

Белорусский государственный университет



Регистрационный № УД-6326 /уч.

СТАТИСТИЧЕСКИЙ АНАЛИЗ ДАННЫХ

Учебная программа учреждения высшего образования по учебной дисциплине для специальности первой степени высшего образования:

**1-31 03 06 Экономическая кибернетика (по направлениям)
направления специальности**

1-31 03 06 - 01 Экономическая кибернетика (математические методы и компьютерное моделирование в экономике)

2018 г.

Учебная программа составлена на основе образовательного стандарта высшего образования ОСВО 1-31 03 06-2013 и учебного плана G31-166/уч. от 30.05.2013, G31и-191/уч. от 30.05.2013.

СОСТАВИТЕЛИ:

В.И. Малюгин, доцент кафедры математического моделирования и анализа данных Белорусского государственного университета, кандидат физико-математических наук, доцент.

Ю.С. Харин, профессор кафедры математического моделирования и анализа данных Белорусского государственного университета, доктор физико-математических наук, чл.-корр. НАНБ, профессор.

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой математического моделирования и анализа данных Белорусского государственного университета (протокол № 13 от 29 марта 2018);

Научно-методическим Советом Белорусского государственного университета (протокол № 5 от 4 мая 2018 г.).

Григорьев И.А., зав. кафедрой ММД

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Учебная дисциплина «Статистический анализ данных» для специальности «Экономическая кибернетика» предполагает изучение методов статистического анализа данных в экономических, финансовых и других приложениях, в предположении, что данные имеют вероятностную (стохастическую) природу, а для их описания и анализа используются многомерные вероятностно-статистические модели и методы.

Теоретический курс поддерживается лабораторным практикумом, предполагающим использование статистических пакетов прикладных программ (ППП) и языков программирования с развитыми возможностями статистического анализа данных (языки R и Python).

Учебная дисциплина «Статистический анализ данных» знакомит студентов с классическими и современными методами решения задач статистического анализа многомерных данных. К ним относятся: предварительный дескриптивный и графический анализ данных, проверка гипотез о свойствах моделей данных, исследование и построение моделей статистических зависимостей между анализируемыми переменными, анализ аномальных наблюдений; анализ неоднородных данных с помощью методов статистической классификации.

Целью дисциплины «Статистический анализ данных» является:

- освоение студентами теоретических основ статистического анализа;
- формирование навыков их практического применения для решения задач анализа экономических данных с помощью статистического программного обеспечения.

Задачами дисциплины «Статистический анализ данных» являются:

- изучение теоретических основ статистического анализа многомерных данных;
- овладение типовыми инструментальными средствами, включая статистические ППП и язык R;
- формирование практических навыков решения типовых задач статистического анализа экономических данных с помощью стандартного статистического программного обеспечения.

Учебная дисциплина «Статистический анализ данных» относится к циклу дисциплин специализаций.

Учебная дисциплина основывается на учебных дисциплинах «Введение в компьютерный и интеллектуальный анализ данных», «Теория вероятностей и математическая статистика». Знания, полученные в результате изучения дисциплины, будут использованы при изучении дисциплины «Имитационное и статистическое моделирование».

В результате освоения учебной дисциплины студент должен:

знать:

- методы статистического оценивания многомерных распределений;
- методы статистической проверки гипотез;

- методы предварительного статистического анализа данных;
- методы корреляционного анализа;
- методы регрессионного анализа данных;
- методы дискриминантного анализа многомерных данных;
- методы кластерного анализа многомерных данных;

уметь:

- осуществлять предварительный статистический анализ данных с целью установления модели данных, выявления кластерной структуры данных и аномальных наблюдений;
- проводить статистический анализ многомерных данных с целью установления статистических зависимостей методами корреляционного, регрессионного и дисперсионного анализа между экономическими переменными;
- осуществлять статистический анализ однородности многомерных данных на основе графического анализа и статистических критериев, а также классификацию неоднородных данных с помощью методов дискриминантного и кластерного анализа.

владеть:

- методами решения основных задач статистического анализа многомерных данных;
- навыками по подготовке данных и решения типовых задач статистического анализа данных;
- навыками применения современных ППП для решения задач статистического анализа многомерных данных в экономических и других приложениях;
- навыками по подготовке отчетов с результатами статистического анализа данных, включающих содержательную (экономическую) интерпретацию результатов анализа, комментарии, выводы и рекомендации.

Освоение учебной дисциплины «Статистический анализ данных» должно обеспечить формирование следующих академических, социально-личностных и профессиональных компетенций:

академические компетенции:

АК-1. Уметь применять базовые научно-теоретические знания для решения теоретических и практических задач.

АК-2. Владеть системным и сравнительным анализом.

АК-3. Владеть исследовательскими навыками.

АК-4. Уметь работать самостоятельно.

АК-5. Быть способным породить новые идеи (обладать креативностью).

АК-6. Владеть междисциплинарным подходом при решении проблем.

АК-7. Иметь навыки, связанные с использованием технических устройств, управлением информацией и работой с компьютером.

АК-8. Владеть навыками устной и письменной коммуникации.

АК-9. Уметь учиться, повышать свою квалификацию в течение всей жизни.

социально-личностные компетенции:

СК-6. Уметь работать в команде.

профессиональные компетенции:

ПК-1. Работать с научно-технической, нормативно-справочной и специальной литературой.

ПК-2. Заниматься аналитической и научно-исследовательской деятельностью в области экономической кибернетики.

ПК-4. Профессионально ставить задачи, выработать и принимать решения.

ПК-5. Владеть современными методами математического и компьютерного моделирования систем и процессов, участвовать в исследованиях и разработке новых методов и технологий.

ПК-13. Анализировать и оценивать собранные данные.

ПК-15. Готовить доклады, материалы к презентациям и представлять их на них.

ПК-27. Работать с научной, технической и патентной литературой.

Структура содержания учебной дисциплины включает такие дидактические единицы, как темы (разделы), в соответствии с которыми разрабатываются и реализуются соответствующие лекционные и лабораторные занятия. Примерная тематика занятий приведена в информационно-методической части.

Дисциплина изучается в 5 семестре. Всего на освоение учебной дисциплины «Статистический анализ данных» отведено 154 часов, в том числе 68 аудиторных часов, из них: лекции – 34 часа, лабораторные занятия – 30 часов, УСР – 4 часа.

Трудоемкость учебной дисциплины составляет 4 зачетные единицы.

Форма текущей аттестации – экзамен.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Тема 1. Общая характеристика статистического анализа данных (САД). Принципы и общая схема статистического анализа. Типы данных. Шкалы измерения. Классификация задач и методов САД.

Тема 2. Предварительный статистический анализ данных. Математические модели данных. Наблюдения как случайные величины, векторы, функции. Модель данных «случайная выборка». Альтернативные модели данных: неоднородная выборка, выборка с засорениями, зависимые наблюдения. Методы предварительного анализа данных. Примеры.

Тема 3. Применение методов статистического оценивания и проверки гипотез в задачах анализа данных. Функциональные и числовые характеристики вероятностных моделей данных и их статистические оценки. Параметрические методы оценивания и подстановочный принцип. Принципы теории статистической проверки гипотез. Общий вид статистического критерия. Понятие Р-значения. Адекватность адекватности моделей данных на основе статистических критериев согласия.

Тема 4. Многомерное нормальное (гауссовское) распределение как модель многомерных данных. Многомерное нормальное распределение как модель многомерных данных. Функциональные и числовые характеристики, маргинальные распределения, линейные преобразования гауссовских случайных векторов.

Тема 5. Условное распределение гауссовского вектора. Функция регрессии и ее оптимальные свойства. Частный и множественный коэффициенты корреляции: определение и свойства.

Тема 6. Статистическое оценивание параметров многомерной гауссовской модели. Оценивание параметров многомерной гауссовской модели по методу максимального правдоподобия. Свойства и распределение оценок.

Тема 7. Корреляционный анализ. Выборочный коэффициент корреляции: свойства и применения. Выборочный частный коэффициент корреляции: свойства и применения. Выборочный множественный коэффициент корреляции: свойства и применения. Статистические выводы о значениях коэффициентов корреляции.

Тема 8. Статистическая проверка гипотез на основе T^2 -статистики Хоттелинга. Многомерный T^2 -критерий Стьюдента. Свойства статистики Хоттелинга. Применение статистики Хоттелинга в анализе данных. Проверка

гипотезы о значениях вектора математического ожидания. Построение доверительной области. Проверка гипотезы о равенстве средних в двух выборка. Проблема Беренса-Фишера.

Тема 9. Регрессионный анализ данных. Регрессионная модель данных. Статистическое оценивание параметров многомерной линейной регрессии. Свойства оценок параметров. Проверка гипотез относительно параметров модели многомерной линейной регрессии.

Тема 10. Методы статистической классификации многомерных данных. Задачи статистической классификации многомерных данных. Дискриминантный анализ данных при наличии обучающей выборки. Подстановочное байесовское решающее правило. Оптимальная классификация гауссовских случайных векторов. Кластерный анализ данных: общая характеристика алгоритмов классификации.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов			Количество часов УСР	Форма контроля знаний
		Лекции	Семинарские занятия	Лабораторные занятия		
1	Общая характеристика статистического анализа данных (САД)	2				Устный опрос
2	Предварительный статистический анализ данных	4		4		Защита подготовленного студентом реферата. Отчет по заданию с устной защитой
3	Применение методов статистического оценивания и проверки гипотез в задачах анализа данных	4		4	2	Устный опрос
4	Многомерное нормальное (гауссовское) распределение как модель многомерных данных	4		2		Контрольная работа №1. Отчет по заданию с устной защитой
5	Условное распределение гауссовского вектора	2		2		Устный опрос
6	Статистическое оценивание параметров многомерной гауссовской модели	2		2		Коллоквиум
7	Корреляционный анализ	4		4		Отчет по заданию с устной защитой

8	Регрессионный анализ данных	4		4	2	Защита подготовленного студентом реферата. Отчет по заданию с устной защитой
9	Статистическая проверка гипотез на основе T^2 -статистики Хотеллинга	4		4		Контрольная работа №2. Отчет по заданию с устной защитой
10	Методы статистической классификации многомерных данных	4		4		Защита подготовленного студентом реферата
ИТОГО		34		30	4	

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Перечень основной литературы

1. Малюгин, В. И. Введение в компьютерный и интеллектуальный анализ данных. Учебно-методические материалы / В. И. Малюгин. – Минск : БГУ, 2014. – 160 с.
2. Харин, Ю. С. Математическая и прикладная статистика / Ю. С. Харин, Е. Е. Жук. – Минск : БГУ, 2005. – 457 с.
3. Харин, Ю. С. Математические и компьютерные основы статистического моделирования анализа данных / Ю. С. Харин, В. И. Малюгин, М. С. Абрамович. – Минск : БГУ, 2008. – 455 с.
4. Харин, Ю. С. Теория вероятностей, математическая и прикладная статистика / Ю. С. Харин, Н. М. Зуев, Е. Е. Жук. – Минск : БГУ, 2011. – 463 с.
5. Айвазян, С. А. Прикладная статистика. Основы эконометрики. Том 1. Теория вероятностей и прикладная статистика / С. А. Айвазян, В. С. Мхитарян. – Москва : Юнити-Дана, 2001. – 656 с.
6. Anderson, T. W. An introduction to multivariate statistical analysis. 3rd Edition / T. W. Anderson. – N. Y. : Wiley, 2003. – 478 p.
7. Härdle, W. K. Applied Multivariate Statistical Analysis. 4th Edition / W. K. Härdle, L. Simar. – N.-Y. : Springer, 2015. – 457 p.

Перечень дополнительной литературы

1. Прикладная статистика. Основы моделирования и первичная обработка данных / С. А. Айвазян [и др.]. – М. : Финансы и статистика, 1983. – 459 с.
2. Прикладная статистика. Исследование зависимостей / С. А. Айвазян [и др.]. – М. : Финансы и статистика, 1985. – 487 с.
3. Прикладная статистика. Классификация и снижение размерности / С. А. Айвазян [и др.]. – М. : Финансы и статистика, 1989. – 607 с.
4. Малюгин, В. И. Методы анализа многомерных эконометрических моделей с неоднородной структурой. Научное издание / В. И. Малюгин. – Минск : БГУ, 2014. – 351 с.
5. Малюгин, В. И. Компьютерный анализ данных и моделирование на IBM SPSS. Учебно-методические указания по лабораторному практикуму «Компьютерный анализ данных и моделирование» / В. И. Малюгин, Н. В. Гринь. – Минск : БГУ, 2013. – 42 с.
6. Han, J. Data Mining: Concepts and Techniques. 2nd ed / J. Han, M. Kamber. – Morgan Kaufmann, 2006. – 512 p.
7. Chambers, J. M. Programming with R. Software for Data Analysis / J. M. Chambers. – Springer, 2008. – 387 p.

8. Venables, W. N., Smith D. M. An Introduction to R / W. N. Venables, D. M. Smith. – N.Y., 2013. – 408 p.
9. Система статистических кредитных рейтингов предприятий: методика построения, верификации и применения / В. И. Малюгин [и др.] // Банковский Вестник. Исследования банка. – № 5, 2013. – 73 с.
10. Malugin, V. I. Statistical analysis and econometric modelling of the creditworthiness of non-financial companies / V. I. Malugin, N. V. Hryn, A. Y. Novopolitsev // Int. J. Computational Economics and Econometrics. – 2014. – Vol. 4(1/2). – P. 130.
11. Малюгин, В. И. Об эффективности статистических алгоритмов кредитного скоринга / В. И. Малюгин, Н. В. Гринь // Банковский Вестник – № 4 – 2010. – С. 39-46.
12. Малюгин, В. И. Об оптимальности классификации случайных наблюдений, различающихся уравнениями регрессии / В. И. Малюгин, Ю. С. Харин // Автоматика и телемеханика. – 1986. – № 7. – С. 35-46.
13. Гринь, Н. В. Исследование точности методов классификации многомерных данных в задачах кредитного скоринга / Н. В. Гринь, В. И. Малюгин // Вестник ГрГУ. – Сер. 2. – 2008. – № 1. – С. 77-85.

Примерный перечень тем для коллоквиумов

- 1) Статистическое оценивание параметров многомерной гауссовской модели.

Рекомендуемая тематика контрольных работ

- 1) Контрольная работа №1. *Многомерное нормальное (гауссовское) распределение как модель многомерных данных.*
- 2) Контрольная работа №2. *Статистическая проверка гипотез на основе T2-статистики Хотеллинга.*

Методические рекомендации по организации самостоятельной работы обучающихся

Для организации самостоятельной работы студентов по учебной дисциплине следует использовать современные информационные технологии: разместить в сетевом доступе комплекс учебных и учебно-методических материалов (учебно-программные материалы, ссылки на учебные издания для теоретического изучения дисциплины, методические указания к лабораторным занятиям, материалы текущего контроля и текущей аттестации, позволяющие определить соответствие учебной деятельности обучающихся требованиям образовательных стандартов высшего образования и учебно-программной документации, в т.ч. вопросы для подготовки к зачету, задания, тесты, вопросы для самоконтроля, тематика рефератов и др., список рекомендуемой литературы, информационных

ресурсов и др.). Эффективность самостоятельной работы студентов проверяется в ходе текущего и итогового контроля знаний. Для общей оценки качества усвоения студентами учебного материала рекомендуется использование рейтинговой системы.

Примерный перечень заданий для управляемой самостоятельной работы:

1. Многомерное нормальное (гауссовское) распределение как модель многомерных данных [1-6];
2. Статистическое оценивание параметров многомерной гауссовской модели [1-6];
3. Методы предварительного анализа данных [1, 3, 7].
4. Методы исследования зависимостей [1, 3, 5, 6, 9].
5. Методы статистической классификации [1, 2, 10, 12].
6. Примеры практического применения методов статистического анализа данных в экономике [11, 16-20].
7. Статистический анализ данных с использованием языка R [14, 15].

Перечень рекомендуемых средств диагностики

На лекционных занятиях по дисциплине «Статистический анализ данных» рекомендуется особое внимание обращать внимание на установлении связей между теоретическим Темами курса и использованием, изучаемых методов и алгоритмов для решения практических задач анализа данных.

Целесообразно использовать форму контроля самостоятельной работы студентов в виде отчетов по заданиям практикума. Условия для самостоятельной работы студентов, в частности, для развития навыков самоконтроля, способствующих интенсификации учебного процесса, обеспечиваются наличием и полной доступностью электронных (и бумажных) вариантов курсов лекций, учебно-методических пособий, монографий и сборников задач по основным темам дисциплины.

Текущий контроль усвоения знаний в течение семестра рекомендуется осуществлять в виде проведения коллоквиума. Для закрепления и проверки знаний и умений студентов (практическая часть курса) рекомендуется проведение компьютерных тестовых занятий, а также защиту отчетов по основным темам дисциплины.

Успеваемость студентов в рамках дисциплины рекомендуется оценивать в конце семестра в форме экзамена, основываясь на текущей оценке успеваемости.

Для диагностики компетенций в рамках учебной дисциплины рекомендуется использовать следующие формы:

– устная форма: собеседования, устные промежуточные и итоговый зачеты.

- письменная форма: тесты, контрольные опросы, контрольная работа.
- устно-письменная форма: отчеты по домашним практическим упражнениям с их устной защитой.

Контрольные мероприятия проводятся в соответствии с учебно-методической картой дисциплины. В случае неявки на контрольное мероприятие по уважительной причине студент вправе по согласованию с преподавателем выполнить его в дополнительное время. Для студентов, получивших неудовлетворительные оценки за контрольные мероприятия, либо не явившихся по неуважительной причине, по согласованию с преподавателем и с разрешения заведующего кафедрой мероприятие может быть проведено повторно.

Методика формирования итоговой оценки

Формой текущей аттестации по учебной дисциплине «Статистический анализ данных» учебным планом предусмотрен экзамен.

При оценивании реферата внимание обращается на:

- содержание, корректность и последовательность изложения – 35%;
- релевантность и полноту раскрытие темы – 20 %;
- самостоятельность суждений – 35%;
- оформление – 10%.

Рекомендуется использовать рейтинговую оценку знаний студента, дающую возможность проследить и оценить динамику процесса достижения целей обучения. Рейтинговая оценка предусматривает использование весовых коэффициентов для текущего контроля знаний и текущей аттестации студентов по дисциплине. Примерные весовые коэффициенты, определяющие вклад текущего контроля знаний в рейтинговую оценку:

- подготовка реферата – 15 %;
- работа на лабораторных занятиях – 35 %;
- контрольные работы – 30 %;
- коллоквиум – 20 %.

Итоговая оценка формируется на основе:

- 1) Правил проведения аттестации студентов (Постановление Министерства образования Республики Беларусь № 53 от 29 мая 2012г.);
- 2) Положение о рейтинговой системе оценки знаний по дисциплине в БГУ (Приказ ректора БГУ от 18.08.2015 № 382-ОД);
- 3) Критериев оценки знаний студентов (письмо Министерства образования от 22.12.2003).

ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ УВО

Название учебной дисциплины, с которой требуется согласование	Название кафедры	Предложения об изменениях в содержании учебной программы учреждения высшего образования по учебной дисциплине	Решение, принятое кафедрой, разработавшей учебную программу (с указанием даты и номера протокола)
Теория вероятностей и математическая статистика	Математического моделирования и анализа данных	нет	Оставить содержание учебной дисциплины без изменения, протокол № 13 от 29.03.2018 г.
Имитационное и статистическое моделирование	Математического моделирования и анализа данных	нет	Оставить содержание учебной дисциплины без изменения, протокол № 13 от 29.03.2018 г.

**ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ ПО
ИЗУЧАЕМОЙ УЧЕБНОЙ ДИСЦИПЛИНЕ**
на ____ / ____ учебный год

№ п/п	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры
_____ (протокол № ____ от _____ 20__ г.)

Заведующий кафедрой

УТВЕРЖДАЮ
Декан факультета
