

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ
ФЕДЕРАЦИИ**

**ФГАОУ ВО "Новосибирский национальный
исследовательский государственный университет"**

Факультет естественных наук

УТВЕРЖДАЮ

Декан ФЕН НГУ, профессор



Резников В.А.

«29» августа 2014 г.

Хеометрика

Курс 4–й, 7 семестр

Учебно-методический комплекс

Новосибирск 2014

Учебно-методический комплекс «Хемотрика» предназначен для студентов 4 курса ФЕН, специальность 020201 «Фундаментальная и прикладная химия». В состав комплекса включены рабочая программа курса, приведены примеры заданий, выполняемых студентами самостоятельно с использованием персонального компьютера и учебной литературы.

Составитель:

канд. хим. наук, доц. Т. Н. Дребущак

Учебно-методический комплекс

© Новосибирский государственный университет, 2014
© Дребущак Т. Н., 2014

Содержание

Аннотация рабочей программы	4
1. Цели освоения дисциплины	5
2. Место дисциплины в структуре ООП	5
3. Компетенции обучающегося, формируемые в результате освоения дисциплины	6
4. Структура и содержание дисциплины	6
5. Образовательные технологии	9
6. Учебно-методическое обеспечение самостоятельной работы студентов. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины	10
7. Учебно-методическое и информационное обеспечение дисциплины	18
8. Материально-техническое обеспечение дисциплины	18

Аннотация рабочей программы

Дисциплина «Хемометрика» относится к вариативной части (профильные дисциплины) профессионального (специального) цикла ООП по специальности 020201 «Фундаментальная и прикладная химия». Дисциплина реализуется на Факультете естественных наук Федерального государственного автономного образовательного учреждения высшего образования «Новосибирский национальный исследовательский государственный университет» кафедрой химии твердого тела.

Содержание дисциплины охватывает круг вопросов, связанных с анализом экспериментальных данных в области химии, с математическими методами анализа данных, с организацией и планированием экспериментов при проведении научных исследований.

Дисциплина нацелена на формирование общекультурных компетенций ОК-8, профессиональных компетенций ПК-1, ПК-12, ПК-15 выпускника.

Преподавание дисциплины предусматривает следующие формы организации учебного процесса: лекции, практические занятия в компьютерном классе, выполнение заданий с использованием персональных компьютеров, самостоятельная работа студента.

Программой дисциплины предусмотрены следующие виды контроля:

текущий контроль успеваемости в форме приема заданий в течение семестра, сдаются задания в три приема, сдача всех заданий - необходимое условие получения положительной оценки;

промежуточная аттестация в форме дифференцированного зачета.

Общая трудоемкость дисциплины составляет 2 зачетные единицы. Программой дисциплины предусмотрены 18 часов лекционных и 18 часов практических занятий, 22 часа самостоятельной работы студентов, 6 часов на проведение текущего контроля (сдача заданий) и 8 часов на проведение промежуточной аттестации (подготовка к зачету и сдача дифференцированного зачета).

Объем контактной работы обучающегося с преподавателем (занятия лекционного типа, семинарского типа, прием заданий, зачет) составляет 44 часа.

Работа с обучающимися в интерактивных формах составляет 36 часов.

1. Цели освоения дисциплины

Курс ставит своей целью усвоение студентами понятий, связанных с использованием математических, статистических и других методов для конструирования оптимальных измерительных процедур и для извлечения достоверной химической информации из экспериментальных данных.

Данный курс знакомит студентов с наиболее распространенными математическими методами анализа экспериментальных данных в области химии. Анализ данных необходим в любой области науки. Можно выделить некие общие принципы анализа данных, общие методы обработки информации. Часто эти методы основаны на статистических процедурах, так как экспериментальные данные, как правило, включают случайную компоненту. При проведении измерений практически невозможно учесть все факторы, влияющие на измеряемую величину. Практическая хемометрика – это компьютеризация, каждый шаг предполагает использование программного обеспечения. Студенты знакомятся с удобными средствами работы с большими массивами численных данных, со специализированными программами, в которых реализованы методы многомерной статистики. Материал иллюстрируется конкретными примерами из области химии твердого тела. В практической части курса даются навыки обработки экспериментальных данных с использованием электронных таблиц.

Для достижения поставленной цели выделяются задачи курса: знакомство с базовыми математическими методами анализа данных, овладение навыками обработки экспериментальных данных с использованием электронных таблиц на конкретных примерах из области химии.

2. Место дисциплины в структуре ООП

Дисциплина «Хемометрика» относится к вариативной части (профильные дисциплины) профессионального (специального) цикла ООП по специальности 020201 «Фундаментальная и прикладная химия».

Дисциплина «Хемометрика» опирается на следующие дисциплины данной ООП:

- Теория вероятностей и математическая статистика;
- Основы компьютерной грамотности;
- Аналитическая химия;
- Физическая химия;
- Химия твердого тела.

Результаты освоения дисциплины «Хемометрика» используются в следующих дисциплинах данной ООП:

- Методы кристаллоструктурных исследований;
- Термический анализ;
- Научно-исследовательская практика;
- Итоговая государственная аттестация.

3. Компетенции обучающегося, формируемые в результате освоения дисциплины «Хемометрика»:

- общекультурные компетенции: ОК-8 (*умеет работать с компьютером на уровне пользователя и способен применять навыки работы с компьютерами как в социальной сфере, так и в области познавательной и профессиональной деятельности*);
- профессиональные компетенции: ПК-1 (*понимает сущность и социальную значимость профессии, основных перспектив и проблем, определяющих конкретную область деятельности*), ПК-12 (*умеет применять основные законы химии при обсуждении полученных результатов, в том числе с привлечением информационных баз данных*), ПК-15 (*владеет методами регистрации и обработки результатов химических экспериментов*).

В результате освоения дисциплины обучающийся должен:

- знать основные понятия хемометрики, наиболее распространенные математические, статистические и другие методы извлечения достоверной химической информации из экспериментальных данных;
- уметь выбирать оптимальные процедуры обработки экспериментальных данных и выполнять практические задания с использованием современного компьютерного программного обеспечения;
- владеть навыками обработки экспериментальных данных с использованием электронных таблиц.

4. Структура и содержание дисциплины

Общая трудоемкость дисциплины составляет 2 зачетные единицы, 72 академических часа. Программой дисциплины предусмотрены 18 часов лекционных и 18 часов практических занятий, 22 часа самостоятельной работы студентов, 6 часов на проведение текущего контроля (сдача заданий) и 8 часов на проведение промежуточной аттестации (подготовка к зачету и сдача дифференцированного зачета).

№ п/п	Раздел дисциплины	Семестр	Неделя семестра	Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часах)	Формы текущего контроля успеваемости (по неделям семестра)

				Лекция	Практические занятия	Задания	Самост. работа	Подготовка к зачету и диф.зачет	Форма промежуточной аттестации (по семестрам)
1	Что такое «хеометрика», цели и задачи. Измерения в химии. Признаки и типы шкал.	7	1	2			1		
	Категории ошибок в химии. Погрешности измерения и причины их возникновения.		2	2			2		
2	Визуализация данных. Оценка точности измерений, распространение погрешностей.	7	3-4		4		1	1	Подготовка к зачету
3	Случайные величины, классификация. Параметры случайных величин и их оценки. Точечное оценивание.	7	5	2			2		
4	Виды распределений случайных величин. Интервальное оценивание.	7	6	2		2	2		Проверка 1-го задания
5	Представление различных видов распределений в электронных таблицах. Построение доверительных интервалов.	7	7		2		1	1	Подготовка к зачету
6	Теория статистического вывода, ошибки I и II рода.	7	8	2			2		
7	Проверка соответствия экспериментального и теоретического распределений, критерий хи-квадрат.	7	9		2		1	1	Подготовка к зачету
8	Поиск взаимосвязи двух и более случайных величин. Корреляция и регрессия.	7	10	2			2		
9	Методы сравнения экспериментальных данных (t-тесты, непараметрические тесты). Сравнение точности измерений.	7	11		2	2	2		Проверка 2-го задания
10	Метод наименьших квадратов, линейная	7	12	2			2		

	модель и анализ остатков, гетероскедактичность данных и варианты перехода к гомоскедактичному случаю.								
11	Сравнение двух процедур методами регрессионного анализа. Определение систематических ошибок.	7	13		2		1	1	Подготовка к зачету
12	Краткий обзор многомерных методов анализа данных. Определение источника вариации данных. Предсказание и оценивание, градуировка. Примеры проведения калибровочных процедур	7	14-15	4	2		1	1	Подготовка к зачету
13	Однофакторный и двухфакторный дисперсионный анализ с повторениями и без повторений, примеры использования этих процедур в химии.	7	16-17		4	2	1	1	Проверка 3-го задания, подготовка к зачету
14		7	18				1	2	Дифференцированный зачет
				18	18	6	22	8	

Итого: 72 часа.

Содержание курса:

Что такое «хеометрика», цели и задачи. Измерения в химии, признаки и типы шкал, допустимые преобразования шкал. Погрешности измерения и причины возникновения ошибок, классификация, систематические и случайные погрешности.

Визуализация данных, форматы представления данных. Организация данных в электронных таблицах. Импорт и экспорт файлов различных форматов, сортировка и фильтрация данных. Графическое представление данных, гистограммы, диаграммы рассеяния (примеры их химии – диаграммы Музера-Пирсона, Рамачадрана, проекции поверхности Хиршфельда).

Случайные величины, параметры случайных величин и их оценки. Свойства оценок. Определение основных статистик по выборке, точечное оценивание параметров случайных величин. Различные виды оценки одного и того же параметра в зависимости от типов шкал.

Некоторые виды распределений случайных величин, их представление в электронных таблицах. Графическое представление. Построение частотных распределений. Сравнение экспериментальных и теоретических распределений.

Интервальное оценивание, построение доверительных интервалов для различных параметров. Теория статистического вывода, 4 этапа проверки

статистических гипотез, ошибки первого и второго рода, определение мощности критерия.

Методы сравнения экспериментальных данных. Проверка соответствия экспериментального распределения и теоретического (или ожидаемого) распределения по критерию согласия хи-квадрат. Использование различных t-тестов для сравнения двух процедур. Выявление систематических ошибок эксперимента. Непараметрические тесты. Области применимости различных критериев проверки статистических гипотез.

Поиск взаимосвязи двух и более случайных величин, некоторые свойства двумерного нормального распределения. Корреляция и регрессия. Метод наименьших квадратов, линейная модель и анализ остатков, гетероскедастичность данных и варианты перехода к гомоскедастичному случаю. Сравнение двух процедур проведения эксперимента методами регрессионного анализа, выявление мультипликативных и аддитивных систематических ошибок.

Краткий обзор многомерных методов анализа. Стратегия обработки многомерных данных, представление о матрице корреляции и матрице близости. Классификация и построение моделей по таблице исходных данных. Множественная регрессия, коэффициент детерминации. Метод главных компонент, уменьшение размерности данных.

Определение источника вариации данных. Однофакторный дисперсионный анализ, сравнение нескольких процедур методами дисперсионного анализа. Двухфакторный дисперсионный анализ без повторений и с повторениями, определение перекрестного влияния факторов.

Общее представление о факторном, дискриминантном и кластерном анализе многомерных данных, области применения разных методов.

Предсказание и оценивание методами корреляционного и регрессионного анализа. Калибровочные (градуировочные) процедуры, выбор модели и проверка достоверности, анализ ошибок, весовые факторы. Проверка линейности регрессионной модели методами дисперсионного анализа.

5. Образовательные технологии

Занятия проводятся в лекционной форме и в форме практических занятий с использованием информационных технологий. Студенты выполняют задания в компьютерном классе и самостоятельно. Часть заданий предполагает работу в команде, например, при выборе оптимальных процедур обработки данных (студенты разбиваются на группы по 2-3 человека, каждая группа реализует свой путь решения проблемы). Часть заданий выполняется в режиме индивидуального обучения. В течение семестра проводится контроль выполнения заданий. Обязательным условием получения положительной оценки на дифференцированном зачете является сдача всех заданий преподавателю, т. е. студент должен не только показать правильно выполненное задание, но и уметь пояснить, как он это сделал, аргументированно ответить на все вопросы, поставленные в задании.

6. Учебно-методическое обеспечение самостоятельной работы студентов. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины

На протяжении всего курса запланирована самостоятельная работа студентов: работа с литературой, выполнение заданий. Примеры заданий, даваемых студентам для самостоятельной работы, приведены в учебном пособии Дребущак Т.Н. «Введение в хемометрику», 2013 г., изданного в НГУ. Все задания условно разбиты на три части (1 часть – задания 1-6 из пособия, 2 часть – задания 7-14, 3 часть – задания 15-20), после выполнения каждой из частей проводится прием заданий, на котором студенты демонстрируют выполненные в электронных таблицах задания, объясняют, какими методами обработки данных они пользовались, отвечают на дополнительные вопросы по поводу организации своих вычислений и сделанных выводов.

Примеры заданий:

Задание 1. Построение гистограммы. Дано 200 измерений некоторой величины (например, длины однотипных межатомных связей в разных кристаллических структурах):

1,160	1,175	1,181	1,185	1,189	1,193	1,197	1,200	1,205	1,211
1,163	1,176	1,181	1,185	1,189	1,193	1,197	1,201	1,206	1,211
1,164	1,176	1,181	1,185	1,189	1,193	1,197	1,202	1,206	1,212
1,165	1,177	1,182	1,186	1,189	1,193	1,197	1,202	1,207	1,213
1,166	1,177	1,182	1,186	1,189	1,193	1,197	1,202	1,207	1,213
1,166	1,178	1,182	1,186	1,190	1,193	1,197	1,202	1,207	1,214
1,167	1,178	1,183	1,186	1,190	1,193	1,197	1,202	1,208	1,215
1,168	1,178	1,183	1,187	1,190	1,194	1,198	1,202	1,208	1,215
1,168	1,178	1,183	1,187	1,190	1,194	1,198	1,202	1,208	1,215
1,169	1,178	1,183	1,187	1,191	1,194	1,198	1,203	1,208	1,216
1,169	1,178	1,183	1,187	1,191	1,194	1,198	1,203	1,208	1,217
1,170	1,179	1,183	1,187	1,191	1,194	1,199	1,203	1,208	1,218
1,171	1,179	1,184	1,188	1,191	1,195	1,199	1,203	1,208	1,219
1,172	1,179	1,184	1,188	1,191	1,195	1,199	1,204	1,209	1,219
1,173	1,180	1,184	1,188	1,191	1,196	1,199	1,204	1,209	1,220
1,173	1,180	1,184	1,188	1,192	1,196	1,199	1,204	1,210	1,221
1,174	1,180	1,184	1,188	1,192	1,196	1,199	1,205	1,210	1,222
1,175	1,180	1,185	1,189	1,192	1,196	1,200	1,205	1,210	1,223
1,175	1,181	1,185	1,189	1,192	1,197	1,200	1,205	1,210	1,225
1,175	1,181	1,185	1,189	1,193	1,197	1,200	1,205	1,210	1,230

Определить среднее значение, дисперсию и стандартное отклонение величины (использовать соответствующие функции в электронных таблицах).

Построить частотную таблицу. По частотной таблице построить гистограмму. Является ли распределение одномодальным?

Задание 2. Элементарные вычисления и построение диаграмм рассеяния. Дана таблица изменения параметров элементарной ячейки ромбической модификации парацетамола в зависимости от температуры. Провести анализ этих данных:

- внести данные в электронную таблицу (стандартные отклонения, которые стоят в скобках, внести в отдельную таблицу);
- построить график изменения объема в зависимости от температуры, правильно оформить;
- вычислить относительные изменения параметров ячейки (за исходное значение взять значение при $T = 300$ К); вычислить ошибки относительных изменений с учетом законов распространения погрешностей (взять формулу для предела погрешности);
- построить на одном графике относительные изменения всех параметров ячейки, правильно оформить.

Параметры элементарной ячейки ромбического парацетамола при разных температурах

T, K	$a, \text{Å}$	$b, \text{Å}$	$c, \text{Å}$	$V, \text{Å}^3$
100	7,1986(17)	11,782(10)	17,183(4)	1457,4(10)
150	7,2433(18)	11,793(11)	17,175(5)	1467,1(11)
200	7,2927(14)	11,806(8)	17,169(3)	1478,3(8)
250	7,3467(14)	11,818(8)	17,165(3)	1490,2(8)
260	7,3583(12)	11,826(7)	17,163(3)	1493,5(7)
270	7,3697(15)	11,823(9)	17,165(4)	1495,6(9)
280	7,3812(16)	11,830(9)	17,164(4)	1498,8(9)
300	7,4049(13)	11,835(8)	17,162(3)	1504,3(8)
320	7,4308(13)	11,839(7)	17,163(3)	1509,9(7)
330	7,4430(16)	11,847(9)	17,160(4)	1513,2(9)
340	7,4556(16)	11,853(9)	17,159(4)	1516,3(9)
350	7,4683(15)	11,855(9)	17,158(4)	1519,1(9)
360	7,4807(15)	11,853(9)	17,160(4)	1521,6(9)

Формула расчета относительных изменений параметра в процентах:

$$\Delta l / l(\%) = \frac{l - l_0}{l_0} * 100.$$

Сделать выводы: носит ли изменение объема линейный характер в исследуемом диапазоне температур, положительные или отрицательные коэффициенты термического расширения (объемный и линейные), в каком направлении структура деформируется сильнее всего, существует ли направление, в котором деформация отсутствует?

Задание 3. Работа с текстовой информацией, сортировка и фильтрация. Импортировать текстовый файл с исходными данными с монокристалльного четырехкружного дифрактометра STADI4 в таблицу.

- Скопировать рефлексы без профилей и углов (до колонки угла пси) на отдельный лист, поставить автофильтр. Пользуясь автофильтром, посчитать, сколько рефлексов имеют неодинаковый фон, сколько слишком широких рефлексов, сколько находилось не в центре при сканировании.

- Посчитать общее количество измеренных рефлексов без учета стандартов (отсортировав по условию, скопировать этот массив на новый лист и воспользоваться функцией «счет»), найти отношение средней интенсивности к среднему значению σ . Найти количество сильных рефлексов ($I/\sigma > 4$). Для этого в колонке рядом вычислить отношение I/σ и поставить фильтр. Отсортировав по условию, скопировать этот массив на новый лист и воспользоваться функцией «счет».

- Навести статистику по сильным рефлексам (по каким группам рефлексов есть погасания). По погасаниям определить возможные ПГС кристаллической структуры.

Задание 4. Графическое представление нормального распределения. Построить график стандартного нормального распределения. Задать значения по x от -3 до 3 с шагом $0,1$. Построить интегральное распределение, воспользовавшись функцией нормального распределения в электронных таблицах. Построить дифференциальное распределение (плотность вероятности), поставив соответствующий статус в функции нормального распределения.

Задание 5. Определение численных значений нормального распределения по таблицам и их интерпретация. Используя функции электронных таблиц, ответить на следующие вопросы.

- Какова вероятность того, что случайное измерение (нормальное распределение) попадет в интервал $\pm 1\sigma$, $\pm 2\sigma$, $\pm 3\sigma$?

- Пусть среднее значение равно 100 , стандартное отклонение равно 15 (нормальное распределение). Какова вероятность того, что мы получили значение 120 и выше случайно? А 140 и выше?

- Для стандартного нормального распределения $N(0, 1)$ найти значение процентной точки z_α , если известна вероятность α случайно получить значение $x > z_\alpha$: $\alpha = 0,1$; $\alpha = 0,05$; $\alpha = 0,01$?

- Найти значение процентной точки $z_{\alpha/2}$, если известна вероятность случайно получить значение x в интервалах либо больше $z_{\alpha/2}$, либо меньше $-z_{\alpha/2}$ (двусторонний случай) для тех же вероятностей α .

Задание 6. Различные виды распределений случайных величин, основанных на нормальном, и их представление в электронных таблицах.

Построить график хи-квадрат-распределения (степень свободы задать в отдельной ячейке и использовать ее как аргумент в формуле). Задать значения

по x от 0 до 10 с шагом 0,1. Меняя степень свободы, проследить, как меняется положение максимума.

Какова вероятность того, что случайное измерение при объеме выборки 8 попадет в интервал $\pm 1\sigma$, $\pm 2\sigma$, $\pm 3\sigma$ (использовать распределение Стьюдента)? Как будет меняться эта вероятность при увеличении объема выборки? При каком объеме выборки различие между нормальным распределением и распределением Стьюдента станет меньше 1 % для интервала $\pm 2\sigma$?

Допустим, что дисперсии двух выборок одинакового объема одинаковы. Какова вероятность случайно получить оценку отношения дисперсий 1, 3, 12 (взять объем выборки 10)? Для ответа воспользоваться табличными значениями F -распределения.

Задание 7. Интервальное оценивание (построение доверительных интервалов). Анализируется международный стандарт вещества новым методом. Известна концентрация $\mu_0 = 0,500$ мг/г. Получили следующие результаты:

0,559; 0,594; 0,531; 0,465; 0,509; 0,487; 0,421; 0,328; 0,428; 0,601; 0,450; 0,440; 0,409; 0,419; 0,357.

Построить доверительные интервалы для среднего и для дисперсии. Попадает ли истинное значение в этот доверительный интервал? Можно ли считать, что новый метод дает удовлетворительный результат на уровне значимости 5 %, 1 % (проверка статистической гипотезы, сравнение со стандартом)?

Задание 8. Проверка соответствия распределений. Проверить по критерию согласия хи-квадрат соответствие экспериментального и нормального распределений (по таблице из задания 1). Сделать это двумя способами: а) вычислить по формулам, которые даны в теоретической части и проверить соответствующую статистическую гипотезу по хи-квадрат-распределению; б) воспользоваться статистической функцией хи-тест. Сравнить результаты. Пояснение: хи-тест выдает значение вероятности, соответствующее критерию (площади «хвоста» функции плотности вероятности распределения).

Задание 9. Ошибка второго рода при проверке статистической гипотезы. Предположим, есть основания считать, что среднее значение случайной величины x известно ($\mu_x = 8$). Предположим далее, что дисперсия x также известна ($\sigma_x^2 = 2$). Найти размер выборки, позволяющей построить критерий проверки гипотезы с 5 %-м уровнем значимости и 5 %-й ошибкой второго рода для выявления 10 %-х отклонений от гипотетического значения. Построить область принятия гипотезы для данного критерия.

Задание 10. Проверка статистических гипотез, сравнение двух серий измерений. Результаты анализа двухфазного образца с известным содержанием фаз 1 : 1 двумя разными методами (одна из фаз в %) даны в таблице.

Дают ли методы одинаковый результат? Есть ли систематические ошибки методов? Какой метод точнее? (Проверить, одинаковы ли дисперсии этих измерений по F -тесту).

Проба	Метод А	Метод В
1	49,9	50,1
2	49,9	49,5
3	50,0	50,3
4	50,1	49,6
5	50,0	49,9
6	50,0	50,2

Задание 11. Проверка статистических гипотез, сравнение двух серий измерений. Содержание полиморфной модификации А в 10 анализируемых образцах, определенное двумя методами (R и T), дано в таблице:

Образец	X_R	X_T
1	57	58
2	25	21
3	50	48
4	10	5
5	45	47
6	53	50
7	50	48
8	48	51
9	80	75
10	55	52

Дает ли проверяемый метод R такой же результат, как и тестовый (t -тест)?

Задание 12. Проверка статистических гипотез, сравнение двух серий измерений. Концентрация активного вещества в 10 пробах до и после обработки дана в таблице:

Проба	До	После
1	20,5	21,3
2	19,9	20,4
3	21,1	20,8
4	21,5	22,1
5	20,0	20,7
6	19,1	18,9
7	20,1	20,5
8	20,5	20,4
9	19,4	19,7
10	19,1	19,4

Эффективна ли обработка, если цель обработки увеличить содержание активного вещества (сравнение двух парных измерений)? Проверить гипотезу на разных уровнях значимости.

Задание 13. Непараметрические методы сравнения, тест Вилкоксона. Содержание полиморфной модификации А в 10 анализируемых образцах, определенное двумя методами (R и T), дано в таблице задания 11. Проверить по тесту Вилкоксона, дает ли проверяемый метод R такой же результат, как и тестовый T ?

Задание 14. Непараметрические методы сравнения, U -тест. Сравнить две группы измерений, дают ли они одинаковый результат?

А: 1,34; 1,64; 1,78; 1,33; 1,80; 1,93; 2,08; 1,31; 1,30; 1,40.

В: 1,31; 1,34; 1,45; 1,49; 1,86; 1,75; 1,62; 1,30.

Для проверки использовать U -тест.

Задание 15. Сравнение двух процедур, регрессионный анализ. По таблице из задания 11 проверить с помощью регрессионного анализа, дает ли проверяемый метод R такой же результат, как и тестовый? Есть ли систематическая ошибка? Проанализировать возможный тип систематической ошибки (проверить соответствующую статистическую гипотезу на каждый тип).

Задание 16. Регрессионный анализ. Сделать выборку измеренных стандартов из задания 3. Вычислить относительные изменения интенсивностей стандартов (отсчитывать от первого стандарта из группы). Есть или нет тренд (тенденция падения или возрастания) стандартов? Сравнить с усредненной по всем стандартам зависимостью (вычислить наклоны линий регрессии и ошибки, сделать выводы).

Задание 17. Однофакторный дисперсионный анализ. Определено содержание действующего компонента М в семи лабораториях:

а	б	в	г	д	е	ж
16	46	12	15	60	62	33
29	28	19	27	39	38	38
35	30	29	34	43	55	55
18	45	11	20	58	42	49
22	31	29	34	40	53	45

Дают ли все лаборатории одинаковый результат (есть или нет систематическая лабораторная ошибка)?

Задание 18. Двухфакторный дисперсионный анализ без повторений. Концентрация загрязняющего вещества вблизи участка размещения промышленных отходов в зависимости от расстояния от города и от глубины взятия пробы дана в таблице:

Глубина , м	Расстояние от города, км			
	1	2	3	4
0,0	25,0	15,3	10,1	5,2
0,5	23,0	15,2	9,0	4,0
1,0	22,5	13,8	12,5	3,0

Влияет ли местоположение взятия пробы, глубина взятия пробы? Можно ли определить перекрестное влияние?

Задание 19. Двухфакторный дисперсионный анализ с повторениями. Содержание активного вещества в пробах разное. Пробы обрабатываются двумя разными методами. Проверить, одинаковые ли результаты дают разные методы, влияет ли содержание активного вещества в пробах на результат и существует ли перекрестное влияние «метод и концентрация активного вещества»?

	Проба 1	Проба 2	Проба 3
Метод 1	92	101	89
	114	118	120
	107	98	110
	99	105	115
Метод 2	91	101	79
	74	68	88
	65	59	55
	90	70	93

Задание 20. Градуировочные процедуры. Построить градуировочную прямую по данным таблицы, построить доверительные интервалы для линии регрессии. Данные калибровки:

Конц., x	Пик, y
0,144	21,333
0,280	40,521
0,044	7,702
0,018	3,707
0,232	32,165
0,120	17,094
0,035	4,758

0,022	3,396
0,035	3,829
0,058	9,057
0,104	14,765
0,070	10,285
0,208	28,149
0,128	19,396
0,080	11,340
0,144	21,333

На дифференцированном зачете предлагается ответить на вопрос, проиллюстрировав ответ примерами из заданий, которые выполнялись в течение семестра.

Примеры вопросов к зачету:

1. Измерения в химии. Типы шкал. Допустимые преобразования шкал.
2. Погрешности измерения и причины их возникновения. Классификация ошибок. Распространение погрешностей.
3. Что такое случайная величина. Нормальное распределение.
4. Виды распределений случайных величин, основанных на нормальном.
5. Параметры случайных величин и их оценки. Точечное оценивание параметров. Свойства оценок.
6. Интервальное оценивание (построение доверительных интервалов). Доверительные интервалы для среднего, для дисперсии.
7. Проверка статистических гипотез, основные этапы. Ошибки первого и второго рода.
8. Проверка соответствия экспериментального распределения и теоретического (или ожидаемого) распределения по критерию согласия хи-квадрат.
9. Методы сравнения двух процедур по t-тестам. Сравнение точности процедур.
10. Корреляция и регрессия, метод наименьших квадратов. Сравнение двух процедур проведения эксперимента методами регрессионного анализа. Определение типа систематической ошибки.
11. Непараметрические методы сравнения двух процедур. Типы шкал. Процедура ранжирования.
12. Определение источника вариации данных. Однофакторный дисперсионный анализ.
13. Определение источника вариации данных. Двухфакторный дисперсионный анализ.
14. Калибровочные процедуры. Регрессия и проверка на линейность, поправка на гетероскедастичность.

7. Учебно-методическое и информационное обеспечение дисциплины

а) основная литература:

1. *Дребуцак Т. Н.* Введение в хемометрику. Новосиб. гос. ун-т, Новосибирск, 2013.
2. *Брандт З.* Анализ данных. Статистические и вычислительные методы для научных работников и инженеров. Пер. с англ. М.: Мир, ООО «Изд-во АСТ», 2003.
3. *Hughes I. G., Hase T. P. A.* Measurements and their uncertainties. Oxford Univ. Press Inc., 2010.

б) дополнительная литература:

1. *Боровков В.* STATISTICA: искусство анализа данных на компьютере. Для профессионалов. СПб: Питер, 2001.
2. *Brereton, R. G.* Chemometrics: Data analysis for the laboratory and chemical plant. Wiley, Chichester, UK, 2003.
3. *Massart D. L., Vandeginste B. G. M., Deming S. N., Michotte Y., Kaufman L.* Chemometrics: a textbook. Elsevier, 1988.
4. *Бендат Дж., Пирсол А.* Прикладной анализ случайных данных. Пер. с англ. М.: Мир, 1989.
5. *Эсбенсен К.* Анализ многомерных данных, сокр. пер. с англ. под ред. О. Родионовой, Изд-во ИПХФ РАН, 2005.
6. *Шараф М. А., Иллман Д. Л., Ковальски Б. Р.* Хемометрика. Пер. с англ. Л.: Химия, 1989.
7. *Загоруйко Н. Г.* Прикладные методы анализа данных и знаний. Новосибирск: Изд-во Ин-та математики, 1999.
8. *Александров В. В., Алексеев А. И., Горский Н. Д.* Анализ данных на ЭВМ (на примере системы СИТО). М.: Финансы и статистика, 1990.
9. *Ким Дж.-О., Мьюллер Ч. У., Клекка У. Р. и др.* Факторный, дискриминантный и кластерный анализ: Пер. с англ. М.: Финансы и статистика, 1989.
10. Измерения в промышленности: Справ. изд.: В 3 кн.: Пер. с нем. / Под ред. П. Профоса. 2-е изд., перераб. и доп. М.: Металлургия, 1990. Кн. 1: Теоретические основы.

в) программное обеспечение и Интернет-ресурсы:

1. *Дребуцак Т. Н.* Введение в хемометрику. http://www.htt.nsu.ru/wp-content/uploads/2014/12/Posobie_Chemometr.pdf
2. Chemometrics World. Доступно на сайте <http://www.wiley.com/legacy/wileychi/chemometrics/about.html>.
3. Российское хемометрическое общество. Доступно на сайте <http://rsc.chemometrics.ru>.

8. Материально-техническое обеспечение дисциплины

- Ноутбук, медиа-проектор, экран, компьютерный класс на 10 оборудованных персональными компьютерами мест.
- Программное обеспечение для демонстрации слайд-презентаций,

электронные таблицы.

Программа составлена в соответствии с требованиями ФГОС ВПО с ОС ВПО, принятым в ФГАОУ ВО «Новосибирский национальный исследовательский государственный университет», с учетом рекомендаций ООП ВПО по специальности 020201 «Фундаментальная и прикладная химия».

Автор: Дребущак Татьяна Николаевна, к.х.н., доцент кафедры химии твердого тела ФЕН, н.с. ИХТТМ СО РАН

Программа одобрена на заседании кафедры химии твердого тела 29 августа 2014 г.

Секретарь кафедры, к.х.н.  Т. Н. Дребущак