

IX. Статистические характеристики

- [Описательная статистика](#)
- [Сводка наблюдений](#)

Статистические характеристики вычисляются в основном для переменных, относящихся к интервальной шкале. Для этого используются следующие четыре команды меню.

Analyze (Анализ) Descriptive Statistics (Дескриптивные статистики) Descriptives..., (Описательная статистика). Analyze (Анализ) Descriptive Statistics (Дескриптивные статистики) Frequencies... (Частоты). Analyze (Анализ) Descriptive Statistics (Дескриптивные статистики) Explore... (Исследовать) Analyze (Анализ) Reports (Отчеты) Case summaries... (Итоги по наблюдениям)

Создание частотных таблиц рассматривается в главе 6, а исследование данных — в главе 10.

В нижеследующей таблице приведен обзор характеристик, рассчитываемых в SPSS. В меню Descriptives... можно также провести стандартизацию переменных (z-преобразование).

Характеристика	Descriptives	Frequencies	Explore	Case summaries
Среднее значение	X	X	X	X
Сумма	X	X		X
Медиана		X	X	X
Групповая медиана		X		X
Квартиль		X		
Процентиль		X	X	
Мода		X		
Стандартное отклонение	X	X	X	X
Стандартная ошибка	X	X	X	X
Дисперсия	X	X	X	X
Минимум	X	X	X	X
Максимум	X	X	X	X
Размах	X	X	X	X
Межквартильная широта			X	
Эксцесс (вариация)	X	X	X	X
Асимметрия	X	X	X	X
Стандартная ошибка эксцесса	X	X	X	X
Стандартная ошибка асимметрии	X	X	X	X
Доверительный интервал			X	
Гармоническое среднее				X
Геометрическое среднее				X
M-оценка (Хампеля)			X	
Выброс			X	
Усеченное среднее			X	

Статистические характеристики, которые задаются в меню Case summaries, можно также вычислить отдельно по категориям группирующих переменных, относящихся к номинальной или порядковой шкале.

В качестве примера для этой и следующей главы мы рассмотрим исследование, относящееся к области медицины — анализ действия двух различных лекарств (с вымышленными названиями альфасан и бетасан) на снижение кровяного давления у гипертоников. Эти данные хранятся в файле `hureg.sav`, содержащем 174 наблюдения и значения следующих переменных:

nr	Номер пациента
med	Лекарство (1 = альфасан, 2 = бетасан)
g	Пол (1 = мужской, 2 = женский)
a	Возраст, лет
gr	Рост, см
gew	Вес, кг
rrs0	Систолическое кровяное давление, исходное значение
rrs1	то же, через 1 месяц
rrs6	то же, через 6 месяцев
rrs12	то же, через 12 месяцев
rrd0	Диастолическое кровяное давление, исходное значение
rrd1	то же, через 1 месяц
rrd6	то же, через 6 месяцев
rrd12	то же, через 12 месяцев
chol0	Холестерин, исходное значение
chol1	то же, через 1 месяц
chol6	то же, через 6 месяцев
chol12	то же, через 12 месяцев
bz0	Сахар в крови, исходное значение
bz1	то же, через 1 месяц
bz6	то же, через 6 месяцев
bz12	то же, через 12 месяцев
ak	Возрастной класс (1 = до 55 лет, 2 = 56-65 лет, 3 = 66-75 лет, 4 = более 75)

9.1. Описательная статистика

Для ознакомления с характеристиками описательной статистики рассмотрим переменную *a*, отражающую возраст.

- Загрузите файл *hyper.sav* и выберите команды меню Analyze (Анализ) Descriptive Statistics (Дескриптивные статистики) Descriptives... (Описательная статистика) Откроется диалоговое окно Descriptives.

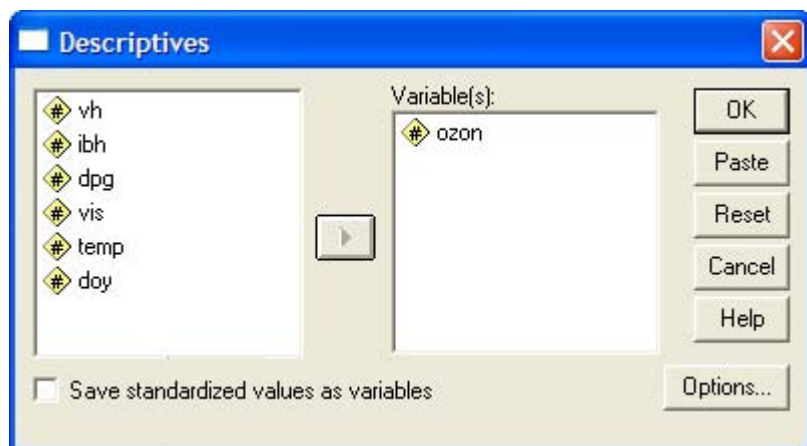


Рис. 9.1: Диалоговое окно *Descriptives*

- Перенесите переменную *a* в список тестируемых переменных, и щелкните на кнопке *Options...* (Параметры).

Здесь можно задать вычисление следующих статистических характеристик:

- Среднего значения,
- Суммы,
- Стандартного отклонения,
- Стандартной ошибки,
- Дисперсии,
- Минимума,
- Максимума,
- Размаха,
- Эксцесса (вариации),
- Асимметрии.
- Установите флажки для вывода следующих характеристик: Mean (Среднее значение), Minimum (Минимум), Maximum (Максимум) и S.E. mean (Стандартная ошибка).

Если анализируется несколько переменных, можно также задать последовательность вывода:

- в порядке возрастания средних значений,
- в порядке убывания средних значений,
- по алфавиту (по именам переменных),
- согласно списку выбранных целевых переменных.

По умолчанию выбран последний вариант. Если имеется только одна переменная, как в данном примере, порядок не имеет значения.

- Пометив желаемые характеристики, щелкните на кнопке *Continue...* (Далее). В главном диалоговом окне укажите, чтобы стандартизованные значения были сохранены в новой переменной открытого файла данных, для чего установите флажок *Save standardized values as variables*.

- Запустите вычисление, щелкнув на кнопке ОК. Результат будет показан в окне просмотра:

Descriptive Statistics (Описательная статистика)

	N	Minimum	Maximum	Mean	
Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic
Возраст	174	36	87	62,11	,88
Valid N (listwise) (Допустимых значений (по списку))	174				

О значении отдельных характеристик описательной статистики можно прочесть в главе 6.

Видно, что в файле данных появилась новая переменная *z*. Она содержит нормированные значения переменной *a* (Возраст). По умолчанию к имени исходной переменной спереди дописывается буква *z*. При этом стандартизация (*z*-преобразование) значения *x* выполняется по формуле

$$z = \frac{x - m}{s}$$

Здесь *m* — среднее значение переменной, а *s* — стандартное отклонение.

Проведение стандартизации переменных может быть целесообразным при использовании некоторых статистических методов. Его также можно выполнять в тех случаях, когда несколько переменных, которые имеют различный размах или отличаются на порядки по значению, должны быть приведены к общему показателю. В подобной ситуации сначала необходимо провести стандартизацию этих переменных, а затем, путем усреднения, вывести общее значение из полученных стандартизованных значений (*z*-значений).

9.2. Сводка наблюдений

Этот пункт меню позволяет как выводить значения переменных по наблюдениям, так и вычислять статистические характеристики.

Первую из этих возможностей мы рассмотрели в разделе 4.7; сейчас мы опишем вычисление статистических характеристик. В качестве примера снова выберем файл *hyper.sav*.

- Загрузите файл *hyper.sav* и выберите команды меню Analyze (Анализ) Reports (Отчеты) Case summaries... (Сводка наблюдений)

Откроется диалоговое окно Summarize Cases (Вывести сводку наблюдений) (см. рис. 9.2).

- Перенесите переменную *a* в правый список и снимите флажок Display Cases (Показывать наблюдения).
- Щелкните на кнопке Statistics... (Статистика). Откроется диалоговое окно Summary Report: Statistics (Сводка: Статистика) (см. рис. 9.3).
- Выберите в списке вычисление среднего значения (Mean), медианы (Median), гармонического среднего (Harmonic Mean) и геометрического среднего (Geometric Mean).
- Кнопка Options... позволяет задать заголовок для сводной таблицы и способ обработки пропущенных значений.

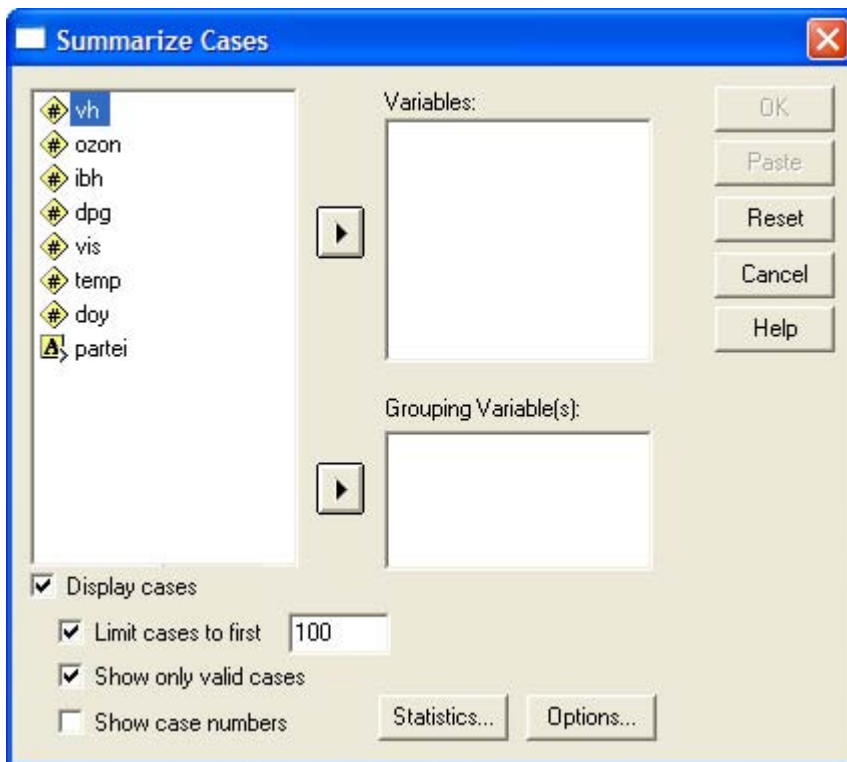


Рис. 9.2: Диалоговое окно Summarize Cases

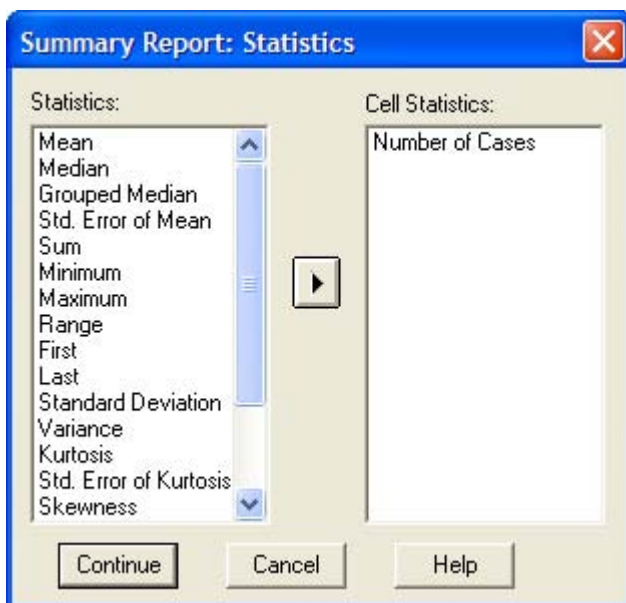


Рис. 9.3: Диалоговое окно Summary Report: Statistics

В окне просмотра будут показаны следующие результаты:

Case Processing Summary (Обработанные наблюдения)

	Cases (Случаи)					
	Included (Включенные)		Excluded (Исключенные)		Total (Всего)	
N	Percent	N	Percent	N	Percent	
Возраст	174	100,0%	0	,0%	174	100,0%

Case Summaries (Сводка наблюдений)

Возраст

Mean	Median	Harmonic Mean	Geometric	Mean
62,11	63,00	59,80	60,98	

Описательные характеристики можно также вычислить отдельно по категориям группирующей переменной.

- Выберите в качестве тестируемой переменной chol0, а в качестве группирующей переменной — g.
- Задайте вычисление среднего значения, стандартного отклонения, стандартной ошибки среднего (Std. Error of Mean) и медианы.

В окне просмотра будут показаны следующие результаты:

Case Processing Summary

	Cases					
	Included		Excluded		Total	
N	Percent	N	Percent	N	Percent	
Холестерин, исходный * Пол	174	100,0%	0	,0%	174	100,0%

Case Summaries

Холестерин, исходный				
Пол	Mean	Std. Deviation	Std. Error Mean	Median
мужской	228,95	54,63	7,11	216,00
женский	241,54	46,19	4,31	241,00
Total	237,27	49,42	3,75	234,50

О настройках, предназначенных для вывода значений по наблюдениям см. раздел 4.8.

Раздельное вычисление по категориям группирующей переменной можно также выполнить при помощи команд меню Analyze (Анализ) Compare Means (Сравнение средних) Means... (Средние). Analyze (Анализ) Reports (Отчеты) OLAP Cubes... (OLAP-кубы)

Здесь доступны те же характеристики, что и в меню Case summaries...

Метод вычисления в форме OLAP-кубов (Online Analytical Processing) впервые появился в версии 9 SPSS. Он отличается тем, что таблицы, получающиеся при разбиении по группирующим переменным, можно активировать, пользуясь мобильными таблицами.