

Белорусский государственный университет

УТВЕРЖДАЮ

Проректор по учебной работе и
образовательным инновациям

О.И.Чуприс

« 09 » января 2018 г.

Регистрационный № УД-6395/уч.

МЕТОДЫ СТАТИСТИЧЕСКОГО АНАЛИЗА СЛОЖНЫХ ДАННЫХ

**Учебная программа учреждения высшего образования
по учебной дисциплине для специальности второй ступени
высшего образования (магистратуры):**

1-31 81 12 Прикладной компьютерный анализ данных

2018 г.

Учебная программа составлена на основе образовательного стандарта высшего образования ОСВО 1-31 81 12-2015 и учебного плана УВО № G31-251/уч. от 26.05.2017 г.

СОСТАВИТЕЛЬ:

А.Ю. Харин, заведующий кафедрой теории вероятностей и математической статистики Белорусского государственного университета, кандидат физико-математических наук, доцент

РЕЦЕНЗЕНТЫ:

Б.А. Залесский, заведующий лабораторией обработки и распознавания изображений Объединенного института проблем информатики Национальной академии наук Беларуси, доктор физико-математических наук;

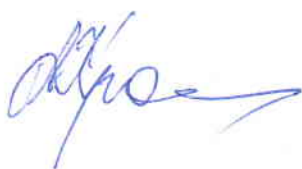
В.В. Мушко, доцент кафедры технологий программирования Белорусского государственного университета, кандидат физико-математических наук

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой теории вероятностей и математической статистики Белорусского государственного университета

(протокол № 5 от 21.11.2017 г.);

Научно-методическим Советом Белорусского государственного университета (протокол № 3 от 19.12.2017 г.)



ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Цель учебной дисциплины «Методы статистического анализа сложных данных» – ознакомление студентов II ступени высшего образования (магистрантов) с основными современными вероятностными моделями, методами и алгоритмами статистического исследования данных, относящихся в категорию сложных.

При изложении учебной дисциплины важно уделить внимание компьютерной реализации основных методов изучаемой теории.

В рамках поставленной цели **задачи учебной дисциплины** состоят в следующем:

1. Изучение теоретических основ – математических моделей и методов статистического анализа данных сложной структуры;
2. Формирование практических навыков решения прикладных задач анализа сложных данных с использованием свободно доступного современного программного обеспечения в области статистического анализа.

Учебная дисциплина «Методы статистического анализа сложных данных» относится к государственному компоненту цикла дисциплин специальной подготовки.

Основой для изучения дисциплины «Методы статистического анализа сложных данных» является дисциплина первой ступени «Теория вероятностей и математическая статистика» (или «Высшая математика» с включением соответствующих разделов), а также изучаемая ранее дисциплина второй ступени «Методы статистического анализа многомерных данных». Кроме того, лабораторный практикум предполагает предварительное изучение курса «Основы компьютерного анализа данных с использованием языка R». Дисциплина «Методы статистического анализа сложных данных» будет полезна для изучения дисциплины «Последовательный статистический анализ и мониторинг потоков данных», а также способствует успешному прохождению практики и написанию магистерских диссертаций.

В результате освоения учебной дисциплины студент магистратуры должен:

знать:

- современные вероятностные модели сложных данных, возникающие при решении прикладных задач;
- эффективные статистические методы анализа моделей сложных данных;
- правила и свойства применения указанных методов;

уметь:

- подбирать подходящую модель для решения прикладных задач статистического анализа данных;
- оценивать потенциальную эффективность применения конкретного статистического метода для решения задачи анализа данных;

владеть:

- знаниями основных моделей сложных статистических данных;
- навыками выбора и обоснования модели при решении конкретной задачи;
- умениями компьютерной реализации основных методов решения задач.

Освоение учебной дисциплины «Методы статистического анализа сложных данных» должно обеспечить формирование следующих академических, социально-личностных и профессиональных компетенций:

академические компетенции:

АК-1 Способность к самостоятельной профессиональной деятельности (анализ, сопоставление, систематизация, абстрагирование, моделирование, проверка достоверности данных, принятие решений и др.), готовность генерировать и использовать новые идеи.

социально-личностные компетенции:

СЛК-1 Учитывать социальные и нравственно-этические нормы в социально-профессиональной деятельности.

СЛК-2 Быть способным к сотрудничеству и работе в команде.

СЛК-3 Владеть коммуникативными способностями для работы в междисциплинарной и международной среде.

профессиональные компетенции:

ПК-1 Квалифицированно использовать современные достижения по разработке и анализу математических моделей, методов компьютерного анализа данных и современные информационные технологии.

ПК-3 Самостоятельно разрабатывать эффективные численные методы и алгоритмы, а также интегрировать их в компьютерные системы анализа данных.

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

Дисциплина изучается во втором семестре (II ступень). Всего на изучение учебной дисциплины «Методы статистического анализа сложных данных» отведено 168 часов, в том числе 56 аудиторных часов, из них: лекции – 20 часов, семинарские занятия – 18 часов, лабораторные занятия – 18 часов.

Трудоемкость учебной дисциплины составляет 4 зачетные единицы.

Форма текущей аттестации – зачет и экзамен.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Раздел 1. Некоторые современные модели сложных данных

Тема 1.1. Особенности анализа сложных данных, типовые модели.

Введение. Примеры современных прикладных задач, показывающие принципиальные сложности статистического анализа данных классическими методами. Классификация современных моделей сложных данных.

Раздел 2. Байесовский подход в статистическом анализе

Тема 2.1. Байесовское оценивание параметров и прогнозирование.

Схема байесовского подхода в статистике. Байесовское оценивание параметров. Байесовское прогнозирование временных рядов с трендом и авторегрессионных временных рядов.

Раздел 3. Последовательный статистический анализ

Тема 3.1. Последовательный подход в статистике, проверка гипотез.

Построение последовательного критерия отношения правдоподобия. Анализ его характеристик. Обобщение на случай данных, образующих цепи Маркова. Последовательные тесты проверки сложных гипотез. Применение в медицине.

Раздел 4. Анализ данных с пропусками

Тема 4.1. Модели пропусков, анализ данных с пропусками.

Различные модели пропусков в данных. Цензурирование. ЕМ-алгоритм и его свойства.

Раздел 5. Анализ лонгитюдных и кросс-секционных данных

Тема 5.1. Кросс-секционные и лонгитюдные данные.

Примеры кросс-секционных и длительных статистических исследований. Панели. Когортный анализ.

Тема 5.2. Модели с фиксированными эффектами, модели со случайными эффектами.

Оценивание параметров, анализ точности.

Раздел 6. Полупараметрические и непараметрические модели

Тема 6.1. Полупараметрические и непараметрические модели.

Полупараметрические модели (сплайны). Модель ядерной регрессии.

Раздел 7. Робастные статистические выводы

Тема 7.1. Понятие робастности.

Классификация типов искажений. Характеристики робастности.

Тема 7.2. Робастность оценивания параметров.

Методы робастного статистического оценивания параметров модели.

Тема 7.3. Робастность проверки гипотез и прогнозирования.

Робастная статистическая проверка гипотез. Робастность в статистическом прогнозировании. Заключение.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов					Количество часов УСР	Форма контроля знаний
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия	Иное		
1	2	3	4	5	6	7	8	9
1	Некоторые современные модели сложных данных	2						
1.1	Особенности анализа сложных данных, типовые модели	2						Устный опрос
2	Байесовский подход в статистическом анализе	2		2	2			
2.1	Байесовское оценивание параметров и прогнозирование	2		2	2			Отчет по лабораторной работе №1
3	Последовательный статистический анализ	2		2	2			
3.1	Последовательный подход в статистике, проверка гипотез	2		2	2			Контрольная работа №1
4.	Анализ данных с пропусками	2		2	2			
4.1	Модели пропусков, анализ данных с пропусками	2		2	2			Командная игра для творческого решения заданий открытого типа
5	Анализ лонгитюдных и кросс-секционных данных	4		4	4			
5.1	Кросс-секционные и лонгитюдные данные	2		2	2			Коллоквиум

5.2	Модели с фиксированными эффектами, модели со случайными эффектами	2		2	2			Контрольная работа №2
6	Полупараметрические и непараметрические модели	2		2	2			
6.1	Полупараметрические и непараметрические модели	2		2	2			Командная игра для творческого решения заданий открытого типа
7	Робастные статистические выводы	6		6	6			
7.1	Понятие робастности	2		2	2			Устный опрос
7.2	Робастность оценивания параметров	2		2	2			Устный опрос
7.3	Робастность проверки гипотез и прогнозирования	2		2	2			Отчет по лабораторной работе №2
	Всего: 56 часов	20		18	18			

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Перечень основной литературы

1. Харин, А.Ю. Робастность байесовских и последовательных статистических решающих правил / А.Ю. Харин. – Минск : БГУ. – 2014. – 207 с.
2. Johnson, R.A. Applied multivariate statistical analysis / R.A. Johnson, D.W. Wichern. – New York : Pearson. – 2017. – 800 p.
3. Харин, Ю.С. Эконометрическое моделирование / Ю.С. Харин, В.И. Малюгин, А.Ю. Харин. – Минск : БГУ. – 2004. – 313 с.
4. Hastie, T. The elements of statistical learning / T. Hastie, R. Tibshirani, J. Friedman. – New York : Springer. – 2018. – 758 p.

Перечень дополнительной литературы

1. Харин, Ю.С. Теория вероятностей, математическая и прикладная статистика / Ю.С. Харин, Н.М. Зуев, Е.Е. Жук. – Минск : БГУ. – 2011. – 463 с.
2. Everitt, B. An introduction to applied multivariate analysis with R / B. Everitt, T. Hothorn. – New York: Springer. – 2018. – 274 p.
3. Харин, Ю.С. Математические и компьютерные основы статистического моделирования и анализа данных / Ю.С. Харин, В.И. Малюгин, М.С. Абрамович. – Минск : БГУ. – 2008. – 455 с.

Перечень рекомендуемых средств диагностики

Для текущего контроля качества усвоения знаний студентами используется следующий диагностический инструментарий:

- устные опросы;
- командная игра для творческого решения заданий открытого типа;
- проведение коллоквиума;
- контрольные работы;
- защита отчетов по лабораторным работам.

Методика формирования итоговой оценки

Итоговая оценка формируется на основе:

1. Правил проведения аттестации студентов (Постановление Министерства образования Республики Беларусь № 53 от 29 мая 2012 г.);
2. Положения 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 г.).

В течение семестра проводятся коллоквиум, две контрольные и две лабораторные работы. На выполнение коллоквиума отводится до 30 минут. Задания оцениваются в соответствии с их сложностью, максимальная сумма баллов за все задания в коллоквиуме равна 10.

Итоговая оценка по дисциплине рассчитывается как средневзвешенная оценка коллоквиума, двух контрольных и двух лабораторных работ. Весовой коэффициент для оценки коллоквиума – 0,2; для оценки каждой контрольной и лабораторной работы – 0,2.

Примерная тематика лабораторных работ

Лабораторная работа № 1. Компьютерная реализация байесовского оценивания параметров и прогнозирования.

Лабораторная работа № 2. Компьютерная реализация и анализ робастности статистических процедур, построение робастных статистических процедур.

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) ¹
Последовательный статистический анализ и мониторинг потоков данных	Теории вероятностей и математической статистики	Нет	Оставить содержание учебной дисциплины без изменения, протокол № 11 от 27.03.2018 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 _____ 201_ г.)
 (название кафедры)

Заведующий кафедрой

 (ученая степень, ученое звание) (подпись) (И.О.Фамилия)

УТВЕРЖДАЮ
 Декан факультета

 (ученая степень, ученое звание) (подпись) (И.О.Фамилия)