

# СТАТИСТИЧЕСКИЙ АНАЛИЗ НАБЛЮДЕНИЙ ИНСТРУМЕНТАЛЬНЫМИ СРЕДСТВАМИ ПРОГРАММНОГО КОМПЛЕКСА SPSS

*Спирков С. Н., кандидат  
технических наук, профессор,  
декан учетно-финансового  
факультета  
Бусыгин Д. Ю., аспирант*

**Резюме.** Анализируются особенности универсального статистического пакета SPSS. Предлагается процедура статистической обработки данных, включающая в себя их проверку на соответствие закону нормального распределения, выявление и оценку связи между показателями, непосредственную обработку данных. Определены возможности и применение пакета SPSS как интеграционного программного продукта.

**Summary.** The peculiarities of universal statistical packet SPSS have been analyzed. The procedure of statistical data processing has been suggested, including their checking on compliance with law of normal distribution as well as revealing and assessment of connection between indices, immediate data processing. Determination of the capability and use of SPSS packet as integration software product.

Процесс автоматизированной обработки статистической информации представляет собой совокупность операций, выполняемых в определенной последовательности – от сбора исходной информации до получения результатной информации. Под операцией, в данном случае, понимается комплекс действий, выполняемых над информацией на одном рабочем месте.

Важная роль в этом процессе отведена информационным технологиям, ориентированным на получение, обработку и распространение информации. Разработка технологического процесса обеспечивает его максимальную автоматизацию при использовании различных технологических средств и высокую достоверность получения результатной информации при минимальных трудовых и стоимостных затратах. При определении состава операций процесса обработки создается возможность выбора наиболее рационального способа обработки информации. В процессе автоматизированной обработки статистической информации прослеживаются три основных этапа: первичный, основной и заключительный.

Операции первичного этапа связаны со сбором и подготовкой статистических отчетов к машинной обработке. От качества выполнения этих операций во многом зависит эффективность и достоверность непосредственно автоматизированной обработки.

Операции основного этапа обеспечивают ввод и обработку первичных отчетов, а также печать сводных отчетов в соответствии с заданным алгоритмом решения статистической задачи. Автоматизированная обработка статистической информации может осуществляться с применением отдельных ПЭВМ или входящих в состав локальной информационно-вычислительной сети.

На заключительном этапе выполняются операции контроля и выпуска сводных отчетов с целью проверки качества их обработки и печати, дальнейшего размножения и рассылки различным потребителям.

В настоящее время, в связи с развитием информационных технологий, совершенствуются методы обработки экономической информации, анализа и синтеза экономико-математических моделей, решения прикладных экономических задач. Современные экономико-математические модели отличаются высокой сложностью, большим объемом входной информации, наличием неопределенных и трудно формализуемых факторов, разнообразием матема-

тических методов. Так, математический анализ и теория дифференциальных уравнений используются при исследовании производственной функции, моделей экономического равновесия и роста; линейная алгебра – в балансовых задачах; теория вероятностей и математическая статистика – при построении прогнозов, анализе экономической информации и моделей; математическое программирование, теория игр и теория оптимального управления – в принятии решений. С одной стороны, практика экономико-математического моделирования развивается во многом благодаря росту вычислительных возможностей ПЭВМ и функциональных возможностей программного обеспечения. С другой – потребности экономистов в математическом инструментарии учитываются при разработке новых версий программных продуктов, благодаря чему расширяются возможности построения моделей. Таким образом, наблюдается положительная обратная связь между развитием экономико-математического моделирования и программного инструментария.

При математическом моделировании и решении прикладных задач все больший интерес концентрируется вокруг пакетов прикладных задач. Связано это с тем, что кустарная реализация не только сложных, но и внешне простых вычислений может привести к большим ошибкам и неустойчивости решений. Программное обеспечение используется как для моделирования и анализа, так и для организации научной работы, сбора и хранения информации, работы с научными источниками, оформления результатов исследований и т. п. Указанное программное обеспечение можно разделить на две основные группы:

1. Универсальные пакеты, позволяющие строить математические модели и решать задачи широкого спектра.

2. Специализированные программы и библиотеки функций, ориентированные на решение определенных типов задач.

Если математические пакеты общего назначения используются для реализации математических моделей, то основной функцией пакетов статистического анализа является обработка числовой информации в интерактивном режиме. Современные пакеты содержат встроенные средства для ввода и изменения исходных данных в форме таблиц. При проведении статистических исследований не требуется дополнительное программное обеспечение – все этапы статистического анализа, начиная с ввода исходных данных и их преобразований и заканчивая подготов-

кой отчета или написанием собственных процедур обработки, можно выполнить, используя только установленный пакет. Вызов команд обработки данных производится посредством меню, панелей инструментов и удобных диалоговых окон. Как правило, пакеты статистического анализа содержат более широкий набор статистических функций, чем обычные математические программы.

Пакет SPSS является самой распространенной за рубежом программой для обработки статистической информации. Исторически пакет SPSS разрабатывался для проведения социологических исследований, что отражено в названии – Statistical Package for the Social Science. С появлением персональных компьютеров была разработана также и PC-версия SPSS, с 1983 года появилась PC-версия SPSS/PC+, рассчитанная на MS-DOS. Для того, чтобы отразить возможность использования программы во всех областях, имеющих отношение к статистическому анализу, исходной аббревиатуре присвоено новое значение: Superior Performance Software System (система программного обеспечения высшей производительности).

Надо отметить, что в справочной системе SPSS недостаточно уделяется внимания описанию положений математической статистики. Предполагается, что пользователь пакета достаточно компетентен в предметной области своих исследований. В помощь начинающему пользователю предлагается лишь два типа подсказок по математической статистике: краткая информация в контекстном меню интерактивных таблиц и серия диалоговых окон, с помощью которых определяется нужная пользователю статистическая процедура. На современном этапе отличительной чертой программы SPSS является использование ее не только как статистического пакета, но и как интеграционного пакета, который может быть использован на различных шагах процесса маркетинговых исследований. SPSS можно применять при идентификации проблемы, разработке подхода к решению, разработке проекта исследований, сборе, подготовке, анализе данных, подготовке и презентации отчета.

Овладение приемами работы с программой SPSS требует предварительных познаний в области статистики. Рассмотрим основные понятия, которые должен знать пользователь, программы SPSS. В первую очередь, сюда относятся предварительные оценки, которые выполняются перед проведением любого статистического теста: проверка наличия нормального распределения; выделение независимых и зависимых выборов; классификация

переменных по статистическим шкалам. Необходимо рассмотреть также часто проводимые процедуры проверки гипотезы о среднем значении вероятности ошибки  $p$ .

Все наблюдения содержат случайную составляющую. Поэтому результаты, получаемые при статистической обработке наблюдений, могут различаться по степени точности и обоснованности. Степень обоснованности и точности результатов зависит от количества, характера данных, метода их обработки, степени соответствия наблюдений и методов друг другу. Количество наблюдений, необходимых для статистической обработки, не может быть менее 5-8. Для получения надежных оценок необходимо не менее 30-40 наблюдений. Если наблюдения охватывают не все изучаемые объекты, а только некоторую их часть, т. е. используется выборочный метод, то минимальное число измерений должно быть установлено с требуемым уровнем точности и доверительной вероятности [4, с. 65-74; 5, с. 223-268].

Большинство статистических методов и программ в статистических пакетах для обработки наблюдений на компьютерах основываются на гипотезе о нормальном законе распределения данных, используемых для обработки. Поэтому первым этапом анализа данных является их проверка на соответствие закону нормального распределения. Классические правила статистического теста, который исходит из гипотезы о нормальном распределении, заключаются в следующем [4, с. 133-137]:

1. Выбираются данные о количественных показателях социально-экономических явлений (объем производства в стоимостном выражении, размеры прибыли, величины рентабельности и годового прироста объемов производства и т. п.);

2. Производится статистическая обработка данных, для чего составляется таблица данных и пар переменных, например, рентабельности и прироста производства в процентах;

3. Строится гистограмма для каждой переменной, показывающая, сколько значений попадает в каждый из интервалов, на которые можно разделить весь диапазон изменения рассматриваемой величины;

4. С помощью критериев согласия А.Н. Колмогорова, В.И. Романовского, К. Пирсона проверяется гипотеза о соответствии данных нормальному закону распределения: критерий согласия К. Пирсона, который равномерно реагирует на отличия распределений во всем диапазоне, но чувствителен к объему выборки и числу интервалов, на которые разбит диапазон переменной; критерий согласия А.Н. Колмогорова, который реагирует на наибольшую

разницу между теоретическим и эмпирическим распределениями вблизи максимума плотности распределения; критерий согласия В. И. Романовского, который является более простым способом оценки эмпирического распределения нормальному.

При нормальном распределении большая часть значений группируется около некоторого среднего значения, по обе стороны от которого частота наблюдений равномерно снижается. Выборки, строго подчиняющиеся нормальному распределению, на практике, как правило, не встречаются. Поэтому почти всегда необходимо выяснить, можно ли реальное распределение считать нормальным и насколько значительно заданное распределение отличается от нормального распределения. Перед применением любого метода, который предполагает существование нормального распределения, наличие последнего нужно проверять в первую очередь. Классическим примером статистического теста, который исходит из гипотезы о нормальном распределении, можно назвать  $t$ -тест Стьюдента, с помощью которого сравнивают две независимые выборки. Если же данные не подчиняются нормальному распределению, следует использовать соответствующий непараметрический тест, в случае двух независимых выборок –  $U$ -тест Манна и Уитни [2, с. 234-236; 6, с. 114-116]. Если визуальное сравнение реальной гистограммы с кривой нормального распределения кажется недостаточным, можно применить тест Колмогорова-Смирнова [2, с. 237]. Еще одну возможность проверки наличия нормального распределения дает построение графика нормального распределения, в котором наблюдаемые значения сопоставляются с ожидаемыми значениями при нормальном распределении.

Вторым этапом анализа данных является выявление и оценка связи между различными показателями. Степень тесноты связи оценивается коэффициентами, изменяющимися в пределах от 0 до +1. Малое значение коэффициента свидетельствует о слабой связи, близкое к единице – характеризует очень сильную связь и часто позволяет предположить наличие функциональной причинно-следственной связи.

Две выборки зависят друг от друга, если каждому значению одной выборки можно закономерно и однозначным образом поставить в соответствие ровно одно значение другой выборки. Аналогично определяется зависимость нескольких выборок. Чаще всего зависимые выборки возникают, когда измерение проводится для нескольких моментов времени. Зависимые выборки образуют

значения параметров изучаемого процесса, соответствующие различным моментам времени. В SPSS зависимые (также связанные, спаренные) выборки будут представляться разными переменными, которые сопоставляются друг с другом в соответствующем тесте в одной и той же совокупности наблюдений. Если закономерное и однозначное соответствие между выборками невозможно, эти выборки являются независимыми. В SPSS неза-

висимые выборки содержат разные наблюдения (например, относящиеся к различным респондентам), которые обычно различаются с помощью групповой переменной, относящейся к номинальной шкале. В наиболее распространенной ситуации, когда требуется сравнить друг с другом разные выборки по их средним значениям или медианам, обычно применяется один из восьми следующих тестов [2, с. 86–87], приведенных в таблицах 1.1 и 1.2:

Таблица 1.1

**Переменные, относящиеся к интервальной шкале и подчиняющиеся нормальному распределению**

Количество сравниваемых выборок	Зависимость	Тест
2	Независимые	t-тест Стьюдента
2	Зависимые	t-тест для зависимых выборок
>2	Независимые	Простой дисперсионный анализ
>2	Зависимые	Простой дисперсионный анализ с повторными измерениями

Таблица 1.2

**Переменные, относящиеся к порядковой шкале или переменные, относящиеся к интервальной шкале, но не подчиняющиеся нормальному распределению**

Количество сравниваемых выборок	Зависимость	Тест
2	Независимые	U-тест Манна и Уитни
2	Зависимые	Тест Уилкоксона
>2	Независимые	H-тест Крускала и Уоллиса
>2	Зависимые	Тест Фридмана

Труды Минского института управления. 2005. №1

Если следовать разделению статистики на описательную и аналитическую, то задача аналитической статистики – предоставить методы, с помощью которых можно было бы объективно выяснить, является ли наблюдаемая разница в средних значениях или взаимосвязь (корреляция) выборок случайной или нет. Например, если сравниваются два средних значения выборок, то можно сформулировать соответственно две предварительные гипотезы: нулевая гипотеза, когда наблюдаемые различия между средними значениями выборок находятся в пределах случайных отклонений; единичная (альтернативная) гипотеза, когда наблюдаемые различия между средними значениями нельзя объяснить случайными отклонениями.

В аналитической статистике разработаны методы вычисления так называемых тестовых (контрольных) величин, которые рассчитываются по определенным формулам на основе данных, содержащихся в выборках или полученных из них характеристик. Эти тестовые величины соответствуют определенным теоретическим распределениям (t-распределению, F-распределению,  $\chi^2$ -распределению и т. д.), которые позволяют вычислить

вероятность ошибки, находящейся в диапазоне от 0 до 1. В практической статистике она также часто выражается в процентах.

Вероятность ошибки, при которой допустимо отвергнуть нулевую гипотезу и принять альтернативную гипотезу, зависит от каждого конкретного случая. В значительной степени эта вероятность определяется характером исследуемой ситуации. Чем больше требуемая вероятность, с которой надо избежать ошибочного решения, тем более узким выбираются границы вероятности ошибки, при которой отвергается нулевая гипотеза, так называемый доверительный интервал вероятности. Высказывания, имеющие вероятность ошибки  $p \leq 0,05$ , называются значимыми; высказывания с вероятностью ошибки  $p \leq 0,1$  – очень значимыми, а высказывания с вероятностью ошибки  $p \leq 0,001$  – максимально значимыми.

В большинстве случаев перед применением статистического теста ставится вопрос о характере заданных условий. В частности, необходимо выяснить следующие моменты: к какой статистической шкале относится данная переменная; если речь идет о переменных с интервальной шкалой, то подчиняют-

ся ли они закону нормального распределения; являются ли сравниваемые выборки зависимыми или независимыми. Можно сказать, что существует три основных вида статистических шкал, на которых могут сравниваться численные значения [2, с. 82–85]: номинальная; порядковая; масштабная.

Переменные, относящиеся к номинальной шкале, не имеют никакой эмпирической значимости. Они представляют информацию без какого-либо дополнительного упорядочивания. Числовые значения номинальной переменной могут служить и для представления различных категорий (например, пол: 1=мужской; 2=женский). Если мы не задаем тип переменной, то программа объявляет создаваемую переменную номинальной.

Эмпирической значимостью переменных, относящихся к порядковой шкале, является порядок чисел. Порядковые переменные представляют категории с каким-либо внутренним порядком (например, “низкий”, “средний”, “высокий”). Кроме частотного анализа, переменные с порядковой шкалой допускают также вычисление определенных статистических характеристик, таких как медиана. В некоторых случаях возможно вычисление среднего значения. Если должна быть установлена связь (корреляция) с другими переменными такого рода, для этой цели можно использовать коэффициент ранговой корреляции.

Масштабные переменные представляют собой значения, принадлежащие какому-либо интервалу или шкале отношений (например, возраст, доход, стаж), и должны быть числовыми. Переменные, у которых разность (интервал) между двумя значениями имеет эмпирическую значимость, относятся к интервальной шкале. Они могут обрабатываться любыми статистическими методами без ограничений. Например, среднее значение является полноценным статистическим показателем для характеристики таких переменных. К шкале отношений относятся все интервальные переменные, которые имеют абсолютную нулевую точку. Поэтому переменные, относящиеся к интервальной шкале, как правило, имеют и шкалу отношений. На практике, в том числе в SPSS, различие между переменными, относящимися к интервальной шкале и шкале отношений, обычно несущественно. То есть, в дальнейшем всегда речь будет идти о переменных, относящихся к интервальной шкале.

Как уже отмечалось, переменные, относящиеся к номинальной шкале, допускают весьма ограниченные возможности для проведения анализа. Исключение в некоторых случаях составляют дихотомические переменные. Для них

можно, по крайней мере, определять ранговую корреляцию. Однако если переменные, относящиеся к номинальной шкале, не являются дихотомическими, вычисление коэффициентов ранговой корреляции не имеет смысла.

Прежде чем применить статистические методы или строить графики, следует представить собранные данные в форме, пригодной для обработки. При этом рекомендуется придерживаться следующего плана действий [3, с. 96; 6, с. 51; 7, с. 159–192]: провести структурирование набора данных, прежде всего выяснив, к каким категориям относятся наблюдения и к каким – переменные; определить шкалу, к которой относятся переменные; составить кодировочную таблицу; ввести данные в Редактор данных, учитывая кодировочную таблицу; проверить введенные данные на отсутствие ошибок и осмысленность; установить, подчиняются ли нормальному распределению переменные, относящиеся к интервальной шкале.

Теперь можно начинать статистическую обработку введенных данных. Дескриптивный анализ включает описательное представление переменных. К нему относятся создание частотной таблицы, вычисление статистических характеристик или графическое представление данных. Частотные таблицы строятся для переменных, относящихся к номинальной шкале, и для порядковых переменных, имеющих не слишком много категорий. Для переменных, относящихся к номинальной шкале, но не подчиняющихся нормальному распределению, вычисляются медианы и оба квартиля; при небольшом числе категорий можно использовать вариант для концентрированных данных. Для переменных, относящихся к интервальной шкале и подчиняющихся нормальному распределению, чаще всего вычисляются среднее значение и стандартное отклонение или стандартная ошибка. Однако, следует выбрать только одну из этих двух характеристик разброса. Для переменных, относящихся ко всем статистическим шкалам, можно построить большое разнообразие графиков, на которых представлены частоты, среднее значение или другие характеристики. Практически любой статистический анализ, наряду с чисто описательными операциями, включает те или иные аналитические методы, при применении которых определяется вероятность ошибки  $p$ . Большая группа тестов служит для выяснения того, различаются ли две или более различных выборки по своим средним значениям или медианам. При этом учитывается разница между независимыми и зависимыми выборками. В зависимости от количества

выборки (две или более), от того, зависимы ли выборки или нет, относятся ли переменные к интервальной или порядковой шкале, подчиняются ли нормальному распределению – применяются специализированные тесты. Очень часто встречается ситуация, когда сравниваются различные группы наблюдений или значений переменных, относящихся к номинальной шкале. В этом случае строятся таблицы сопряженности. Другая группа тестов касается исследования связей между двумя переменными, то есть выявления корреляций и восстановления регрессий.

Кроме этих довольно простых статистических методов, существуют также более сложные методы многомерного анализа, в которых обычно одновременно используется большое количество переменных. Например, если требуется свести большое количество переменных к меньшему количеству, называемых факторами, то проводится факторный анализ. Если же цель противоположна – объединить заданные наблюдения, образовав из них кластеры, то применяется кластерный анализ [6, с. 467].

Весьма актуальной задачей совершенствования функционирования статистической информационной системы является автоматизация статистической деятельности на предприятиях и организациях, что позволит получить необходимую статистическую информацию для их управления, а также существенно сократить затраты на подготовку исходных данных в органах государственной статистики. Использование пакетов прикладных программ подготовки и первичной обработки статистических отчетов на уровне предприятия (организации) позволяет формировать их сразу в электронном виде, что, в конечном счете, повысит эффективность функционирования статистической информационной системы. В настоящее время на первых этапах внедрения статистической информационной системы пакеты прикладных программ для проведения аналитических исследований являются более универсальными по сравнению с пакетами, реализующими регламентные задачи, и, следовательно, в условиях рыночной экономики могут быть более успешно использованы на уровне предприятий и организаций. В связи с этим, следует отметить, что пакет SPSS позволяет организовать полный цикл исследований по статистическому анализу и прогнозированию данных на основе широкого набора современных методов прикладной статистики.

Программа SPSS представляет собой интегрированный пакет различных модулей, ко-

торые позволяют использовать его на различных этапах исследования. 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) представляют еще два модуля: Advanced Models (продвинутые модели) и Regression Models (регрессионные модели). Эти три модуля охватывают тот спектр методов анализа, который входил в раннюю версию программы для больших ЭВМ. SPSS Base (базовый модуль) 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, дисперсионный анализ, анализ пригодности (анализ надежности) и многомерное шкалирование. Regression Models включает в себя различные методы регрессионного анализа, такие как: бинарная и мультиномиальная логистическая регрессия, нелинейная регрессия и пробит-анализ. В Advanced Models входят различные методы дисперсионного анализа (многомерный, с учетом повторных измерений), общая линейная модель, анализ выживания, включая метод Каплана-Майера и регрессию Кокса, логит модели. Модуль Table служит для создания презентационных таблиц. Здесь предоставляются более широкие возможности по сравнению с упрощенными частотными таблицами и таблицами сопряженности, которые строятся в SPSS Base.

Для расширения программного комплекса SPSS предлагается еще следующий ряд модулей [2, с. 16–18]. Amos (Analysis of moment structures – анализ моментных структур) включает методы анализа с помощью линейных структурных уравнений. Целью программы является проверка сложных теоретических связей между различными признаками случайного процесса и их описание при помощи подходящих коэффициентов. Проверка проводится в форме причинного анализа и анализа траектории. При этом пользователь в графическом виде должен задать теоретическую модель, в которую, вместе с данными непосредственных наблюдений, могут быть



включены дополнительные расширения SPSS, как приемника LISREL (Linear Structural Relationships – линейные структурные взаимоотношения). Answer Tree (дерево решений) включает четыре различных метода автоматизированного деления данных на отдельные группы (сегменты). Деление проводится таким образом, что частотные распределения целевой (зависимой) переменной в различных сегментах значительно различаются. Типичным примером применения данного метода является создание характерных профилей покупателей при исследовании потребительского рынка. Answer Tree является преемницей программы CHAID (Chisquared interaction Detector – детектор взаимодействий на основе  $\chi^2$ -распределения). Модуль Categories содержит различные методы для анализа категориальных данных, а именно: анализ соответствий и три различных метода оптимального шкалирования (анализ однородности, нелинейный анализ главных компонент, нелинейный канонический корреляционный анализ). Clementine – это программа для data mining (добычи знаний), в которых пользователю предлагаются многочисленные подходы к построению моделей, например, нейронные сети, деревья решений, различные виды регрессионного анализа. Clementine представляет собой “верстак” аналитика, при помощи которого можно визуализировать процесс моделирования, перепроверять модели, сравнивать их между собой. Для удобства пользования программой имеется вспомогательная среда внедрения результатов. Conjoint (совместный анализ). Совместный анализ применяется при исследовании рынка для изучения потребительских свойств продуктов на предмет их привлекательности. При этом опрашиваемые респонденты по своему усмотрению должны рас-

положить предлагаемые наборы потребительских свойств продуктов в порядке предпочтения, на основании которого можно будет затем вывести так называемые детализированные показатели полезности отдельных категорий каждого потребительского свойства. Программа Data Entry предназначена для быстрого составления вопросников, а также ввода и чистки данных. Заданные на этапе создания вопросника вопросы и категории ответов потом используются в качестве меток переменных и значений. Exact Tests (точные тесты). Данный модуль служит для вычисления точного значения вероятности ошибки (величины  $p$ ) в условиях ограниченности данных при проверке по критерию  $\chi^2$  (Chi-Quadrat-Test) и при непараметрических тестах, в случае необходимости для этого может быть применен метод Монте-Карло. Программа Goldmine содержит специальную регрессионную модель для регрессионного анализа упорядоченных зависимых и независимых переменных. При помощи Sample Power может быть определен оптимальный размер выборки для большинства методов статистического анализа, реализованных в SPSS. SPSS Missing Value Analysis. Данный модуль служит для анализа и восстановления закономерностей, которым подчиняются пропущенные значения. Он предоставляет различные варианты замены недостающих значений. Модуль Trends содержит различные методы для анализа временных рядов, такие как: модели ARIMA, экспоненциальное сглаживание, сезонная декомпозиция и спектральный анализ.

Комбинация базового модуля с другими позволит использовать его не только как статистический пакет, но и найти большое применение в маркетинговых и социологических исследованиях.

### Литература

1. Информационные технологии в статистике: Учеб. / Под ред. В. П. Божко. М.: Финстатинформ, 2002. 144 с.
2. Ахим Бююль, Петер Цефель: SPSS СПб: ООО “ДиаСофтЮП”, 2002. 608с.
3. Нареш Малхотра: Маркетинговые исследования и эффективный анализ статистических данных. К.: ООО “ТИД”ДС”, 2002. 768 с.
4. Спирков С. Н.: Теория статистики: Учеб. комплекс. Мн.: Изд-во МИУ, 2004. 216 с.
5. Теория статистики: Учеб. / Под ред. Р. А. Шмойловой. М.: Финансы и статистика, 2001. 557 с.
6. Тюрин Ю. Н., Макаров А. А.: Анализ данных на компьютере / Под ред. В. Э. Фигурнова. М.: ИНФРА-М, 2003. 544 с.
7. Методические рекомендации по использованию современных программных средств для математического моделирования и решения экономических задач. Мн.: НИЭИ Минэкономики Республики Беларусь. 2000. 241 с.
8. Информатика в статистике: Словарь справочник. М.: Финансы и статистика, 1994.
9. www.Diasoft.kiev.ua.
10. www.spss.ru