

перерабатывающей промышленности отпадает надобность в создании консультационных служб на уровне регионов (областей, районов). Функцию консультативных служб могут успешно выполнять научно-исследовательские учреждения и высшие учебные заведения, которые способны своевременно предоставлять информацию о новейших достижениях в сфере производства сельскохозяйственной продукции и ее переработки, развития племенного животноводства и элитного семеноводства, в области защиты животных и растений от болезней и вредителей и др. Руководители и специалисты предприятий и организаций АПК, фермеры и владельцы личных подсобных хозяйств и даже садоводы-любители на своих автоматизированных рабочих местах, используя самые современные информационные технологии, смогут получать интересующую их разнообразную информацию, а также, при необходимости, осуществлять ее обмен.

Реализация этого проекта (АСНТИАПК) позволит создать интегрированную информационную систему, объединяющую республиканские, региональные и отраслевые информационные ресурсы и обеспечивающую сбор, накопление, обработку, поиск и представление информации в интересах государственных органов управления АПК, субъектов экономики, науки и образования, общества в целом.

КОМПЬЮТЕРЫ РАСШИРЯЮТ ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ СТАТИСТИКИ

Грабауров В.А., д.т.н., профессор, БГАТУ, г. Минск

Статистика как средство накопления знаний

В последние годы резко возросла ценность знаний. Начали появляться интеллектуальные предприятия, в которых на вершине пирамиды ценностей организации находятся знания (рис. 1).



Рис. 1. Пирамида ценностей интеллектуальной организации

Гуру экономики Питер Друкер выразился весьма категорично: «Сейчас только тот является менеджером, кто заставляет знания работать». Соответственно возникло новое направление Knowledge Management – Управление знаниями, и организации все шире стали использовать системы управления знаниями. Эти системы применяются для автоматизации продаж, управления документами, управление проектами, внутреннего документооборота, отслеживания претензий, Web-публикаций, доставки новостей, в отчетах о затратах, информационных порталах и т.д.

Но знания не приходят сами по себе. В процессе работы мы имеем дело с числами, которые мы сначала упорядочиваем в виде таблицы и только потом принимаем решения. Фактически существует информационная пирамида, в основании которой находятся данные – отдельные символы. На втором уровне располагается информация, отвечающая на вопросы «кто», «что». «где» и «когда». И на третьем уровне находятся знания – применение данных и информации, которые включают стратегию, практику, метод или подход и отвечают на вопросы «как». Переход с нижних на более высокие информационные уровни осуществляется за счет обработки накопленной информации. Таким образом, средством получения знаний является статистика.

Отношение к статистике характеризуется двумя крайностями. С одной стороны, ее обожествляют. Вспомним И.Ильфа, Е.Петрова «Двенадцать стульев»: «Статистика знает все... Все граждане обоого пола занесены в толстые аккуратные книги. Известно, сколько в стране охотников, балерин, станков, собак всех пород, велосипедов, памятников, девушек и швейных машинок...». С другой стороны, статистику называют особой формой лжи, так как мы не всегда доверяем статистическим данным и так часто не сбываются прогнозы.

Истина, очевидно, посредине. Тем не менее, над статистикой висит Дамоклов меч возможной ошибки. Проблема в том, что при изучении сложного объекта мы никогда не знаем всех сопутствующих обстоятельств. И можно лишь стремиться уменьшить опасность ошибки.

Какими путями действуют профессионалы по статистической обработке? Они пытаются создать как можно более совершенный метод обработки для повышения достоверности. Но более совершенный метод обработки – это, как правило, более сложный метод. И возникает следующее противоречие: чем более сложен метод обработки, тем менее он понятен и доступен обычному пользователю! Т.е. возникает тупик. Современная статистика становится доступной только для профессионалов. Поэтому необходимо найти выход из тупика. Таким выходом является использование компьютерных программ, которые сочетают сложные методы обработки и легкий, доступный интерфейс.

Сейчас получили распространение универсальные компьютерные программы типа Excel со встроенными средствами статистической обработки и специальные программы: Statgraphics, SPSS, Statistica for Windows и др.

Компьютерные средства статистической обработки в учебных занятиях

Для проведения занятий со студентами мы используем пакет Statistica, который рассчитан как на профессиональных статистиков, так и на обычных пользователей. Среди применяемых методов статистической обработки есть

такие широко распространенные, как регрессионный анализ, а также менее распространенные многомерные методы, которые без компьютерных программ почти невозможно использовать.

Регрессионный анализ. С помощью регрессионного анализа осуществляется построение модели по результатам наблюдений. При этом выбирается вид зависимости с учетом статистической достоверности модели, оцениваемой посредством коэффициента множественной корреляции. В качестве примера студенты берут из пакета MACRO статистические данные об изменении ВВП США с 1930 по 1980 годы. После построения модели рассчитывается прогнозное значение ВВП в 1990 г. и сравнивается с реальным (рис. 2). Когда выясняется, что ошибка прогноза составляет около 1 % (что, конечно, случайно), на студентов оказывает большое впечатление возможность рассчитывать прогнозные показатели.

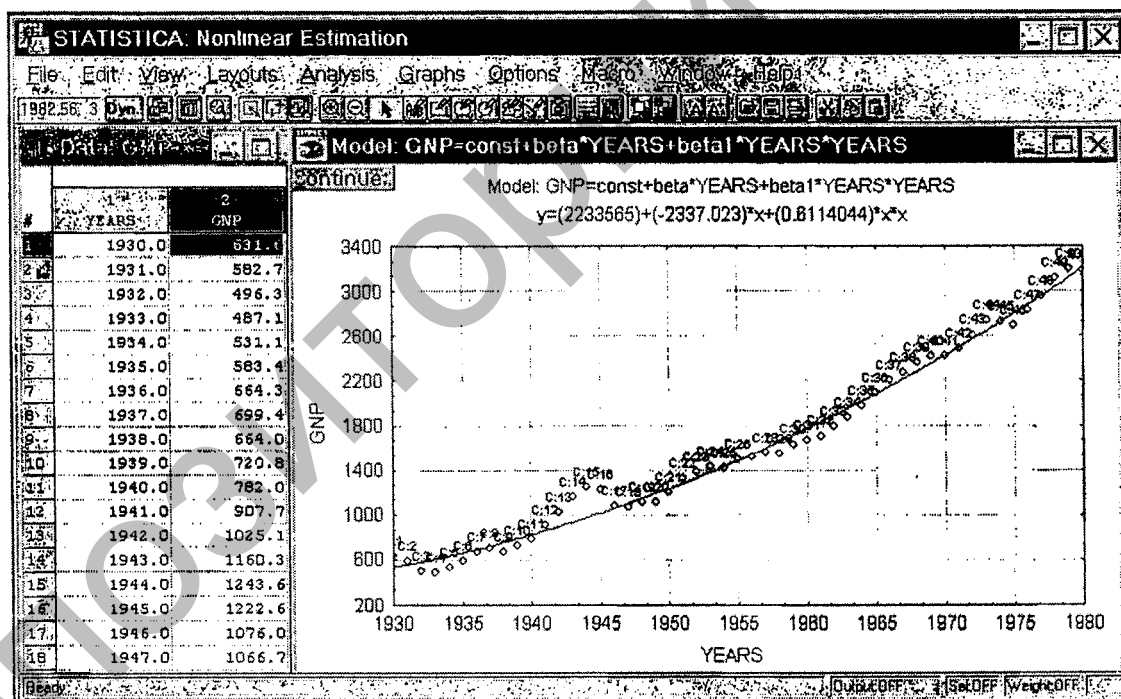


Рис. 2. Построение регрессионной модели изменения ВВП

Экономический объект – это многомерный объект, на который действуют много факторов, в том числе время, и его состояние описывается множеством

параметров, поэтому получили распространение многомерные статистические методы.

Многомерный факторный анализ (Factor Analysis). Основная идея многомерного факторного анализа: дисперсия выходного параметра может быть представлена как сумма дисперсий входных параметров:

$$S^2_y = S^2_{y_1} + S^2_{y_2} + \dots + S^2_{y_k} + S^2_{y_\lambda} + S^2_{y_\delta}$$

где: $\sum_{i=1}^k S^2_{y_i}$ – общая дисперсия (внешние факторы);

$S^2_{y_\lambda}$ – специфическая дисперсия (внутренние факторы);

$S^2_{y_\delta}$ – дисперсия, обусловленная ошибкой измерения.

После преобразований получим: $1 = F_{y_1}^2 + F_{y_2}^2 + \dots + F_{y_k}^2 + \Lambda_y^2 + \Delta_y^2$.

Квадраты факторных нагрузок показывают доли дисперсии измеряемого показателя. Студенты выбирают в пакете MACRO (рис. 3) 7-8 макроэкономических показателей и высказывают свои предположения о том, какие из них в наибольшей степени влияют на ВВП. Все варианты ранжирования факторов записываем на доске. Затем проводим исследования методом многомерного факторного анализа в пакете Statistica, сравниваем полученный результат с выдвинутыми предположениями и обсуждаем причины возможных ошибок.

<input type="checkbox"/> Real GNP	<input type="checkbox"/> Real Consumption Exp.	<input type="checkbox"/> Long-Term Int. Rate
<input type="checkbox"/> Real GNP growth	<input type="checkbox"/> Real Investment Exp.	<input type="checkbox"/> Capital Stock
<input type="checkbox"/> Potential GNP	<input type="checkbox"/> Real Net Exports	<input type="checkbox"/> Employment
<input type="checkbox"/> GNP Gap	<input type="checkbox"/> Real Government Exp.	<input type="checkbox"/> Wages
<input type="checkbox"/> Unemployment Rate	<input type="checkbox"/> Avg Prop to Consume	<input type="checkbox"/> Real Wages
<input type="checkbox"/> Nominal GNP	<input type="checkbox"/> Real Gov. Deficit	<input type="checkbox"/> Investment % of GNP
<input type="checkbox"/> Nominal GNP growth	<input type="checkbox"/> Money Supply	<input type="checkbox"/> Gov Deficit % of GNP
<input type="checkbox"/> GNP Deflator	<input type="checkbox"/> Money Growth Rate	<input type="checkbox"/> Net Exports % of GNP
<input type="checkbox"/> Inflation (GNP)	<input type="checkbox"/> Real Money Supply	<input type="checkbox"/> Savings % of GNP
<input type="checkbox"/> Consumer Price Index	<input type="checkbox"/> Velocity of M1	<input type="checkbox"/> Exchange Rate
<input type="checkbox"/> Inflation Rate (CPI)	<input type="checkbox"/> Short-Term Int. Rate	
<input type="checkbox"/> Disposable Income	<input type="checkbox"/> Real Short Rate	

Рис. 3. Макроэкономические показатели в пакете MACRO

В экономическом анализе большое значение имеет группировка объектов. По результатам группировки можно выделить успешные и отстающие предприятия и принимать соответствующие решения. Проблема группировки

заключается в том, что приходится проводить анализ по множеству разнородных показателей за различные промежутки времени. Наиболее распространенными являются кластерный и дискриминантный анализы. Различие между ними заключается в том, что если задачей кластерного анализа является классификация объектов по сходным признакам, то в дискриминантном анализе имеются признаки нескольких объектов, и нужно определить, к какой группе относится каждый объект. Т.е. дискриминантный анализ – это постановка диагноза.

Кластерный анализ. В качестве примера применения кластерного анализа использовались характеристики 20 инвестиционных фондов США, в которых учитывались показатели прибыли за 5 лет, отдельно за каждый год, риск, затраты и налоги. Так как эти показатели имеют различную размерность, таблица данных в программе Statistica стандартизовывалась с помощью специальной процедуры. Конечный результат анализа представлен в виде кластера (рис. 4). Все инвестиционные фонды распределены по трем группам: наиболее устойчивые, которым можно доверять, неустойчивые, которым доверять не следует и третья группа, занимающая промежуточное положение.

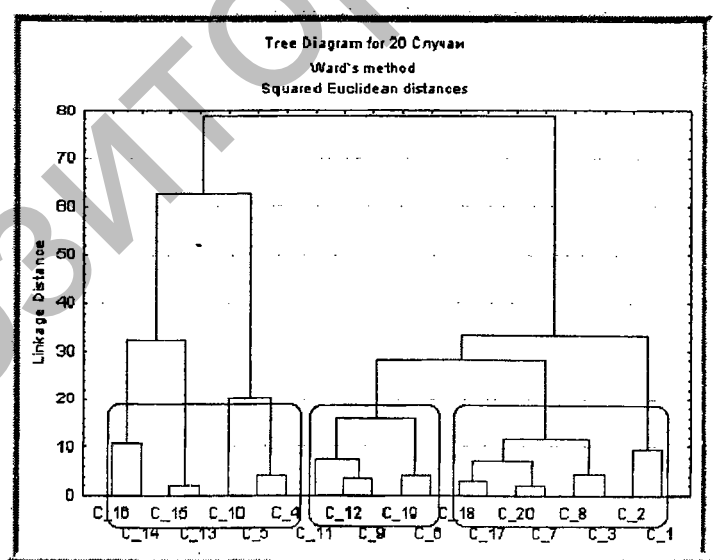


Рис. 4. Результаты кластерного анализа инвестиционных фондов

Дискриминантный анализ. Для дискриминантного анализа (Discriminant Analysis) использовались данные группировки

сельскохозяйственных объектов по 5 группам. При группировке использовались 7 показателей деятельности объектов:

- X_1 – прибыль;
- X_2 – продукция на 1 работника;
- X_3 – продукция на 1 га;
- X_4 – производство молока на 1 га;
- X_5 – производство мяса на 1 га;
- X_6 – выручка на 1 работника;
- X_7 – выручка на 1 га.

Объекты уже были распределены по 5 классам до начала исследования, и задачей анализа являлось выяснение достоверности проведенной классификации. Правильность классификации оценивалась по критерию максимума отношения дисперсии между классами к дисперсии внутри классов. В программе Statistica предусматривается возможность перемещения объектов между классами и проверяется достоверность результатов. Анализ заканчивается, когда достигается 100% правильность. В конечном окне результатов дискриминантного анализа (рис. 5) появляются классификационные функции по каждому из 5 классов. Эти функции дают возможность классифицировать новые объекты.

Переменная	Classification Functions; grouping: CLASS (Discrimin				
	G 1:1 p=,10000	G 2:2 p=,10000	G 3:3 p=,15000	G 4:4 p=,30000	G 5:5 p=,35000
X 1	0,162	0,142	0,0794	0,0372	-0,01163
X 2	-0,020	-0,017	-0,0087	-0,0031	0,00373
X 3	1,076	0,933	0,5278	0,2793	-0,05179
X 4	-1,450	-1,226	-0,6641	-0,3523	0,10154
X 5	11,692	10,199	5,5922	3,1011	-0,54263
X 6	22,971	19,927	11,3611	5,6352	-1,51277
X 7	-0,189	-0,168	-0,0990	-0,0538	0,00998
Постоянн	-362,296	-286,051	-95,2500	-36,1104	-9,78125

Рис. 5. Окно результатов дискриминантного анализа

Например, «зависимая» переменная Y , определяющая мнение эксперта о группировке для первого класса: $Y_1 = -326,296 + 0,162x_1 - 0,020x_2 + 1,076x_3 - 1,450x_4 + 11,692x_5 + 22,971x_6 - 0,189x_7$. Новые случаи будут относиться к тому классу, для которого классификационное значение Y будет максимальным.

Дисперсионный анализ ANOVA (*Analysis of Variance*). В дисперсионном анализе исследуемый фактор известен заранее. Цель анализа – выявление значимости фактора. Таким фактором может быть уровень заработной платы, различные виды упаковки товара, применение новых средств обучения и т.д. Так как вариации результирующего показателя зависят не только от исследуемого фактора, но и от неизвестных, необходимо выяснить, что же влияет на вариации результирующего показателя.

Для проведения анализа результаты эксперимента помещают в таблицу (табл. 1), в каждой колонке которой размещены данные по уровням исследуемого фактора.

Таблица 1. Данные для однофакторного дисперсионного анализа

Уровни фактора	1	2	...	k	...	m
	x_{11}	x_{12}	...	x_{1k}	...	x_{1m}

Результаты измерений	x_{i1}	x_{i2}	...	x_{ik}	...	x_{im}

	x_{n1}	x_{n2}	...	x_{nk}	...	x_{nm}

Таких уровней может быть два или несколько. Данные внутри каждой колонки различаются, так как на результаты влияют неизвестные факторы, например, индивидуальные качества работников. Цель анализа – выяснить, существенно ли отличие дисперсии между колонками (уровнями исследуемого фактора) от дисперсии внутри колонок (неизвестные факторы). Для этого

вычисляется F-отношение по распределению Фишера и сравнивается с табличным значением для выбранного уровня риска. Пример результатов дисперсионного анализа в программе Statistica показан на рис. 6.

Эффект	Тест	Значени	F	Эффект df	Ошибка df	p
OTREZOK	Wilks	0,001105	226,0000	4	1	0,049843

Additional text in the window: Multivariate Tests of Significance (Anova), Sigma-ограниченная параметризация, Effective hypothesis decomposition.

Рис. 6. Окно результатов дисперсионного анализа

Дополнительные возможности статистических программ

Кроме приведенных выше распространенных статистических процедур, которые мы применяем в учебном процессе, статистические программы в процессе совершенствования наращивают дополнительные возможности. Такие возможности появляются как внутри универсальных статистических программ, так и появляются новые, специальные программы.

Статистика и качество. Как известно, по распределению Стьюдента вероятность попадания случайной величины в диапазоне $\pm 3\sigma$ составляет 99,7%. Т.е. за пределами этого диапазона остается только небольшая часть – 0,3%. Отсюда появилась и получила широкое распространение во всем мире программа качества под названием «шесть сигм» - 99,7% годных изделий.

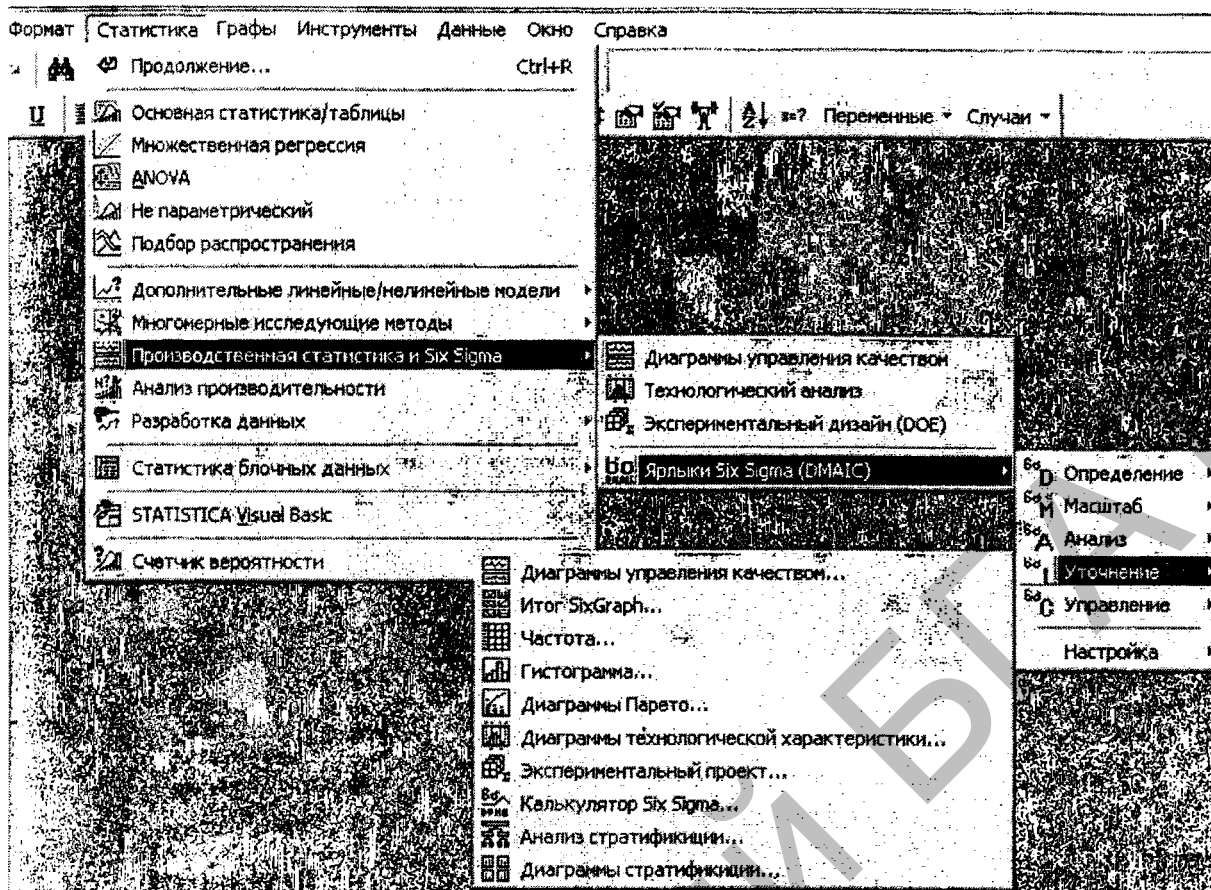


Рис. 7. Окно модуля «Производственная статистика и Six Sigma» в пакете Statistica

В пакете Statistica имеется специальный модуль под названием «Производственная статистика и Six Sigma» (рис. 7). Этот модуль содержит множество специальных статистических процедур, применяемых для обработки производственной информации, таких как построение диаграмм управления качеством, Парето-диаграмм, технологический анализ и др.

Углубленный анализ данных. В статистической обработке данных приходится иметь дело с различными масштабами информации. Если объемы данных измеряются тысячами, то применяются обычные программы. Но иногда появляется необходимость обрабатывать громадные массивы, измеряемые миллионами данных. Для таких случаев разработаны специальные программы, работающие по принципам нейронных сетей. Эти программы получили название Data Mining - «Переработка руды данных». В таких программах обработка ведется по другим принципам. В отличие от обычной оперативной статистической обработки, где вопросы ставятся прямо: «каковы такие-то

показатели?», в Data Mining вопросы ставятся иначе: «существует ли вероятность чего либо?». Различие задач, решаемых при помощи обычной статистической обработки и Data Mining показано в таблице 2.

Табл. 2. Сравнение задач для обычной статистики и Data Mining

Обычная обработка	Data Mining
Каковы средние показатели травматизма для курящих и некурящих?	Встречаются ли точные шаблоны в описаниях людей, подверженных повышенному травматизму?
Каковы размеры телефонных счетов существующих клиентов в сравнении со счетами клиентов, отказавшихся от услуг телефонной компании?	Имеются ли характерные портреты клиентов, которые, по всей вероятности, собираются отказаться от услуг телефонной компании?
Какова средняя величина ежедневных покупок по украденной и не украденной кредитной карточке?	Существуют ли стереотипные схемы покупок для случаев мошенничества с кредитными карточками?

В программах Data Mining проводится обнаружение скрытой информации за счет просеивания огромного количества данных. Эти программы применяются для рыночной сегментации, составления характеристик клиентов, обнаружения мошенничества, анализа потребительской корзины, анализа тренда и т.д.

Использование компьютерных статистических программ имеет еще одну сторону. На основе статистической обработки с помощью компьютеров создаются математические модели. А ведь именно применение математики и компьютерных программ является признаком научного исследования. Характерны высказывания по этому поводу двух мыслителей.

Иммануил Кант: «Учение о природе будет содержать науку в собственном смысле лишь в той мере, в какой может быть применена в ней математика».

Дональд Кнут: «Наука – это та часть наших знаний, которую мы сумели понять настолько хорошо, что можем обучить этому ЭВМ. Там, где мы еще не

достигли такого понимания, речь пока идет лишь о профессиональном искусстве».

СИСТЕМНАЯ ИНФОРМАТИЗАЦИЯ ПРОЦЕССОВ ПРИНЯТИЯ УПРАВЛЕНЧЕСКИХ РЕШЕНИЙ

**Чернышев В.О., проф., д.т.н., главный научный сотрудник,
Величко Л.Н., зам. ген. директора по маркетингу,
Качура Л.П., зам. ген. директора по развитию,
Метлицкий Ю.Н., зам. ген. директора по проектам
ЗАО «НПП Белсофт» (г. Минск)**

Принятая¹ общегосударственная программа «Электронная Беларусь» предусматривает комплексную (системную) информатизацию всех сфер общественно-полезной жизни и деятельности белорусского народа и создания информационного общества нашей страны. Эта программа в первую очередь предусматривает информатизацию и интеграцию деловых процессов управленческой деятельности отдельных субъектов хозяйствования и отраслей экономики в целом. Основными компонентами управленческой деятельности являются процессы планирования (стратегического и тактического) и прогнозирования, которые определяют создание автоматизированных систем подготовки и поддержки принятия решений (СППР). Эти системы оказывают существенную помощь лицам, принимающим решения (ЛПР)

При реализации задач комплексной информатизации управленческой деятельности возникает необходимость выбора наиболее прогрессивного подхода и современной методологии построения инструментальной платформы СППР. В этих условиях важнейшим инструментом является системный подход к информатизации процессов планирования, прогнозирования и информационной поддержки принимаемых управленческих решений.

Системный подход [1] представляет собой современный инструмент организационной интеграционной деятельности по исследованию (анализу и синтезу) плохоструктуризированных сложных систем как единого целого,

¹ О государственной программе информатизации Республики Беларусь на 2003-2005 годы и на перспективу до 2010 года «Электронная Беларусь». Постановление Совета Министров Республики Беларусь от 27 декабря 2002г., № 1819