

Федеральное государственное бюджетное образовательное
учреждение высшего образования
Московский государственный университет имени М.В. Ломоносова
Факультет наук о материалах

УТВЕРЖДАЮ
Зам. декана ФНМ по учебной
работе
_____/А.В. Кнотько /
«__» _____ 2020 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Наименование дисциплины:

Статистическая обработка эксперимента

Уровень высшего образования:

бакалавриат

Направление подготовки:

04.03.02 Химия, физика и механика материалов

Направленность (профиль)/специализация ОПОП:

общий

Форма обучения:

очная

Рабочая программа рассмотрена и одобрена
Методической комиссией факультета наук о материалах
(протокол №_____, дата)

Москва 2020

Рабочая программа дисциплины разработана в соответствии с самостоятельно установленным МГУ образовательным стандартом (ОС МГУ) для реализуемых основных профессиональных образовательных программ высшего образования по направлению подготовки «Химия, физика и механика материалов» (программы бакалавриата, магистратуры, реализуемых последовательно по схеме интегрированной подготовки) в редакции приказа МГУ от _____20__ г.

1. Место дисциплины в структуре ОПОП ВО: Базовая часть, блок общенаучной подготовки подготовки, модуль «Инженерия материалов», курс предназначен для студентов факультета наук о материалах **4-го года обучения (7-й семестр)**, курс является обязательным

2. Входные требования для освоения дисциплины, предварительные условия (если есть):

Математический анализ
Высшая алгебра и аналитическая геометрия
Теория вероятностей и математическая статистика
Программирование и ЭВМ
Химическая термодинамика и кинетика
Английский язык
Основы научного перевода

3. Результаты обучения по дисциплине:

Знать: способы расчёта погрешностей измерений и доверительных интервалов, методы проверки статистических гипотез, методы линейного и нелинейного регрессионного анализа применительно к методу наименьших квадратов, методы глобальной оптимизации (генетические алгоритмы, метод отжига, символьная регрессия), регуляризации, основы программирования на GNU Octave.

Уметь: решать задачи по расчёту погрешностей измерений, проверке статистических гипотез и регрессионному анализу, используя полученные теоретические знания, электронные таблицы (LibreOffice Calc или MS Excel) и язык программирования GNU Octave.

Владеть: теоретическими знаниями для проверки статистических гипотез, теорией и численными методами для линейного и нелинейного регрессионного анализа, основами программирования в GNU Octave.

Иметь опыт: решения типовых задач по расчёту погрешностей, проверке статистических гипотез и регрессионному анализу, в том числе, имитирующих реальные проблемы, с которыми приходится сталкиваться в практике химических исследований

4. Объем дисциплины составляет 2 з.е. (72 ак.ч.)

5. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий:

5.1. Структура дисциплины по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий (в строгом соответствии с учебным планом)

Вид работы	Семестр				Всего
	1				
Общая трудоёмкость, акад. Часов	72				72
Аудиторная работа:	36				36
Лекции, акад. Часов	0				0
Семинары, акад. Часов	0				0
Лабораторные работы, акад. часов	36				36
Самостоятельная работа, акад. часов	36				36
Вид итогового контроля (зачёт, экзамен)	Зач.				

5.2. Содержание разделов (тем) дисциплины

1. Погрешности измерений и проверка статистических гипотез (8 часов).

Случайные и систематические ошибки. Правила округления. Погрешность косвенных измерений. Статистические гипотезы и их проверка. Статистические критерии и доверительные интервалы. Корреляция и коэффициент корреляции.

2. Основы программирования в GNU Octave (6 часов).

Основы работы с матрицами в GNU Octave. Условные операторы и операторы циклов. Функции и лямбда-выражения (безымянные функции). Построение графиков. Статистические функции в GNU Octave.

3. Линейный метод наименьших квадратов (МНК) (8 часов).

Постановка задачи. Одномерный и многомерный МНК, система нормальных уравнений и методы её решения. Статистическая значимость линейной регрессии, оценка доверительных интервалов коэффициентов, оценка доверительных и предсказательных интервалов модели. Коэффициент детерминации. Регуляризация Тихонова и гребневая (ридж) регрессия.

4. Нелинейный метод наименьших квадратов (МНК) (6 часов).

Линеаризация (сведение к линейному МНК) как метод. Численные методы решения системы нелинейных нормальных уравнений. Матрицы Якоби и Гессе. Метод Гаусса-Ньютона и его демпфированные варианты: методы Левенберга-Марквардта и доверительных областей с ломаным/ключковидным шагом (англ. trust-region dogleg method). Приближенная оценка доверительных и предсказательных интервалов.

5. Погрешности вычислений на ЭВМ, численное и автоматическое дифференцирование (4 часа).

Двоичная система счисления, перевод между двоичной, десятичной, восьмеричной и шестнадцатеричной системами счисления. Двоичные дроби. Арифметика с плавающей запятой и формат IEEE-754. Ошибки округления при расчётах на ЭВМ и оценка оптимального шага численного дифференцирования.

Достоинства и недостатки разных методов дифференцирования. Понятие о дуальных числах и их арифметике. Их использование для автоматического дифференцирования функций.

7. Глобальная оптимизация (4 часа)

Локальный и глобальный минимум функции. Проблема поиска глобального и попадания в глобальный минимум. Функция Растригина как пример. Понятие о методах глобальной оптимизации: генетические алгоритмы, символьная регрессия (в т.ч. обратная польская запись и синтаксические деревья).

6. Фонд оценочных средств (ФОС, оценочные и методические материалы) для оценивания результатов обучения по дисциплине (модулю).

6.1. Типовые контрольные задания или иные материалы для проведения текущего контроля успеваемости, критерии и шкалы оценивания (в отсутствие утвержденных соответствующих локальных нормативных актов на факультете)

Примерный перечень задач на контрольных работах:

Контрольная работа №1 (погрешности, статистические гипотезы и доверительные интервалы)

- Расчёт погрешности косвенного измерения
- Расчёт доверительного интервала по Стьюденту на малой выборке.
- Теоретический вопрос на одну из тем: квантили, функции распределения, проверка статистических гипотез
- Оценка равномерности/неравномерности распределения с использованием критерия Пирсона (на примере игрального кубика).

Контрольная работа №2 (теория линейного и нелинейного МНК)

- Теоретический вопрос о степенях свободы в задаче многомерной линейной регрессии
- Теоретический вопрос по МНК
- Запись матрицы Якоби для конкретного случая нелинейной регрессии

- Линеаризация нелинейной регрессии: пример с логарифмами
- Линеаризация полиномиальной регрессии
- Решение задачи одномерной линейной регрессии без константы вручную. Выборка из 5 точек, ручной вывод формул для коэффициента и его доверительного интервала.

Контрольная работа №3 (численные методы дифференцирования, методы глобальной оптимизации)

- Перевод десятичных дробей в двоичные. Выводы о погрешностях округления при расчётах на ЭВМ с участием этих дробей.
- Сравнение точности двух формул численного дифференцирования с подбором оптимальных шагов дифференцирования.
- Дифференцирование функции одной переменной с использованием дуальных чисел.
- Представление алгебраических выражений в виде обратной польской записи и синтаксического дерева, расчёты в этом представлении.

Примерный перечень задач на домашних работах:

Домашняя работа №1 (погрешности, статистические гипотезы и доверительные интервалы)

- Проверка соответствия эмпирического распределения теоретическому (нормальному, хи-квадрат и т.п.)
- Расчёт доверительного интервала по Стьюденту с исключением грубого промаха
- Применение статистических критериев для сравнения дисперсий и матожиданий двух выборок.

Домашняя работа №2 (основы работы в GNU Octave)

- Задачи по формированию, видоизменению и перемножению матриц.
- График функции одной переменной
- График функции многих переменных
- Реализация расчёта квадратного или кубического корня методом Ньютона (задача на освоение управляющих конструкций языка)
- Интегрирование функции двух переменных методом Монте-Карло (задача на освоение матричной арифметики и логических матриц)

Домашняя работа №3 (линейный и нелинейный МНК)

- Одномерная линейная регрессия
- Многомерная линейная регрессия
- Полиномиальная регрессия: описание теплоёмкости индивидуального соединения
- Нелинейная регрессия, решение методом Левенберга-Марквардта (в т.ч. на примерах данных из ферментативной кинетики).
- Многомерная гребневая регрессия, подбор оптимального параметра регуляризации.

6.2. Типовые контрольные задания или иные материалы для проведения промежуточной аттестации по дисциплине (модулю), критерии и шкалы оценивания (в отсутствие утвержденных соответствующих локальных нормативных актов на факультете)

Вопросы к зачету

1. Правила округления, случайные и систематические погрешности. Погрешность косвенного измерения.
2. Функции распределения и доверительные интервалы. Грубые промахи.
3. Проверка статистических гипотез: квантили, p -value.
4. Сравнение дисперсий, средних, проверка распределения на нормальность: критерии Фишера, Стьюдента, Пирсона.
5. Ковариация и коэффициент корреляции.
6. GNU Octave: интерактивный режим, матричные операции и работа с графиками.
7. GNU Octave: скрипты, статистические функции, операторы циклов и условные операторы.

8. Метод наименьших квадратов: одномерная и многомерная линейная регрессия.
9. Метод наименьших квадратов: нелинейная регрессия и численные методы (Гаусса-Ньютона, Левенберга-Марквардта, доверительных областей).
10. Мультиколлинеарность и регуляризация Тихонова (гребневая или ридж-регрессия).
11. Двоичная система счисления. Двоичные дроби. Плавающая запятая, формат IEEE-754.
12. Глобальный и локальный минимум, проблема поиска глобального. Методы глобальной оптимизации: генетические алгоритмы, алгоритм имитации отжига.
13. Символьная регрессия. Обратная польская запись и синтаксические деревья.

7. Ресурсное обеспечение:

7.1. Перечень основной и дополнительной литературы

Основная литература:

1. Ю. Н. Тюрин, А. А. Макаров, «Анализ данных на компьютере». М., «Форум», 2008.
2. Спиридонов В.П., Лопаткин А.А. Математическая обработка физико-химических данных. Москва, издательство Московского Университета, 1970.
3. А.В. Пантелеев, Д.В. Метлицкая, Е.А. Алёшина. Методы глобальной оптимизации. Метаэвристические стратегии и алгоритмы. М., «Вузовская книга», 2013.

Дополнительная литература

1. К. Дёрффель, «Статистика в аналитической химии». М., «Мир», 1998.
2. «Справочник по прикладной статистике». Под ред. Э. Ллойда и У. Ледермана. М., «Финансы и статистика», 1989.
3. Г. Корн, Т. Корн. Справочник по математике для научных работников и инженеров. Москва, «Наука», 1973.
4. А.В. Гармаш, Н.В. Сорокина «Метрологические основы аналитической химии»
5. Lourakis M. I. A., Argyros A. A. Is Levenberg-Marquardt the Most Efficient Optimization Algorithm for Implementing Bundle Adjustment? 2005. ICCV'05 Proceedings of the Tenth IEEE International Conference on Computer Vision. V. 2. Beijing, China. P.1526-1531. <http://dx.doi.org/10.1109/ICCV.2005.128>.
6. Levenberg K. A method for the solution of certain non-linear problems in least squares // Quarterly of Applied Mathematics. 1944. V. 2. N 2. P.164-168. <https://doi.org/10.1090/qam/10666>
7. Marquardt D. W. An Algorithm for Least-Squares Estimation of Nonlinear Parameters //, J. Soc. Indust. Appl. Math. 1963. V. 11. N 2. P. 431-441. <http://dx.doi.org/10.1137/0111030>
8. R. Gonin, A. H. Money, Nonlinear L_p -norm estimation, Marcel Dekker Inc., New York and Basel.
9. Львовский Е.Н. Статистические методы построения эмпирических формул. Москва, Высшая школа, 1988.
10. Дьяконов В.П. MATLAB. Полный самоучитель. ДМК Пресс, 2012
11. Hansen J.S. GNU Octave beginner's guide. Packt Publishing, Birminham, UK, 2011.
12. Гилл Ф., Мюррей У., Райт М. Практическая оптимизация. Пер. с англ. Москва, «Мир», 1985. 509 с.

7.2. Перечень лицензионного программного обеспечения, в том числе отечественного производства (подлежит обновлению при необходимости)

Устанавливается на компьютерах студентов, для прохождения курса платное ПО не обязательно.

Обязательно: GNU Octave 4.x, LibreOffice

Рекомендуется: Microsoft Excel 2010 и выше

Опционально: LaTeX, MATLAB 8.x или 9.x

7.3. Описание материально-технического обеспечения.

Аудитория с доской, компьютерный проектор, компьютерный класс

8. Соответствие результатов обучения по данному элементу ОПОП результатам освоения ОПОП указано в Общей характеристике ОПОП.

9. Разработчик (разработчики) программы.

к.х.н. Алексей Леонидович Восков