

ФГБОУ ВО Новосибирский ГАУ
Факультет Экономики и управления
Кафедра Информационных технологий и моделирования

Планирование и управление данными

Методические указания для проведения лабораторно-практических занятий и
самостоятельной работы студентов

Новосибирск 2021

УДК 004.451.5 (07)
ББК 32.972.134, я 7
П 372

Составитель: О.В.Агафонова, доцент кафедры Информационных технологий и моделирования, к.э.н.

Рецензент: И.В.Трубчанинова, доцент кафедры Учета и финансовых технологий, к.э.н.

Планирование и управление данными: методические указания для проведения лабораторно-практических занятий и самостоятельной работы студентов / Новосиб. гос. аграр. ун-т. Фак. ЭиУ; сост. О.В. Агафонова - Новосибирск, 2021. – 10 с.

Методические рекомендации для проведения лабораторно-практических занятий и самостоятельной работы студентов по дисциплине «Планирование и управление данными» предназначены для студентов направления подготовки 09.03.03 Прикладная информатика всех форм обучения.

Методические указания утверждены и рекомендованы к изданию учебно-методическим советом факультета экономики и управления (протокол №4 от «28» декабря 2021 г.)

© Новосибирский государственный аграрный университет, 2021

1. Введение

Лабораторно-практическое занятие – форма организации обучения, которая направлена на формирование практических умений и навыков и является связующим звеном между самостоятельным теоретическим освоением обучающимися учебной дисциплины и применением ее положений на практике.

Лабораторно-практические занятия проводятся в целях: выработки практических умений и приобретения навыков в решении задач, выполнении заданий, производстве расчетов, разработке и оформлении документов, практического овладения компьютерными технологиями. Главным их содержанием является практическая работа каждого обучающегося.

Самостоятельная работа обучающихся – вид деятельности, при котором в условиях систематического уменьшения прямого контакта с преподавателем обучающимися выполняются учебные задания.

При этом специфика самостоятельной работы обучающихся заключается в том, чтобы они самостоятельно получали новые знания.

Цель и задачи дисциплины

Целью изучения курса «Планирование и управление данными» является: формирование у обучающегося целостной системы знаний о теоретических, методологических и практических подходах, используемых в работе с данными.

Исходя из цели, в процессе изучения дисциплины решаются следующие задачи:

- применение знаний, полученных в области информационных технологий и программных средств при решении задач профессиональной деятельности;
- использование современных информационных технологий, в том числе отечественного производства при решении задач профессиональной деятельности;
- обладание понятиями и категориями математического моделирования, используемыми при расчете экономических и организационно-технических процессов;
- применение методов теории систем и системного анализа, дискретной математики, теории вероятностей и математической статистики, методов оптимизации и исследования операций, нечетких вычислений, математического моделирования.

Дисциплина «Планирование и управление данными» в соответствии с требованиями ФГОС ВО направлена на формирование следующих компетенций бакалавра:

ОПК-2. Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности;

ОПК-6. Способен анализировать и разрабатывать организационно-технические и экономические процессы с применением методов системного анализа и математического моделирования.

2. Цель проведения лабораторно-практических занятий и самостоятельной работы обучающихся

Лабораторно-практические занятия направлены на закрепление и расширение знаний, полученных на лекциях.

Лабораторно-практические занятия по курсу «Планирование и управление данными» направлены на практическое углубленное изучение учебной дисциплины, привитие обучающимся навыков самостоятельного поиска и анализа учебной информации, формирование и развитие у них научного мышления, умения активно участвовать в творческой дискуссии, делать правильные выводы, аргументировано излагать и отстаивать свое мнение.

Цель самостоятельной работы обучающихся – овладение методами получения новых знаний, приобретение навыков самостоятельного анализа, усиление научных основ практической деятельности.

3. Содержание занятий

1. Анализ данных

1.1 Статистические методы обработки данных

Основы статистического метода исследований. Статистические методы в практике обработки данных. Статистическое изучение взаимосвязей. Изучение динамики общественных явлений. Индексы.

1.2 Анализ многомерных данных

Методы логико-алгебро-геометрического направления: методы исследований и построения зависимостей (множественный корреляционно-регрессионный анализ, множественный дисперсионный и ковариационный анализ, метод канонических корреляций, путевой анализ); методы сжатия информации и снижения размерности (метод главных компонент, факторный анализ); методы классификации (кластерный анализ, дискриминантный анализ)

1.3 Функции анализа, планирования, управления, добычи, визуализации данных в программе Statistica

Операции по вычислению моды, медианы, экстремумов, квартилей и других статистик. Корреляционный метод для определения характеристик взаимосвязи между показателями. Вычисление коэффициентов (Пирсон, Крамер, Кендалл и т.д.), получение различных видов корреляций, вывод табличных результатов и последующая удобная работа с ними.

Графические возможности пакета Statistica. Быстрое построение графиков и способы настройки изображения для визуализации результатов исследования.

Графический анализ таблиц, варианты редактирования графиков, диаграмм и гистограмм, средства для создания собственных настроек изображений. Технические изображения, чертежи, бизнес-графики.

2. Интеллектуальный анализ данных

2.1 Технологии многомерной обработки больших объемов данных

Общая схема организации хранилища данных.

Хранилища данных, интегрирующие данные для анализа информации из нескольких источников на предприятии. Системы оперативной аналитической обработки, позволяющие получить ответы на запросы, охватывающие большие объемы данных в поисках общих тенденций. Приложения добычи данных, служащие для выявления знаний за счет полуавтоматического поиска ранее неизвестных шаблонов и связей в базах данных.

2.2 Основные продукты DataMining

Алгоритмы и технологии DataMining. Определения OLAP, DataMining и взаимосвязи между ними. Аналитическая отчетность и многомерное представление данных. Хранилище данных. Основные операции над кубом данных. Типы задач, решаемые методами DataMining. Критерии выбора аналитических платформ и пакетов DataMining. Основные этапы внедрения систем анализа данных. Категории пользователей аналитических систем. Способы снижения рисков проектов DataMining.

2.3 Аналитическая платформа Deductor: хранилище данных

Платформа Deductor. Хранилище данных. Создание нового хранилища. Преобразование структуры хранилища. Загрузка информации. Извлечение информации.

2.4 Многомерные отчеты и OLAP

Особенности OLAP. Разработка системы аналитической отчетности.

2.5 Логистическая регрессия, деревья решений и ассоциативные правила

Основы методов логистической регрессии и деревьев решений. Техника практического применения в Deductor. Техника применения ассоциативных правил.

4. Задачи для лабораторно-практических занятий

Задание 1. Оценить параметры идентификационной структурной модели на основе условных исходных данных

$$y_1 = a_{10} + b_{12}y_2 + a_{11}x_1 + \varepsilon_1$$

$$y_2 = a_{20} + b_{21}y_1 + a_{22}x_2 + \varepsilon_2$$

№: пп	y_1	x_1	y_2	x_2
1	5	2	7	2
2	4	3	8	2
3	3	1	6	4
4	2	5	4	6
5	7	4	2	10
6	8	3	7	8
7	9	6	4	9
8	10	8	2	7
Итого				
Среднее значение				

Задание 2. Требуется построить регрессионную модель зависимости заработной платы работника (Y от возраста (X) 20 работников одного предприятия

Данные для построения регрессионной модели заработной платы

№	Y-заработная плата работника за месяц,долл.	X-возраст работника,лет	Z=пол, М/Ж
1	300	29	Ж
2	400	40	М
3	300	36	Ж
4	320	32	Ж
5	200	23	М
6	350	45	Ж
7	350	38	Ж
8	400	40	Ж
9	380	50	М
10	400	47	М
11	250	28	М
12	350	30	Ж
13	200	25	М
14	400	48	М
15	220	30	Ж
16	320	40	М
17	390	40	М
18	360	38	М
19	260	29	Ж
20	250	25	М

Задание 3. Количество сверхурочных часов, отработанных группой работников за неделю

Количество сверхурочных часов	0-2	2-4	4-6	6-8	8-10	10-2
Количество работников	4	8	15	23	12	6

Задание: 1. Постройте график и дайте его характеристику.
2. Рассчитайте показатели центральной тенденции и вариации.

5. Контролирующие материалы для аттестации по дисциплине

Задачи для контрольной работы

Задача 1

Осуществлены 2 серии из n_1 и n_2 независимых испытаний, причем в первой событие А произошло m_1 раз, а во второй серии – m_2 раз. Найти оценку максимального правдоподобия неизвестной вероятности p события А в каждом испытании (считая эту вероятность одной и той же постоянной в обеих сериях).

Задача 2.

Найти доверительный интервал для оценки с надежностью 0,99 и 0,95 неизвестного математического ожидания μ нормально распределенного признака X генеральной совокупности, если даны генеральное среднее квадратическое отклонение σ , выборочная средняя \bar{x} и объем выборки n .

а) $\sigma = 4$ $\bar{x} = 10,2$ $n = 16$

б) $\sigma = 5$ $\bar{x} = 16,8$ $N = 25$

Задача 3.

Количество сверхурочных часов, отработанных группой работников за неделю

Количество сверхурочных часов	0-2	2-4	4-6	6-8	8-10	10-2
Количество работников	2	6	13	15	8	5

Задание:

1. Постройте график и дайте его характеристику.
2. Рассчитайте показатели центральной тенденции и вариации.

Обоснуйте решение задач.

Вопросы для экзамена

1. Методология анализа сложных систем.
2. Основные концепции построения хранилищ данных.
3. Построение автоматизированных систем предобработки данных.
4. Алгоритмы и технологии DataMining.
5. Обзор алгоритмов построения деревьев решений.
6. Математические основы нейросетевых технологий.
7. Методология построения регрессионных моделей.
8. Применение технологии деревьев решений для оценки кредитоспособности физических лиц.
9. Применение нейросетевых технологий для оценки кредитоспособности физических лиц.
10. Ассоциативные правила, как инструмент повышения прибыльности в розничной и оптовой торговле.
11. Задачи классификации, как инструмент повышения эффективности бизнеса.
12. Методология построения и верификации прогнозных моделей.
13. Основные методы прогнозирования.
14. Методы оценки качества прогнозных моделей.
15. Методика анализа данных.
16. Определения OLAP, DataMining и взаимосвязи между ними.
17. Особенности OLAP.
18. Аналитическая отчетность и многомерное представление данных.
19. Хранилище данных.
20. Основные операции над кубом данных.
21. Типы задач, решаемые методами DataMining.
22. Алгоритмы, получившие наибольшее распространение для каждого типа задач: самоорганизующиеся карты, деревья решений, линейная регрессия, нейронные сети, ассоциативные правила.
23. Этапы подготовки данных.
24. Методы проведения экспертиз для выявления наиболее значимых факторов.
25. Понятия парциальной и комплексной обработки.
26. Анализ качества полученных моделей.
27. Критерии выбора аналитических платформ и пакетов DataMining.
28. Основные этапы внедрения систем анализа данных.
29. Категории пользователей аналитических систем.
30. Способы снижения рисков проектов DataMining.

6. Рекомендуемая литература

Основная литература

1. Агальцов, В. П. Базы данных: учебник: в 2-х кн. Книга 1. Локальные базы данных. - Москва: ФОРУМ: ИНФРА-М, 2020. - 352 с. (ЭБС «Инфра-М»)
2. Агальцов, В. П. Базы данных: в 2 книгах. Книга 2. Распределенные и удаленные базы данных: учебник / В.П. Агальцов. - Москва: ФОРУМ: ИНФРА-М, 2021. - 271 с. (ЭБС «Инфра-М»)
3. Балдин, К. В. Информационные системы в экономике: учебник / К. В. Балдин, В. Б. Уткин. - 8-е изд., стер. - Москва: Издательско-торговая корпорация «Дашков и К°», 2019. - 394 с.(ЭБС «Инфра-М»)

Дополнительная литература

1. Балдин, К. В. Информационные системы в экономике: учебное пособие / К.В. Балдин. - Москва: ИНФРА-М, 2022. - 218 с. (ЭБС «Инфра-М»)
2. Информационные системы в экономике: учебное пособие / под ред. Д. В. Чистова. - Москва: ИНФРА-М, 2019. - 234 с.(ЭБС «Инфра-М»)
3. Обработка экспериментальных данных на ЭВМ: учебник / О.С. Логунова, П.Ю. Романов, Е.А. Ильина [и др.]. - Москва: ИНФРА-М, 2019. — 326 с. (ЭБС «Инфра-М»)

Информационное обеспечение

1. Электронно-библиотечная система издательства «Лань» // www.e.lanbook.com.
2. Электронно-библиотечная система издательства «ИНФРА-М» // www.znaniium.com.
3. Научная электронная библиотека eLibrary.ru
4. Электронно-библиотечная система НГАУ// <http://nsau.edu.ru/library/ebooks/e-lib-sys-nsau/>
5. Библиотечно-информационный ресурс «Федеральное собрание образовательных материалов для студентов».

Составитель
Агафонова Ольга Витальевна

Методические указания для проведения лабораторно-практических занятий и
самостоятельной работы студентов

Объем 0,63 уч. – изд. л.

Новосибирский государственный аграрный университет

630039, Новосибирск, ул. Добролюбова, 160