**Санкт-Петербургский государственный университет**

**Р А Б О Ч А Я П Р О Г Р А М М А**

**УЧЕБНОЙ ДИСЦИПЛИНЫ**

Введение в регрессионный анализ

Regression Analysis

**Язык(и) обучения**

русский

Трудоемкость в зачетных единицах: 2

Регистрационный номер рабочей программы: 051770

**Раздел 1. Характеристики учебных занятий**

**1.1. Цели и задачи учебных занятий**

В настоящее время математические методы планирования эксперимента получили основательное теоретическое развитие и широко используются на практике. В частности, такие методы используются при анализе производственных процессов для изучения всевозможных зависимостей между величинами, влияющими на ход исследуемого процесса. Искомые зависимости могут быть выведены теоретически или же получены из экспериментальных данных. Раздел статистики, объединяющий практические методы исследования (регрессионной) зависимости между величинами по экспериментальным данным, получил название «регрессионный анализ». Целью регрессионного анализа является определение общего вида уравнения регрессии (модели), построение оценок неизвестных параметров, входящих в это уравнение, и проверка статистических гипотез об этих параметрах.

В первой части курса рассматриваются основные понятия и методы регрессионного анализа, подробно исследуются задачи построения оценок, в частности, для моделей неполных рангов, приводится ряд классических теоретических результатов. Вторая часть посвящена описанию результатов теории оптимального планирования экспериментов. Подробно излагаются классические результаты, касающиеся свойств информационных матриц, теорема Каратеодори, теорема эквивалентности Кифера-Вольфовица и ее следствия. Также приводится ряд новых результатов, основанных на применении теорем эквивалентности.

Цель курса «Введение в регрессионный анализ»– ознакомить студентов с основными понятиями и методами регрессионного анализа в объеме, указанном в разделе 2.2, выработать навыки решения конкретных прикладных задач. В плане практической работы, акцент делается на выработку у студентов навыков самостоятельного написания алгоритмов на языке R и в мат. пакетах Maple и MatLab.

**1.2. Требования подготовленности обучающегося к освоению содержания учебных занятий (пререквизиты)**

Дисциплина изучается обучающимися бакалавриата по направлению подготовки 02.03.01 «Математика и компьютерные науки» в пятом семестре обучения. При чтении курса, предполагается, что студенты обладают базовыми знаниями в области математического и функционального анализа, алгебры, численных методов в объеме курсов бакалаврских программ, читаемых на математико-механическом факультете СПбГУ.

**1.3. Перечень результатов обучения (learning outcomes)**

В результате освоения дисциплины у обучающихся должно выработаться умение использовать математические пакеты и языки программирования для решения конкретных математических задач в объеме, указанном в разделе 2.2.

Кроме того, в результате обучения обучающиеся должны:

аргументировано, логически верно и содержательно ясно строить устную и письменную речь, способен использовать навыки публичной речи, ведения дискуссии и полемики;

знать и уметь учитывать основные тенденции развития современного естествознания;

обладать способностью активно применять общенаучные базовые знания естественных наук, математики и информатики в области прикладной математики и информатики;

обладать способностью работы с информацией из различных источников, включая сетевые ресурсы сети Интернет, для решения профессиональных задач;

обладать способностью применять методы прикладной математики и информатики, для решения задач производственной и технологической деятельности;  
обладать способностью применять в профессиональной деятельности современные языки программирования и языки баз данных, операционные системы, электронные библиотеки и пакеты программ, сетевые технологии;

знать и уметь использовать современные стохастические методы вычислений; иметь навыки построения статистических моделей и анализа данных.

**1.4. Перечень и объём активных и интерактивных форм учебных занятий**

Самостоятельная работа: индивидуальная работа с доступной литературой по программированию на MatLab и R, а также специальной математической литературой (см. раздел 3.4.2) для лучшего освоения дисциплины, её углублённого изучения и удовлетворения личных познавательных потребностей.

**Раздел 2. Организация, структура и содержание учебных занятий**

**2.1. Организация учебных занятий**

**2.1.1 Основной курс**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Трудоёмкость, объёмы учебной работы и наполняемость групп обучающихся | | | | | | | | | | | | | | | | | | |
| Код модуля в составе дисциплины,  практики и т.п. | Контактная работа обучающихся с преподавателем | | | | | | | | | | | | Самостоятельная работа | | | | Объём активных и интерактивных  форм учебных занятий | Трудоёмкость |
| лекции | семинары | консультации | практические  занятия | лабораторные работы | контрольные работы | коллоквиумы | текущий контроль | промежуточная  аттестация | итоговая аттестация | под руководством преподавателя | в присутствии  преподавателя | сам. раб. с использованием  методических материалов | текущий контроль (сам.раб.) | промежуточная аттестация (сам.раб.) | итоговая аттестация  (сам.раб.) |
| ОСНОВНАЯ ТРАЕКТОРИЯ | | | | | | | | | | | | | | | | | | |
| Форма обучения: очная | | | | | | | | | | | | | | | | | | |
| Семестр 5 | 30 |  |  |  |  |  |  |  | 2 |  |  |  | 30 |  | 10 |  | 10 | 2 |
|  | 2-100 |  |  |  |  |  |  |  | 10-25 |  |  |  | 1-1 |  | 1-1 |  |  |  |
| ИТОГО | 30 |  |  |  |  |  |  |  | 2 |  |  |  | 30 |  | 10 |  |  | 2 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Виды, формы и сроки текущего контроля успеваемости и промежуточной аттестации | | | | | | |
| Код модуля в составе дисциплины, практики и т.п. | Формы текущего контроля успеваемости | | Виды промежуточной аттестации | | Виды итоговой аттестации  (только для программ итоговой аттестации и дополнительных образовательных программ) | |
| Формы | Сроки | Виды | Сроки | Виды | Сроки |
| ОСНОВНАЯ ТРАЕКТОРИЯ | | | | | | |
| Форма обучения: очная | | | | | | |
| Семестр 5 |  |  | зачёт, устно, традиционная форма | по графику промежуточной аттестации |  |  |

**2.2. Структура и содержание учебных занятий**

Период обучения (модуль): **Семестр 5**

|  |  |  |  |
| --- | --- | --- | --- |
| **№ п/п** | **Наименование темы (раздела, части)** | **Вид учебных занятий** | **Кол-во часов** |
| 1 | **Тема 1.** Описательная регрессия.  Математическая модель и предпосылки регрессионного анализа. | лекции | 2 |
| сам. раб. с использованием  методических материалов | 2 |
| 2 | **Тема 2.** Теорема Гаусса-Маркова. МНК-оценка в вырожденном случае. | лекции | 2 |
| сам. раб. с использованием  методических материалов | 2 |
| 3 | **Тема 3.** Оценка максимального правдоподобия. Обобщенный метод наименьших квадратов. | лекции | 2 |
| сам. раб. с использованием  методических материалов | 2 |
| 4 | **Тема 4.** Модели с априорной информацией о параметрах. | лекции | 4 |
| сам. раб. с использованием  методических материалов | 4 |
| 5 | **Тема 5.** Основы дисперсионного анализа. | лекции | 4 |
| сам. раб. с использованием  методических материалов | 4 |
| 6 | **Тема 6.** Основные понятия теории оптимального планирования эксперимента. Теорема Каратеодори. | лекции | 4 |
| сам. раб. с использованием  методических материалов | 4 |
| 7 | **Тема 7.** Критерии оптимальности. Теоремы эквивалентности. | лекции | 4 |
| сам. раб. с использованием  методических материалов | 4 |
| 8 | **Тема 8.** Применение теорем эквивалентности. | лекции | 4 |
| сам. раб. с использованием  методических материалов | 4 |
| 9 | **Тема 9.** Факторные планы. | лекции | 4 |
| сам. раб. с использованием  методических материалов | 4 |
| 10 | Промежуточная аттестация | самостоятельная работа | 10 |
| зачёт | 2 |
| **Итого** | | | **72** |

**Тема 1.** *Описательная регрессия. Математическая модель и предпосылки регрессионного анализа*

• Постановка задачи. Различные подходы к оценке параметра. Лемма о разрешимости нормального уравнения.

• Классическая регрессионная модель. Стандартные предположения об ошибках. Наилучшая линейная несмещенная оценка.

**Тема 2.** *Теорема Гаусса-Маркова. МНК-оценка в вырожденном случае.*

• МНК-оценка. Вид дисперсионной матрицы линейной несмещенной оценки. Примеры. Критерий оценивания векторной функции параметров.

• Обобщенно-обратные матрицы. Теорема о свойствах обобщенно-обратной матрицы. Примеры.

**Тема 3.** *Оценка максимального правдоподобия. Обобщенный метод наименьших квадратов.*

• Определение. Связь между МНК-оценкой и оценкой максимального правдоподобия.

• Обобщенная линейная регрессионная модель. Обобщенная МНК-оценка.

**Тема 4.** *Модели с априорной информацией о параметрах.*

• Постановка задачи и классификация моделей. Линейные ограничения на параметры. Линейные ограничения со случайной ошибкой. Ограничения по области.

• Минимаксная, байесовская и гребневая оценки.

**Тема 5.**   *Основы дисперсионного анализа.*

• Объясненная и необъясненная доли дисперсии. Остаточная сумма квадратов.

• Несмещенная оценка для дисперсии. Проверка гипотез о значении параметров.

**Тема 6.** *Основные понятия теории оптимального планирования эксперимента. Теорема Каратеодори.*

• Понятие дискретного и непрерывного плана эксперимента. Стандартные допущения об ошибках и свойствах регрессионной модели.

• Информационная матрица плана. Свойства информационной матрицы. Ограничение сверху для числа точек оптимального плана.

**Тема 7.** *Критерии оптимальности. Теоремы эквивалентности.*

• Основные критерии оптимальности. Теорема Кифера-Вольфовица

• Аналоги теоремы Кифера-Вольфовица для различных критериев оптимальности.

**Тема 8.** *Применение теорем эквивалентности.*

• Явный вид e\_k-оптимальных планов.

• Явный вид L-оптимальных планов.

**Тема 9.** *Факторные планы.*

• МНК-оценка параметров регрессии для измерений, проводимых по полному двухуровневому факторному плану.

**Раздел 3. Обеспечение учебных занятий**

**3.1. Методическое обеспечение**

**3.1.1 Методические указания по освоению дисциплины**

Успешное освоение дисциплины возможно благодаря посещению лекций, участию в обсуждении вопросов, подготовленных к занятию, самостоятельной работе, включающей в себя чтение специальной литературы по разделам темы.

По курсу «Введение в регрессионный анализ» предусмотрено проведение лекционных и семинарских занятий. Все студенты должны быть обеспечены методическими материалами, рекомендованными по курсу. Методические материалы включают в себя следующие типы материалов — учебники, учебные пособия, методические указания для студентов, интернет-ресурсы, электронные учебные пособия, с опорой на которые проводится аудиторная работа.

**3.1.2 Методическое обеспечение самостоятельной работы**

При выполнении домашних, индивидуальных и контрольных заданий студенту необходимо знать содержание лекционных занятий, уметь применять методы, рассмотренные в процессе обучения. При подготовке к самостоятельной работе целесообразно использовать рекомендованную литературу, а также компьютеры с соответствующим программным обеспечением (пакеты Maple, MatLab и R).

**3.1.3 Методика проведения текущего контроля успеваемости и промежуточной аттестации и критерии оценивания**

В процессе обучения обучающиеся выполняют индивидуальные работы и разрабатывают альтернативные алгоритмы решения задач, обсуждаемых на лекциях. По окончании курса проводится проверочная работа (решение задач), а, также, зачет по теории.

Методика проведения зачета

Зачет проводится в устной и письменной форме. Преподаватели имеют набор практических и теоретических заданий для проведения зачета. Зачет выставляется по итогам проведения промежуточной аттестации.

Использование конспектов и учебников, а также электронных устройств хранения, обработки или передачи информации при подготовке и ответе на вопросы зачета не разрешается. В случае обнаружения факта использования недозволенных материалов (устройств) составляется акт, и студент удаляется с зачета.

*Критерий оценивания*.

Зачет ставится за сданные в течение семестра практические задачи (не менее 7 из 10), успешно написанную проверочную работу и по итогам устного опроса в конце семестра. Зачет получают студенты, выполнившие не менее 7 заданий и правильно ответившие на три из пяти теоретических вопроса. (Примерный список вопросов по теории см. в пункте 2.2.)

**Критерии оценивания по ECTS:**

Оценка A:

Студент выполнил 10 задач, выдаваемых в течение семестра, успешно написал проверочную работу\* в конце семестра и правильно ответил на 5 теоретических вопросов на финальном зачете.

Оценка B:

Студент выполнил 9 задач, выдаваемых в течение семестра, успешно написал проверочную работу\* в конце семестра и правильно ответил на 4 теоретических вопроса из 5.

Оценка C:

Студент выполнил 8 задач, выдаваемых в течение семестра, успешно написал проверочную работу\* в конце семестра и правильно ответил на 4 теоретических вопроса из 5.

Оценка D:

Студент выполнил 8 задач, выдаваемых в течение семестра, успешно написал проверочную работу\* в конце семестра и правильно ответил на 3 теоретических вопроса из 5.

Оценка E:

Студент выполнил не менее 7 задач, выдаваемых в течение семестра, удовлетворительно написал проверочную работу\* в конце семестра и правильно ответил на 3 теоретических вопроса из 5.

Оценка F:

Студент выполнил менее 7 задач, выдаваемых в течение семестра, и(или) плохо (не) написал проверочную работу\* в конце семестра и (или) неправильно ответил на 3 теоретических вопроса из 5.

\*Для зачета необходимо получить за проверочную работу оценку выше F. Оценка за зачет в целом, не может быть выше, чем оценка за проверочную работу.

**3.1.4 Методические материалы для проведения текущего контроля успеваемости и промежуточной аттестации (контрольно-измерительные материалы, оценочные средства)**

Преподаватели имеют набор контрольных заданий и тестов для контроля успеваемости студентов. Примерный список вопросов к зачету состоит из тем, перечисленных в разделе 2.2.

**Тема 1.** Описательная регрессия. Математическая модель и предпосылки регрессионного анализа.

**Тема 2.** Теорема Гаусса-Маркова. МНК-оценка в вырожденном случае.

**Тема 3.** Оценка максимального правдоподобия. Обобщенный метод наименьших квадратов.

**Тема 4.** Модели с априорной информацией о параметрах.

**Тема 5.** Основы дисперсионного анализа.

**Тема 6.** Основные понятия теории оптимального планирования эксперимента. Теорема Каратеодори.

**Тема 7.** Критерии оптимальности. Теоремы эквивалентности.

**Тема 8.** Применение теорем эквивалентности.

**Тема 9.** Факторные планы.

**3.1.5 Методические материалы для оценки обучающимися содержания и качества учебного процесса**

Оценка обучающимися содержания и качества учебного процесса по дисциплине осуществляется в установленном в СПбГУ порядке.

**3.2. Кадровое обеспечение**

**3.2.1 Образование и (или) квалификация штатных преподавателей и иных лиц, допущенных к проведению учебных занятий**

К преподаванию дисциплины могут быть допущены преподаватели, имеющие диплом о высшем образовании по соответствующему направлению.

**3.2.2 Обеспечение учебно-вспомогательным и (или) иным персоналом**

Не предусмотрено.

**3.3. Материально-техническое обеспечение**

**3.3.1 Характеристики аудиторий (помещений, мест) для проведения занятий**

В аудиториях, где проводятся занятия, необходимо наличие досок и средств письма на них.

**3.3.2 Характеристики аудиторного оборудования, в том числе неспециализированного компьютерного оборудования и программного обеспечения общего пользования**

В аудитории желательно наличие проектора.

**3.3.3 Характеристики специализированного оборудования**

По желанию преподавателя для подготовки к некоторым занятиям может потребоваться принтер, чтобы распечатать раздаточные материалы, а также проектор для проведения презентаций.

**3.3.4 Характеристики специализированного программного обеспечения**

Математические пакеты Matlab, Maple, R. TeX + (WinEdt или аналогичный редактор).

**3.3.5 Перечень и объёмы требуемых расходных материалов**

Фломастеры цветные, губки, бумага формата А3 (для блокнота-доски), канцелярские товары в объеме, необходимом для организации и проведения занятий. Также, в соответствии с разделом 3.3.3, может потребоваться белая бумага формата А4 для печати на принтере.

**3.4. Информационное обеспечение**

**3.4.1 Список обязательной литературы**

1. Регрессионный анализ: теоретические основы и практические рекомендации. Малов С.В., 2013.
2. MATLAB: Методы вычислений. Колпак Е.П., 2007.
3. Планирование и анализ для регрессионных моделей. Мелас В.Б., Шпилев П.В. СПб, изд. СПбГУ, 2014.

**3.4.2 Список дополнительной литературы**

1. Пепелышев А. Н. Статистические вычисления в пакете MATLAB., 2006.

2. Пепелышев А. Н. Некоторые современные методы исследования оптимальных планов эксперимента, 2006.

3. Мелас В.Б. Общая теория функционального подхода к оптимальному планированию эксперимента., 1999.

4. Мелас В.Б. Е - оптимальные планы эксперимента., 1997.

5. Н. Дрейпер, Г. Смит. Прикладной регрессионный анализ., 2007.

**3.4.3 Перечень иных информационных источников**

1. <http://statmod.ru/wiki/>.

2. https://www.ruhr-uni-bochum.de/mathematik3/en/research/index.html.

**Раздел 4. Разработчики программы**

Шпилев Петр Валерьевич, кандидат физико-математических наук, доцент, Кафедра статистического моделирования, e-mail: p.shpilev@spbu.ru;