

**Белорусский государственный университет**

**УТВЕРЖДАЮ**  
Проректор по учебной работе



А.Л. Толстик

Регистрационный № УД- 1851 /уч.

**КОМПЬЮТЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ ДИСКРЕТНЫХ  
ДАНЫХ**

**Учебная программа учреждения высшего образования  
по учебной дисциплине для специальности:**

**1-98 01 01 Компьютерная безопасность (по направлениям)**

2016 г.

Учебная программа составлена на основе образовательного стандарта высшего образования ОСВО 1-98 01 01-2013 и учебного плана Р 98-138/уч., 30.05.2013.

**Составители:**

Е. Н. Орлова, доцент кафедры математического моделирования и анализа данных Белорусского государственного университета, кандидат физико-математических наук, доцент;

Ю.С. Харин, заведующий кафедрой 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 № 8 от 01 декабря 2015 г.);

Учебно-методической комиссией факультета прикладной математики и информатики Белорусского государственного университета (протокол № 3 от 17 декабря 2015 г.).

## ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

В связи с интенсивной компьютеризацией научных и практических исследований возрастает внимание к проблемам дискретной математики, и, как следствие, к дискретным моделям статистических данных.

Статистический анализ дискретных данных – раздел статистического анализа данных, в рамках которого анализируются данные, представленные дискретными вероятностными моделями, либо данные нечисловой природы.

Примерами объектов нечисловой природы являются:

- значения качественных признаков, то есть результаты кодировки объектов с помощью заданного перечня категорий (градаций);
- упорядочения (ранжировки), описывающие мнения экспертов;
- классификации, то есть разбиения совокупности объектов на группы сходных между собой (кластеры);
- толерантности, то есть бинарные отношения, описывающие сходство объектов между собой;
- результаты парных сравнений или контроля качества продукции по альтернативному признаку («годен»- « брак»), то есть последовательности из 0 и 1;
- слова, предложения, тексты;
- векторы, координаты которых – совокупность значений разнотипных признаков, то есть векторы, в которых часть признаков носит качественный характер, а часть – количественный.

Учебная дисциплина «Компьютерный статистический анализ дискретных данных» знакомит студентов с классическими и современными методами решения задач статистического анализа дискретных данных, к которым относятся описание данных и их первичная статистическая обработка, оценивание неизвестных характеристик и параметров, исследование статистических свойств полученных оценок, исследование взаимосвязей между показателями, проверка статистических гипотез.

**Целью** дисциплины «Компьютерный статистический анализ дискретных данных» является:

- научная и практическая подготовка студентов к применению дискретных математических моделей в технических и социально-экономических системах;
- изучение теоретических основ статистического анализа дискретных данных;
- развитие у студентов навыков исследования соответствующих дискретных вероятностных моделей, а также дискретизации непрерывных моделей;
- формирование практических навыков, решения типовых задач анализа дискретных данных с помощью статистического программного обеспечения;

– формирование у студентов профессиональных знаний в предметной области, подкрепленных владением математическими методами и пакетами прикладных программ.

**Задачами** дисциплины «Компьютерный статистический анализ дискретных данных» являются:

– формирование у студентов навыков построения моделей с использованием дискретной математики и теории вероятностей;

– систематическое изучение теоретических основ статистического анализа дискретных данных;

– приобретение практических навыков для правильного использования арсенала методов статистического анализа дискретных данных в конкретных ситуациях.

В первом разделе дисциплины «Компьютерный статистический анализ дискретных данных» приводятся сведения о математических моделях дискретных данных, о шкалах измерения и об основных дискретных распределениях вероятностей.

Во втором разделе рассматривается полиномиальная модель дискретных данных. Приводятся основные числовые и функциональные характеристики полиномиального распределения вероятностей. Доказывается асимптотическая нормальность вектора абсолютных частот исходов, рассматриваются задачи статистического оценивания параметров и параметрических функций для полиномиальной модели. Рассматриваются также задачи статистической проверки гипотез относительно параметров полиномиального распределения. Приводятся критерии однородности и независимости.

Третий раздел посвящен вопросам статистического анализа пуассоновской модели дискретных данных. Рассматриваются задачи оценивания параметра и параметрических функций пуассоновского распределения вероятностей и задачи проверки статистических гипотез.

В четвертом разделе приводятся сведения методах статистического анализа нечисловых данных.

Пятый раздел посвящен вопросам статистического анализа дискретных временных рядов. Особое внимание уделяется однородным и связным цепям Маркова.

Учебная дисциплина «Компьютерный статистический анализ дискретных данных» относится к циклу дисциплин вузовского компонента и взаимосвязана с учебными дисциплинами «Теория вероятностей и математическая статистика» и «Статистический анализ данных». Методы, излагаемые в дисциплине «Компьютерный статистический анализ дискретных данных», могут быть использованы при изучении ряда других дисциплин по специальности «Компьютерная безопасность».

В результате изучения дисциплины студент должен

**знать:**

- проблематику, основные термины, базовые понятия и методологию статистического анализа дискретных данных;
- основные дискретные вероятностные модели данных;
- методы выявления зависимости между признаками, измеренными в нечисловых шкалах;

**уметь:**

- осуществлять предварительный статистический анализ дискретных данных с целью установления модели данных;
- проводить статистический анализ дискретных данных с целью установления статистических зависимостей;
- осуществлять классификацию неоднородных данных;

**владеть:**

- методами решения основных задач статистического анализа дискретных данных;
- навыками по подготовке данных и решения типовых задач статистического анализа дискретных данных;
- навыками применения современных ППП для решения задач статистического анализа дискретных данных в различных приложениях.

В соответствии с образовательным стандартом специальности 1-98 01 01 «Компьютерная безопасность» учебная программа предусматривает для изучения дисциплины всего 132 часа, из них 68 аудиторных часов, в том числе лекций – 34 часа, лабораторных занятий – 30 часов, УСП – 4 часов (3 курс, 6 семестр).

Форма текущей аттестации по учебной дисциплине – зачет.

## СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

### Раздел I. Математические модели дискретных данных

**Тема 1.1. Шкалы измерения статистических данных.** Количественная шкала. Номинальная шкала. Порядковая шкала. Примеры.

**Тема 1.2. Дискретные модели статистических данных.** Дискретные распределения вероятностей и их свойства. Нечисловые данные. Примеры.

### Раздел II. Полиномиальная модель дискретных данных

**Тема 2.1. Полиномиальное распределение вероятностей.** Полиномиальное распределение как модель многомерных дискретных данных. Числовые характеристики полиномиального распределения. Ковариационная матрица. Производящая и характеристическая функции.

**Тема 2.2. Полиномиальная статистическая модель.** Независимые однородные испытания с конечным множеством исходов. Вектор абсолютных частот исходов, его свойства и распределение вероятностей.

**Тема 2.3. Асимптотический анализ полиномиальной статистической модели.** Асимптотическая нормальность вектора абсолютных частот исходов. Асимптотическая нормальность линейного преобразования вектора абсолютных частот. Маргинальные распределения вероятностей вектора абсолютных частот исходов.

**Тема 2.4. Оптимальные оценки неизвестных параметров полиномиальной модели.** Критерий оптимальности оценок параметров и параметрических функций. Понятие оцениваемой функции. Оцениваемые функции для полиномиальной модели. Эмпирическая функция распределения для полиномиальной модели.

**Тема 2.5. Оценки максимального правдоподобия параметров и параметрических функций в полиномиальной модели.** МП-оценка вектора вероятностей. Существование и свойства МП-оценок параметрических функций.

**Тема 2.6. Доверительное оценивание параметров и параметрических функций полиномиальной модели.** Построение доверительного интервала для параметрической функции. Методы построения доверительных областей для вектора вероятностей.

**Тема 2.7. Проверка статистических гипотез относительно параметров полиномиальной модели.** Критерии хи-квадрат для простых и сложных гипотез. Критерий отношения правдоподобия. Критерий однородности хи-квадрат. Критерий независимости хи-квадрат.

### Раздел III. Пуассоновская модель дискретных данных

**Тема 3.1. Распределение Пуассона и его свойства.** Определение и свойства. Теорема Пуассона для биномиального распределения. Связь пуассоновского распределения с полиномиальным.

**Тема 3.2. Точечные и интервальные оценки неизвестных параметров и параметрических функций в пуассоновской модели.** Оценивание параметрических функций. МП-оценки параметра пуассоновской модели. Эффективность выборочного среднего. Информация Фишера. Доверительные интервалы для параметра.

**Тема 3.3. Проверка статистических гипотез относительно параметров пуассоновской модели.** Проверка гипотезы согласия для пуассоновской модели. Гипотеза однородности для пуассоновской модели. Критерий Неймана-Пирсона. Критерий отношения правдоподобия для простой гипотезы.

**Раздел IV. Статистический анализ данных, измеренных в порядковых и номинальных шкалах**

**Тема 4.1. Статистический анализ данных, измеренных в порядковых шкалах.** Понятие ранговой корреляции. Ранговые коэффициенты корреляции: свойства и применения. Коэффициент конкордации: свойства и применения.

**Тема 4.2. Статистический анализ данных, измеренных в номинальных шкалах.** Таблицы сопряженности. Основные показатели степени тесноты статистической зависимости номинальных признаков: свойства и применение.

**Раздел V. Статистический анализ дискретных временных рядов**

**Тема 5.1. Модели динамики дискретных данных.** Дискретные временные ряды. Цепи Маркова: основные понятия и свойства. Условия стационарности.

**Тема 5.2. Статистический анализ однородных цепей Маркова.** МП-оценки матрицы вероятностей одношаговых переходов и стационарного распределения вероятностей однородной цепи Маркова. Проверка гипотез о цепи Маркова.

**Тема 5.3. Статистический анализ связанных цепей Маркова.** Методы оценивания параметров связанных цепей Маркова. МТД-модели. Скрытые цепи Маркова и их статистический анализ.

## УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

№п/п	Название раздела, темы	Количество часов				Количество часов УСР	Форма контроля знаний
		Аудиторные					
		Лекции	Практ. и сем. занятия	Лаб. занятия	Иное		
<b>1</b>	<b>Математические модели дискретных данных</b>	<b>4</b>		<b>4</b>			
1.1	Шкалы измерения статистических данных	2		2			Опрос на занятии
1.2	Дискретные модели статистических данных	2		2			Опрос на занятии
<b>2</b>	<b>Полиномиальная модель дискретных данных</b>	<b>14</b>		<b>12</b>		<b>2</b>	
2.1	Полиномиальное распределение вероятностей	2				2	Тест
2.2	Полиномиальная статистическая модель	2		2			Опрос на занятии
2.3	Асимптотический анализ полиномиальной статистической модели	2		2			Опрос на занятии
2.4	Оптимальные оценки неизвестных параметров полиномиальной модели	2		2			Опрос на занятии
2.5	Оценки максимального правдоподобия параметров и параметрических функций в полиномиальной модели	2		2			Опрос на занятии
2.6	Доверительное оценивание параметров и параметрических функций полиномиальной модели	2		2			Опрос на занятии
2.7	Проверка статистических гипотез относительно параметров полиномиальной модели	2		2			Опрос на занятии
<b>3</b>	<b>Пуассоновская модель дискретных данных</b>	<b>6</b>		<b>4</b>		<b>2</b>	
3.1	Распределение Пуассона и его свойства	2				2	Тест
3.2	Точечные и интервальные оценки неизвестных параметров и параметрических функ-	2		2			Опрос на занятии



	ций в пуассоновской модели						
3.3	Проверка статистических гипотез относительно параметров пуассоновской модели	2		2			Опрос на занятии
<b>4</b>	<b>Статистический анализ данных, измеренных в порядковых и номинальных шкалах</b>	<b>4</b>		<b>4</b>			
4.1	Статистический анализ данных, измеренных в порядковых шкалах	2		2			Опрос на занятии
4.2	Статистический анализ данных, измеренных в номинальных шкалах	2		2			Опрос на занятии
<b>5</b>	<b>Статистический анализ дискретных временных рядов</b>	<b>6</b>		<b>6</b>			
5.1	Модели динамики дискретных данных	2		2			Опрос на занятии
5.2	Статистический анализ однородных цепей Маркова	2		2			Опрос на занятии
5.3	Статистический анализ однородных цепей Маркова	2		2			Опрос на занятии
<b>ИТОГО</b>		<b>34</b>		<b>30</b>		<b>4</b>	

## ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

### *Рекомендуемая литература*

#### *Основная*

1. Андерсон Т.В. Введение в многомерный статистический анализ. М., Физматгиз, 1963.
2. Андерсон Т.В. Статистический анализ временных рядов. М.: Мир, 1976.
3. Кендалл М.Дж., Стьюарт А. Многомерный статистический анализ и временные ряды. М.: Мир, 1976.
4. Харин Ю.С., Жук Е.Е. Математическая и прикладная статистика. Мн.: БГУ, 2005.
5. Бендат Дж., Пирсол А. Прикладной анализ случайных данных. Москва: Мир 1989.
6. Бриллинджер Д. Временные ряды. М.: Мир, 1990.
7. Харин Ю.С., Малюгин В.И., Абрамович М.С. Математические и компьютерные основы статистического моделирования и анализа данных. Мн.: БГУ, 2008.
8. Харин Ю.С., Абрамович М.С., Малюгин В.И. Компьютерный учебник по статистике. Мн.: БГУ, 2001.
9. Ли Ц., Джадж Д., Зельнер А. Оценивание параметров марковских моделей по агрегированным временным рядам. М.: Статистика, 1977.
10. Харин Ю.С., Степанова М.Д. Практикум на ЭВМ по математической статистике. Мн.: Университетское, 1987.

#### *Дополнительная*

1. Болч Б., Хуань К.Дж. Многомерные статистические методы для экономики. М.: Мир, 1979.
2. Афифи С.А., Эйзен. Статистический анализ. Подход с использованием ЭВМ. М.: Мир, 1982.
3. Айвазян С.А. и др. Прикладная статистика. Основы моделирования и первичная обработка данных. М.: Финансы и статистика, 1983.
4. Айвазян С.А. и др. Прикладная статистика. Исследование зависимостей. М.: Финансы и статистика, 1985.
5. Айвазян С.А. и др. Прикладная статистика. Классификация и снижение размерности. М.: Финансы и статистика, 1989.

6. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики, т. 1. Теория вероятностей и прикладная статистика, М.: ЮНИТИ, 2001.
7. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы. М.: Финансы и статистика, 1998.
8. Харин Ю.С., Степанова М.Д. Практикум на ЭВМ по математической статистике. - Мн.: Университетское, 1987.
9. Харин Ю.С. и др. Математические основы криптологии. – Мн.: БГУ, 1999
10. Харин Ю.С., Агиевич С.В. Компьютерный практикум по математическим методам защиты информации. – Мн.: БГУ, 2001.
11. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере. М.: ИНФРА-М, 1995.
12. Боровиков В.П. STATISTICA. Статистический анализ и обработка данных в среде Windows. М.: Филинь, 1997.
13. Многомерный статистический анализ в экономике / Под ред. В.Н. Тамашевича. М.: ЮНИТИ, 1999.
14. Кнут Д. Искусство программирования на ЭВМ. Т.2: Получисленные алгоритмы. - М.: Мир, 1977.
15. Ивченко Г.И., Медведев Ю.И. Введение в математическую статистику. – М.: URSS, 2009.
16. Ивченко Г.И., Глибоченко А.Ф., Иванов В.А., Медведев Ю.И. Статистический анализ дискретных последовательностей. – М.: МИЭМ, 1994.

## Организация управляемой самостоятельной работы студентов

Управляемая самостоятельная работа (УСР) студентов – это самостоятельная работа, выполняемая по заданию и при методическом руководстве преподавателя, а также контролируемая преподавателем на определенном этапе обучения. Целью УСР является целенаправленное обучение студентов основным навыкам и умению индивидуальной самостоятельной работы.

Результативность самостоятельной работы студентов во многом определяется наличием активных методов ее контроля:

- входной контроль знаний и умений студентов в начале изучения очередной дисциплины;
- текущий контроль, то есть регулярное отслеживание уровня усвоения материала на лекциях, практических и лабораторных занятиях;
- промежуточный контроль по окончании изучения раздела или модуля курса;
- самоконтроль, осуществляемый студентом в процессе изучения дисциплины при подготовке к контрольным мероприятиям;
- итоговый контроль по дисциплине в виде зачета или экзамена;
- контроль остаточных знаний и умений спустя определенное время после завершения изучения дисциплины.

На освоение учебного материала в рамках УСР для дисциплины «Компьютерный статистический анализ дискретных данных» отводится 4 аудиторных часа.

Для самостоятельного изучения в рамках УСР дисциплины «Компьютерный статистический анализ дискретных данных» выносятся следующие темы:

1. Полиномиальное распределение вероятностей [1 - 5];
2. Распределение Пуассона и его свойства [1 - 5].

Контроль осуществляется в виде тестирования.

## **Рекомендации по контролю качества усвоения знаний и проведению аттестации**

На лекционных занятиях по учебной дисциплине «Компьютерный статистический анализ дискретных данных» рекомендуется использовать элементы проблемного обучения: проблемное изложение некоторых аспектов, использование частично-поискового метода.

Для аттестации обучающихся на соответствие их персональных достижений поэтапным и конечным требованиям образовательной программы создаются фонды оценочных средств, включающие типовые задания, контрольные работы и тесты. Оценочными средствами предусматривается оценка способности обучающихся к творческой деятельности, их готовность вести поиск решения новых задач, связанных с недостаточностью конкретных специальных знаний и отсутствием общепринятых алгоритмов.

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

- устная форма: собеседование, устный промежуточный зачет, итоговый зачет;
- письменная форма: тест, контрольный опрос, контрольная работа;
- устно-письменная форма: отчет 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.

Оценка текущей успеваемости рассчитывается как среднее оценок за каждую из письменных контрольных работ, оценки за отчеты по домашним практическим упражнениям и оценки за итоговый тест.

Итоговая аттестация предусматривает проведение зачета. При этом рекомендуется использовать оценивание успеваемости на основе модульно-рейтинговой системы.

## ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ

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	Кафедра математического моделирования и анализа данных	нет	Оставить содержание учебной дисциплины без изменения, протокол № 8 от 01 декабря 2015 г.
Статистический анализ данных	Кафедра математического моделирования и анализа данных	нет	Оставить содержание учебной дисциплины без изменения, протокол № 8 от 01 декабря 2015 г.

**ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ**

на \_\_\_\_ / \_\_\_\_ учебный год

№№ Пп	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры дискретной математики и алгоритмики (протокол № \_\_\_\_ от \_\_\_\_\_ 201\_ г.)

Заведующий кафедрой

\_\_\_\_\_  
(ученая степень, звание)

\_\_\_\_\_  
(подпись)

\_\_\_\_\_  
(И.О. Фамилия)

УТВЕРЖДАЮ

Декан факультета

\_\_\_\_\_  
(ученая степень, звание)

\_\_\_\_\_  
(подпись)

\_\_\_\_\_  
(И.О.Фамилия)