа или ПТУ)	42	17,9	17,9	18,4
	Среднее специальное образование (техникум)	122	52,1	52,1	70,5
	Незаконченное высшее (с 3-го курса ВУЗа)	15	6,4	6,4	76,9
	Высшее образование	54	23,1	23,1	100,0
	Bcero	234	100,0	100,0	

а. Пол респондента = Мужской

Пол респондента = Женский

Статистика^а

Ваше образование?

Ν	Допустимо	266
	Пропущенные	0

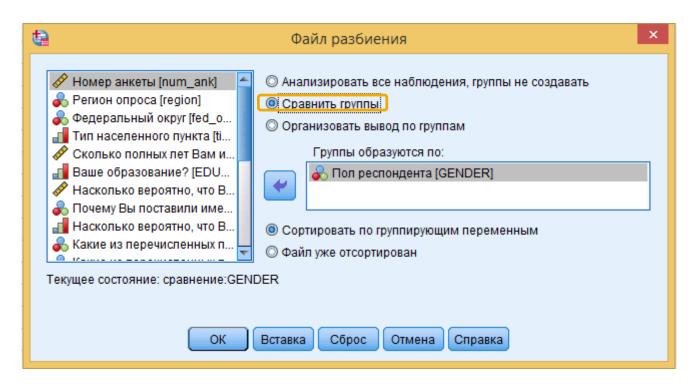
а. Пол респондента = Женский

Ваше образование?а

		Частота	Проценты	Процент допустимых	Накопленны й процент
Допустимо	Неполное среднее образование или ниже	2	,8	,8	.8
	Среднее образование (школа или ПТУ)	34	12,8	12,8	13,5
	Среднее специальное образование (техникум)	124	46,6	46,6	60,2
	Незаконченное высшее (с 3-го курса ВУЗа)	10	3,8	3,8	63,9
	Высшее образование	96	36,1	36,1	100,0
	Bcero	266	100,0	100,0	

а. Пол респондента = Женский

- Можно сделать так, чтобы на выходе информация по мужчинам и женщинам была собрана в одну таблицу. Для этого необходимо вместо «Организовывать вывод по группам» выбрать «Сравнить группы».
- Чтобы убрать разбиение на группы, следует выбрать «Анализировать все наблюдения, группы не создавать».



Выводятся частотные данные по уровню образования в общей таблице для мужчин и для женщин.

Статистика

Ваше образование?

Мужской	Ν	Допустимо	234
		Пропущенные	0
Женский	Ν	Допустимо	266
		Пропущенные	0

Ваше образование?

Пол респо	ндента		Частота	Проценты	Процент допустимых	Накопленны й процент
Мужской	Допустимо	Неполное среднее образование или ниже	1	,4	,4	,4
		Среднее образование (школа или ПТУ)	42	17,9	17,9	18,4
		Среднее специальное образование (техникум)	122	52,1	52,1	70,5
		Незаконченное высшее (с 3-го курса ВУЗа)	15	6,4	6,4	76,9
		Высшее образование	54	23,1	23,1	100,0
		Bcero	234	100,0	100,0	
Женский	Допустимо	Неполное среднее образование или ниже	2	,8	,8	,8,
		Среднее образование (школа или ПТУ)	34	12,8	12,8	13,5
		Среднее специальное образование (техникум)	124	46,6	46,6	60,2
		Незаконченное высшее (с 3-го курса ВУЗа)	10	3,8	3,8	63,9
		Высшее образование	96	36,1	36,1	100,0
		Bcero	266	100,0	100,0	



4. Объединение данных

Объединение данных необходимо в том случае, когда файлы данных создаются разными людьми (например, компаниями, проводящими опрос по одной анкете в разных регионах).

Рекомендации к процедуре объединения данных:

- При добавлении переменных и объектов, убедитесь, что порядок следования наблюдений в рабочем и внешнем файлах одинаков.
- Настройте форматы каждой переменной рабочего файла данных, чтобы они соответствовали данным внешнего файла (если добавляемая переменная содержит буквенные символы, либо замените буквы числами во внешнем файле, либо поменяйте тип переменной на строковую в рабочем файле).
- Перед добавлением данных создайте резервную копию рабочего файла, чтобы к ней можно было вернуться в случае неудачного переноса данных.

4. Объединение данных

Объединить данные из двух файлов можно двумя различными способами:

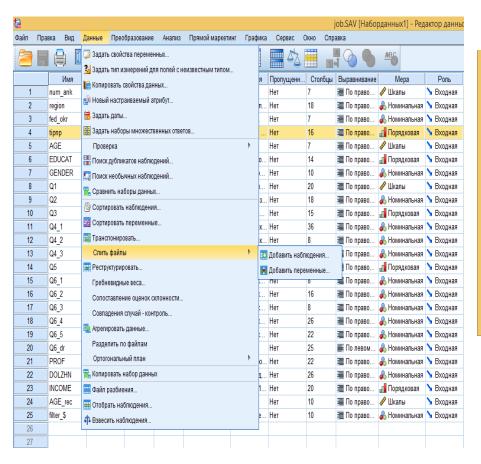
- Слить набор данных с другим набором данных, содержащим такие же переменные, но другие наблюдения.
- Слить набор данных с другим набором данных, содержащим такие же наблюдения, но другие переменные.

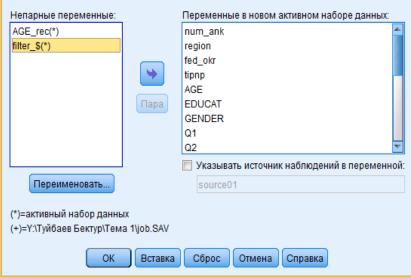
Чтобы слить данные необходимо:

- Выберите в меню «Данные» → «Слить файлы».
- Выберите в зависимости от того, что вам необходимо сделать **«Добавить** наблюдения…» или **«Добавить переменные**…».

4. Объединение данных

Добавление наблюдений





Непарные переменные – переменные, которые не войдут в новый набор данных.



5. Взвешивание данных

При взвешивании (Weighting) каждому наблюдению в базе присваивается **весовой коэффициент**, отображающий значимость наблюдения по сравнению с другими наблюдениями (респондентами).

Цель взвешивания - увеличить либо уменьшить в выборке количество наблюдений с определенными характеристиками.

Для чего применяется?

- Чтобы выборочные данные максимально точно представляли конкретные характеристики генеральной совокупности.
- Чтобы повысить значимость ответов респондентов с определенными признаками (корректировка выборки).

Пример: если проводится опрос для определения, какие изменения стоит вносить в существующую продукцию, исследователь может принять решение присвоить больший весовой коэффициент ответам респондентов, которые пользуются данным товаром чаще других.

5. Взвешивание данных

Математическая процедура: деление процента генеральной совокупности на соответствующий процент выборки.

Когда массив «взвешен», любые последующие статистические процедуры выполняются с учетом полученных коэффициентов.

	Выборка	Население в целом
Мужчины	46%	50%
Женщины	54%	50%



Weight = Population% / Sample%

Weight = 50 / 46 = 1.087 (муж)

50 / 54 = 0.926 (жен)



Меню SPSS:

Data → Weight Cases → Weight cases by...



SPSS syntax:

- compute weight1 = 0.
- if (q52= 1) weight1 = 1.087.
- if (q52= 2) weight1 = 0.926.
- exe.
- weight by weight1.

Литература по Теме 3

- 1) Бююль А., Цеффель П. SPSS: искусство обработки информации. М., 2005
 - Глава 7. Отбор данных
 - Глава 8. Модификация данных
- 2) Наследов A. IBM SPSS Statistics 20 и AMOS: профессиональный статистический анализ данных. СПб., 2013
 - Глава 4. Управление данными



