

ОСНОВЫ СТАТИСТИКИ

Модули SPSS. Вывод статистических
характеристик





Здравствуйте!

В ходе изучения материалы, вы освоите ключевые принципы работы в SPSS, в частности вывод статистических данных и графическое представление данных.

Мы говорим, что, обработав данные мы можем получить ответы на многие, интересующие нас вопросы, но каким образом это сделать? Для этого необходимо, во-первых, знать основу статистики, т.е. базовые методы и инструменты, о которых мы будем с вами говорить, ну и конечно же, без чего невозможна современная обработка данных, это программный продукт, автоматизирующий все расчеты. Их на рынке сейчас немало, но одним из наиболее часто используемых в практике программ является статистический пакет SPSS for Windows. SPSS Statistics (аббревиатура англ. «Statistical Package for the Social Sciences» – «статистический пакет для общественных наук») – компьютерная программа для статистической обработки данных, один из лидеров рынка в области коммерческих статистических продуктов, предназначенных для проведения прикладных исследований в общественных науках. Он выделяется своим хорошим интерфейсом и функциональными возможностями, в основном работают в англоязычной версии, так как при работе с русскоязычной версией могут возникнуть проблемы с пониманием назначения той или иной функции, поэтому при изложении материала вы будете видеть окно с англоязычным интерфейсом, а озвучиваться будут перевод назначения того или иного окна.

Основу программы SPSS составляет SPSS Base (базовый модуль), предоставляющий разнообразные возможности доступа к данным и управления данными. Он содержит методы анализа, которые применяются чаще всего.

Традиционно вместе с SPSS Base (базовым модулем) поставляются ещё два модуля: Advanced Models (продвинутые модели) и Regression Models (регрессионные модели). Также существуют еще дополнительные модели.

SPSS Base (Базовый модуль) входит в базовую поставку. Он включает все процедуры ввода, отбора и корректировки данных, а также большинство предлагаемых в SPSS статистических методов. Наряду с простыми методиками статистического анализа, такими как частотный анализ, расчет статистических характеристик, таблиц сопряженности, корреляций, построения графиков, этот модуль включает t-тесты и большое количество других непараметрических тестов, а также усложненные методы, такие как многомерный линейный регрессионный анализ, дискриминантный анализ, факторный анализ, кластерный анализ, дисперсионный анализ, анализ пригодности (анализ надежности) и многомерное шкалирование.

Regression Models. Данный модуль включает в себя различные методы регрессионного анализа, такие как: бинарная и мультиномиальная логистическая регрессия, нелинейная регрессия и пробит-анализ.

Advanced Models В этот модуль входят различные методы дисперсионного анализа (многомерный, с учетом повторных измерений), общая линейная модель, анализ выживания и другие.



Вывод статистических характеристик

Чтобы получить описательную статистику числовых переменных, можно кликнуть в диалоге Frequencies на кнопку Statistics... (Статистика). Откроется диалоговое окно Frequencies: Statistics (Частоты: Статистика).

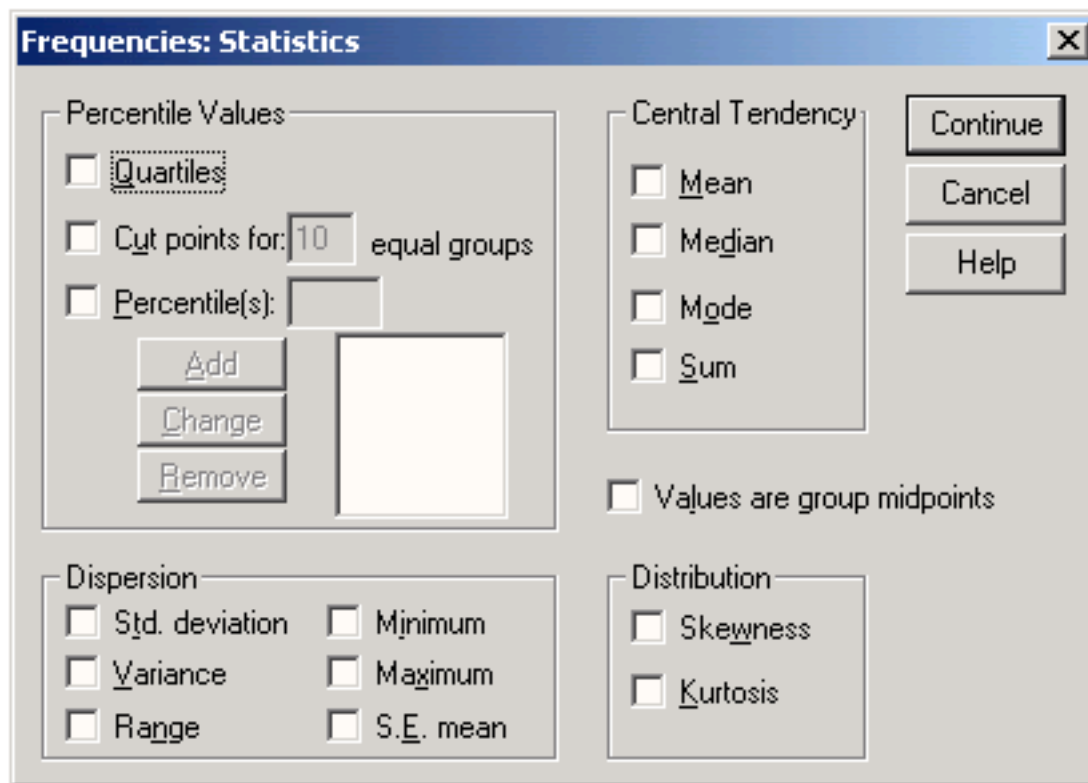


Рисунок 13.1 Диалоговое окно frequencies: Statistics

В группе Percentile Values (Значения процентилей) можно выбрать следующие варианты: Quartiles (Квартили): Будут показаны первый, второй и третий квартили. Первый квартиль (Q1) – это точка на шкале измеренных значений, ниже (левее) которой располагаются 25% измеренных значений. Второй квартиль (Q2) – это точка, ниже которой располагаются 50% измеренных значений. Второй квартиль также называется медианой. Третий квартиль (Q3) – это точка на шкале измеренных значений, ниже которой располагаются 75% значений. Если данные имеются только в форме порядкового отношения, то качестве меры разброса используется межквартильная ширина. Она определяется как:

Cut points (Точки раздела): Будут вычислены значения процентилей, разделяющие выборку на группы наблюдений, которые имеют одинаковую ширину, то есть включают одно и то же количество измеренных значений. По умолчанию предлагается количество групп 10. Если задать, к примеру, 4, то будут показаны квартили, то есть квартили соответствуют процентилем 25, 50 и 75. Видно, что число показываемых процентилей на единицу меньше заданного числа групп. Percentile(s) (Процентили): Здесь имеются в виду значения процентилей, определяемые пользователем. Введите значение процентиля в пределах от 0 до 100 и кликните на кнопку Add (Добавить). Повторите эти действия для всех желаемых значений процентилей. Значения в порядке возрастания будут показаны в списке. Например, если ввести значения 25, 50 и 75, то мы получим квартили. Можно задавать любые значения процентилей, например, 37 и 83. В первом случае (37) будет показано значение выбранной переменной, ниже которого лежат 37%



значений, а во втором случае (83) – значение, ниже которого располагаются 83% значений.

В группе Dispersion (Разброс) можно выбрать следующие меры разброса:

Std. deviation (Стандартное отклонение) – это мера разброса измеренных величин; оно равно квадратному корню из дисперсии. В интервале шириной, равной удвоенному стандартному отклонению, который отложен по обе стороны от среднего значения, располагается примерно 67% всех значений выборки, подчиняющейся нормальному распределению.

Variance (Дисперсия) – это квадрат стандартного отклонения и, следовательно, эта характеристика также является мерой разброса измеренных величин. Она определяется как сумма квадратов отклонений всех измеренных значений от их среднеарифметического значения, деленная на количество измерений минус 1.

Range (Размах) – это разница между наибольшим значением (максимумом) и наименьшим значением (минимумом).

Minimum (Минимум) – Наименьшее значение.

Maximum (Максимум) – Наибольшее значение.

S.E. mean (Стандартная ошибка среднего значения) – в интервале шириной, равной удвоенной стандартной ошибке, отложенному вокруг среднего значения, располагается среднее значение генеральной совокупности с вероятностью примерно 67%. Стандартная ошибка определяется как стандартное отклонение, деленное на квадратный корень из объема выборки.

Обычно мерами разброса переменных, относящихся к интервальной шкале и подчиняющихся нормальному распределению, служат стандартное отклонение и стандартная ошибка. Как было сказано выше, стандартное отклонение позволяет задать диапазон разброса отдельных значений. По так называемому правилу кулака, в одном диапазоне стандартного отклонения (охватывающем ширину стандартного отклонения в обе стороны от среднего значения) располагается примерно 67% значений, в диапазоне удвоенного стандартного отклонения – примерно 95%, а в диапазоне утроенного стандартного отклонения – примерно 99% значений.

С другой стороны, стандартная ошибка позволяет задать доверительный интервал для среднего значения. В диапазоне удвоенной стандартной ошибки по обе стороны от среднего значения с вероятностью примерно 95% находится среднее значение генеральной совокупности. С вероятностью примерно 99% она лежит в диапазоне утроенной стандартной ошибки. Часто указывают только одну из этих двух мер разброса, обычно – стандартную ошибку, так как ее значение меньше. Во всех случаях следует точно выяснить, какая из мер разброса имеется в виду.

В группе Central Tendency (Средние) можно выбрать следующие характеристики:

Mean (Среднее значение) – это арифметическое среднее измеренных значений; оно определяется как сумма значений, деленная на их количество. Например, если имеется 12 измеренных значений и их сумма составляет 600, то среднее значение будет $x = 600 : 12 = 50$.

Median (Медиана) – это точка на шкале измеренных значений, выше и ниже которой лежит по половине всех измеренных значений.

Mode (Мода) – это значение, которое наиболее часто встречается в выборке. Если одна и та же наибольшая частота встречается у нескольких значений, то выбирается наименьшее из них.

Sum (Сумма) – сумма всех значений.

В группе Distribution (Распределение) можно выбрать следующие меры несимметричности распределения:

Skewness (Коэффициент асимметрии) – это мера отклонения распределения частоты от симметричного распределения, то есть такого, у которого на одинаковом удалении от среднего



значения по обе стороны выборки данных располагается одинаковое количество значений. Если наблюдения подчиняются нормальному распределению, то асимметрия равна нулю.

Kurtosis (Коэффициент вариации или эксцесс) – указывает, является ли распределение пологим (при большом значении коэффициента) или крутым. Коэффициент вариации равен нулю, если наблюдения подчиняются нормальному распределению. Как правило, для переменных, относящихся к интервальной шкале и подчиняющихся нормальному распределению, в качестве основной характеристики используют среднее значение, а в качестве меры разброса – стандартное отклонение или стандартную ошибку. Для порядковых или интервальных переменных, не подчиняющихся нормальному распределению – соответственно медиану или первый и третий квартили. Для переменных относящихся к номинальной шкале, нельзя дать других значимых характеристик кроме моды.

В диалоге есть еще один флажок:

Values are group midpoints (Значения являются средними точками групп): Если установить этот флажок, то при вычислении медианы и остальных значений процентилей оценки этих характеристик будут определяться для концентрированных данных.

Для переменной alter (возраст) мы определим следующие характеристики: среднее значение, медиану, моду, квартили, стандартное отклонение, дисперсию, размах, минимум, максимум, стандартную ошибку, асимметрию и эксцесс. Поступите следующим образом:

Выберите в меню команды Analyze (Анализ) / Descriptive Statistics (Дескриптивные статистики) / Frequencies... (Частоты).

В диалоге Frequencies кликните на кнопку Reset (Сброс), чтобы отменить прежние настройки.

Перенесите переменную alter в список выходных переменных.

Кликните на кнопку Statistics... (Статистика).

В диалоге Frequencies: Statistics установите флажки желаемых характеристик. Затем кликните на кнопку Continue (Продолжить). Вы вернетесь в диалог Frequencies.

В диалоге Frequencies деактивируйте опцию Display frequency tables (Показывать частотные таблицы). Кликните на кнопку ОК.

В окне просмотра появятся следующие результаты:

Таблица 13.1 Statistics (Статистика) Alter

N	Valid (Допустимые)	106
	Missing (Утерянные)	2
Mean (Среднее значение)		22,24
Std. Error of Mean (Стандартная ошибка среднего)		21
Median (Медиана)		22,00
Mode (Мода)		21



Std. Deviation (Стандартное отклонение)		2,189
Variance (Дисперсия)		4,791
Skewness (Асимметрия)		0,859
Std. Error of Skewness (Стандартная ошибка асимметрии)		0,235
Kurtosis (Коэффициент вариации / Экцесс)		1,042
Std. Error of Kurtosis (Стандартная ошибка эксцесса)		0,465
Range (Размах)		11
Minimum (Минимум)		18
Maximum (Максимум)		29
Percentiles (Процентили)	25	21,00
	50	22,00
	75	23,00

Респонденты опроса о финансовом состоянии и социальном положении имеют средний возраст 22,24 года. Медиана составляет 22. Большинству респондентов 21 год (это мода). Самому молодому респонденту 18 лет (минимум), самому старшему – 29 лет (максимум). Самый старший респондент на 11 лет старше самого молодого (размах). Стандартное отклонение составляет 2,19. Следовательно, дисперсия – квадрат стандартного отклонения – равна $(2,19)^2 = 4,79$. Асимметрия и коэффициент вариации даны со соответствующими стандартными ошибками.

Графическое представление

Результаты частотного распределения можно представить графически. Для примера мы создадим столбчатую диаграмму для частотного распределения основных специальностей. Поступите следующим образом:

Выберите в меню команды Analyze (Анализ) / Descriptive Statistics (Дескриптивные статистики) / Frequencies (Частоты)

Перенесите переменную fach в список выходных переменных.

Кликните на кнопку Charts... (Диаграммы). Откроется диалоговое окно Frequencies: Charts (Частоты: Диаграммы).

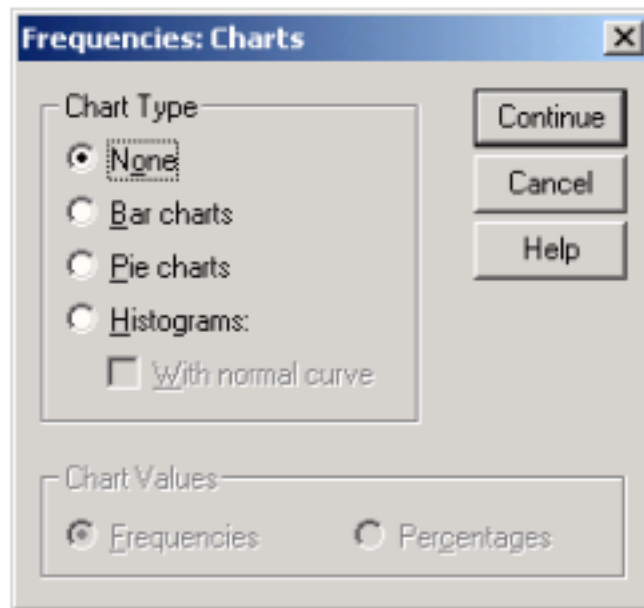


Рисунок 13.2 Диалоговое окно Frequencies: Charts.

Выберите в группе Chart Type (Тип диаграммы) пункт Bar charts (Столбчатая диаграмма), а в группе Chart Values (Значения диаграммы) – пункт Percentages (Проценты). Подтвердите выбор кнопкой Continue (Продолжить). Вы вернетесь в диалог Frequencies.

В диалоговом окне Frequencies снимите флажок Display frequency tables (Показывать частотные таблицы). – Кликните на кнопку ОК. Диаграмма будет показана в окне просмотра (см. рис. 13.3).





Рисунок 13.3 Столбчатая диаграмма в средстве просмотра

Усовершенствуем вид этой диаграммы.

Чтобы начать редактирование, дважды кликните в области столбчатой диаграммы. Диаграмма будет показана в редакторе диаграмм.

На панели инструментов редактора диаграмм кликните на символе меток столбцов:



Откроется диалоговое окно Bar Label Style (Стиль меток столбцов). Выберите пункт Framed (В рамке), кликните на кнопку Apply all (Применить для всех) и затем на Close (Закреть). На каждом столбце появится надпись с его процентным значением.

Кликните мышью на любом из столбцов. На верхней стороне каждого столбца появится по два маленьких черных квадрата. Это означает, что области столбцов готовы для редактирования.

Кликните мышью на символе образца заливки:



Откроется диалоговое окно Fill Patterns (Образцы заливки).

Выберите в нем подходящий образец заливки. Подтвердите выбор кнопкой Apply (Применить) и закройте диалоговое окно.

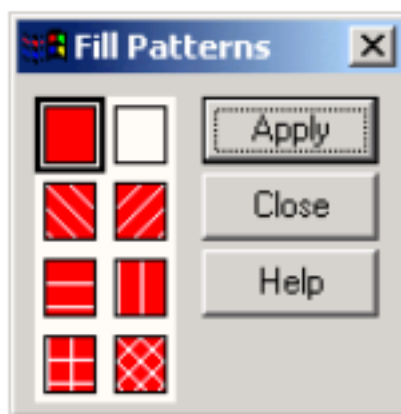


Рисунок 13.4 Диалоговое окно Fill Patterns.

Столбцы будут заполнены выбранной заливкой.

Кликните мышью на символе вида столбцов:

Выберите пункт Drop shadow (Тень), кликните на кнопку Apply all (Применить для всех) и затем на Close (Закреть).

Дважды кликните на заголовке диаграммы Fachbereich. Откроется диалоговое окно Titles (Заголовки) (см. рис. 13.5).

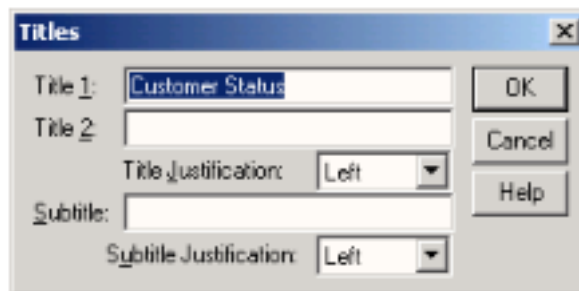


Рисунок 13.5 Диалоговое окно Titles

Измените заголовок на «Основная специальность» и закройте диалог кнопкой ОК.



В меню Chart (Диаграмма) установите флажок Outer Frame (Внешняя рамка). Закройте редактор диаграмм; получившийся график показан на рис. 13.6.

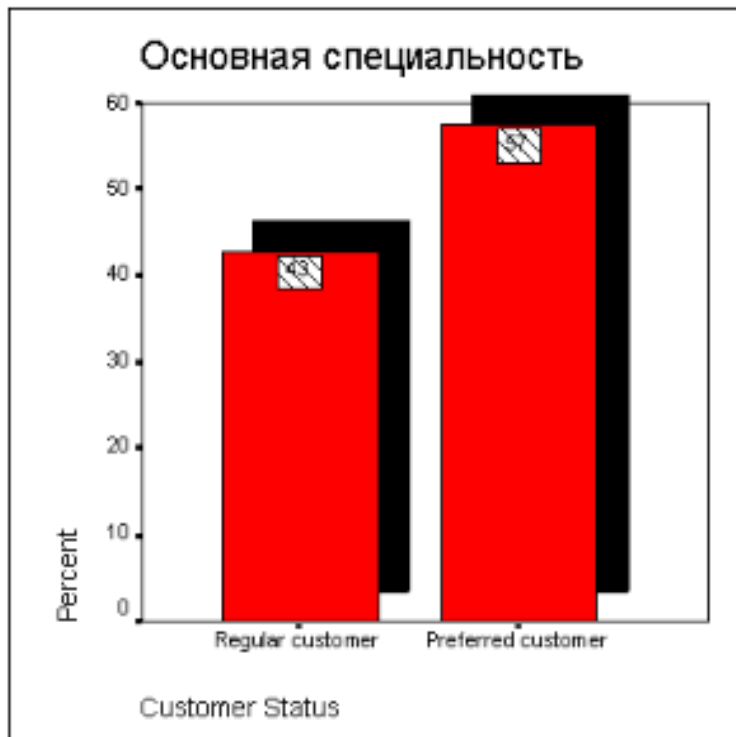


Рисунок 13.6 Отредактированная диаграмма

Рассмотрим другой пример – визуальное представление частотного анализа.

Выберите в меню команды Analyze (Анализ) / Descriptive Statistics (Дескриптивные статистики) / Frequencies (Частоты)

Кликните на кнопку Reset (Сброс), чтобы установить стандартные настройки.

Перенесите переменную sozial (социальное положение) в список выходных переменных.

Кликните на кнопку Charts... (Диаграммы). В диалоговом окне Frequencies: Charts выберите пункт Histograms (Гистограмма). Установите флажок With normal curve (С кривой нормального распределения). Кликните на кнопку Continue.

В диалоговом окне Frequencies снимите флажок Display frequency tables (Показывать частотные таблицы). Кликните на кнопку ОК. Гистограмма будет показана в окне просмотра (см. рис. 13.7).

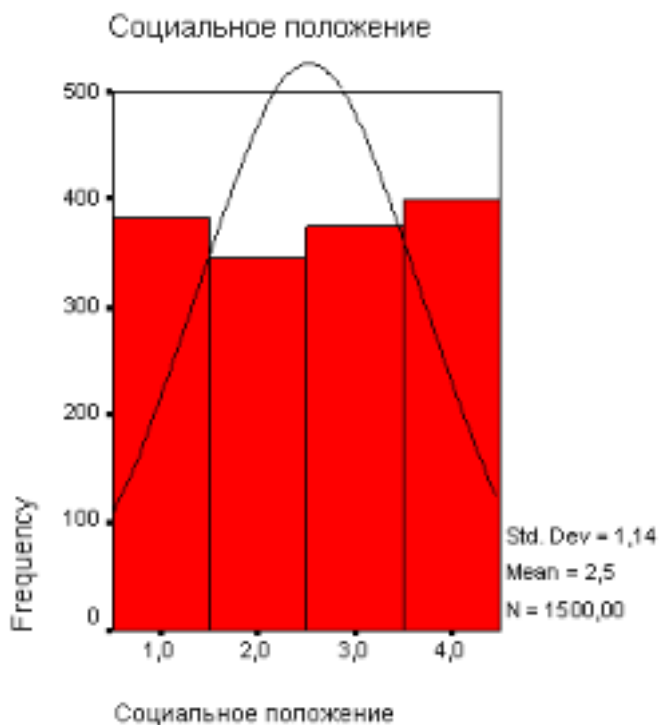


Рисунок 13.7 Гистограмма

Частоты на гистограмме обозначены колонками, которые, в отличие от столбчатой диаграммы, не изолированы, а примыкают друг к другу. Отображаются также стандартное отклонение, среднее значение и общее количество наблюдений (M). Кроме того, показана кривая нормального распределения.

Дважды кликните на области гистограммы – откроется редактор диаграмм, в котором можно придать гистограмме желаемый вид. График отобразится в редакторе диаграмм.

Выберите другой образец заливки и снабдите колонки надписями.

При желании проверьте другие функции редактора диаграмм.

С помощью гистограмм чаще всего отображают переменные, относящихся к интервальной шкале.

Отметим еще ряд возможностей графического представления данных в SPSS в статистике широко используются следующие графические изображения: диаграммы и статистические карты (картограммы).

Диаграммы – наиболее распространенный способ графических изображений. Это графики количественных отношений. Виды и способы их построения разнообразны. Диаграммы используются для наглядного сравнения с прошлым периодом, с другим объектом, с нормативом, с плановым заданием.

Статистические карты (картограммы) – графики количественного распределения по конкретной территории. По своей основной характеристике этот вид графиков близок к диаграммам и специфичен лишь в том отношении, что представляют собой условные изображения статистических данных на контурной географической карте.

Картограммы применяются в пространственной экономике. В бизнес-статистике они важны при решении логистических задач: выявлении сети поставщиков и потребителей, оптимизации



поставок и т.п. Густота тона или штриховки соответствует количественному значению показателя для той или иной территории.

В зависимости от характера решаемых задач выделяют диаграммы сравнения, структурные диаграммы и диаграммы динамики; диаграммы распределения – гистограмма, полигон, кумулята, огиба.

Наиболее распространенным видом диаграмм сравнения являются столбиковые диаграммы (bar chart) – графическое изображение статистических показателей в виде столбиков-прямоугольников. Такие диаграммы широко используются для наглядного сравнения во времени и пространстве, а также для изображения структуры объекта. Столбики могут иметь горизонтальное расположение (tape chart).

При построении столбиковых диаграмм необходимо начертить систему прямоугольных координат, в которой расположатся столбики. Столбиковые диаграммы можно применять для сравнения нескольких показателей.

Широко применяются структурные диаграммы (structure chart). Основное назначение структурных диаграмм заключается в наглядном отражении структуры явления, характеристике удельных весов отдельных частей изучаемого явления, выявлении структурных сдвигов.

Подведем итоги. Программа SPSS позволяет получать данные описательной статистики, в несколько кликов мыши, кроме того возможно использование также статистических функций, который включает в себя классический набор. Также возможно графическое отображение данных, без особых трудностей.