

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«КАМЧАТСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «КамчатГТУ»)

Технологический факультет

Кафедра «Экология и природопользование»

УТВЕРЖДАЮ
Декан технологического факультета
Я.М. Хорошман
«01» 12 2021 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Методы математической статистики в экологии»

направление подготовки
05.03.06 Экология и природопользование
(уровень бакалавриата)

профиль:
«Экология»

Петропавловск-Камчатский
2021

Рабочая программа по дисциплине «Методы математической статистики в экологии» составлена на основании ФГОС ВО направления подготовки 05.03.06 «Экология и природопользование».

Составитель рабочей программы

Доцент кафедры ЭП, к.б.н.  Миловская Л.В.

Рабочая программа рассмотрена на заседании кафедры ЭП
«01» декабря 2021 г., протокол № 6

Заведующий кафедрой
«01» декабря 2021 г.,  Ступникова Н.А.

1. Цели и задачи учебной дисциплины

Целью дисциплины «Методы математической статистики в экологии» является овладение методами количественной оценки получаемых результатов исследований.

Задачи дисциплины:

- изучение основных характеристик варьирующих объектов;
- изучение законов распределения;
- изучение критериев достоверности оценок;
- знакомство с корреляционным анализом;
- знакомство с регрессионным анализом.

Основные *прикладные задачи*, которые позволят решить применение статистических методов в экологических исследованиях, следующие:

- овладение способами группировки первичных данных;
- владение способами расчёта средних величин и ошибки;
- построение кривых распределения;
- выбор уравнений регрессии, имеющих биологический смысл.
- оценка достоверности полученных результатов.
- умение правильно анализировать результаты наблюдений;
- построение графиков зависимости.
- создание доказательной базы проведенного исследования.

2. Требования к результатам освоения дисциплины

Процесс изучения дисциплины направлен на формирование профессиональной компетенции:

– ОПК-3 — Способен применять базовые методы экологических исследований для решения задач профессиональной деятельности.

– ОПК-6 — Способен проектировать, представлять, защищать и распространять результаты своей профессиональной и научно-исследовательской деятельности.

Планируемые результаты освоения практики, соотнесенные с планируемыми результатами освоения образовательной программы, представлены в таблице 1.

Таблица 1 — Планируемые результаты обучения по дисциплине, соотнесенные с установленными в программе бакалавриата индикаторами достижения компетенций

Код компетенции	Наименование компетенции	Код и наименование индикатора достижения	Планируемый результат обучения по дисциплине	Код показателя освоения
1	2	3	4	5
ОПК-3	Способен применять базовые методы экологических исследований для решения задач профессиональной деятельности	ИД-1 _{ОПК-3} : Знает основные методы экологических исследований для решения задач профессиональной деятельности.	Знать:	3(ОПК-3)1
			1. изучение основных характеристик варьирующих объектов;	3(ОПК-3)2
			2. изучение законов распределения;	3(ОПК-3)3
			3. изучение критериев достоверности оценок;	3(ОПК-3)4
			4. знакомство с корреляционным и регрессионным анализом.	3(ОПК-3)5
5. оценку достоверности полученных результатов.				

1	2	3	4	5
ОПК-3		ИД-2 _{ОПК-3} : Владеет навыками проведения экологических исследований для решения задач профессиональной деятельности.	Владеть: 1. способами группировки первичных данных; 2. способами расчёта средних величин и ошибки; 3. построением кривых распределения;	В(ОПК-3)1 В(ОПК-3)2 В(ОПК-3)3
ОПК-6	Способен проектировать, представлять, защищать и распространять результаты своей профессиональной и научно-исследовательской деятельности	ИД-1 _{ОПК-6} : Знает основные методы проектирования профессиональной и научно-исследовательской деятельности	Знать: 1. построение графиков зависимости. 2. выбор уравнений регрессии, имеющих биологический смысл.	З(ОПК-6)1 З(ОПК-6)2
ОПК-6		Владеет навыками представлять, защищать и распространять результаты своей профессиональной и научно-исследовательской деятельности	Владеть: 1. доказательной базой проведенного исследования	В (ОПК-6)1

3 Место дисциплины в структуре образовательной программы

Учебная дисциплина «Методы математической статистики в экологии» является дисциплиной обязательной части в структуре образовательной программы.

При изучении дисциплины «Основы математической статистики в экологических исследованиях» используются знания по таким дисциплинам, как:

Биология — основы организации живых организмов и особенностях их функционирования на молекулярном, клеточном, тканевом, организменном, популяционном, экосистемном и биосферном уровнях; формировании представлений о роли живых организмов в общей структуре и взаимодействии сфер Земли, разнообразие организмов.

Биоразнообразие — структура, уровни и методы измерения биоразнообразия, современная картина биоразнообразия.

Общая экология — влияние факторов окружающей среды на отдельные организмы, популяции и виды; структурные и функциональные характеристики, динамика численности популяций, внутривидовые группировки и их взаимоотношения;), пути формирования биоценозов, их взаимодействие с внешней средой.

Математика — логарифмическое, дифференциальное и интегральное исчисление.

Знания по дисциплине «Методы математической статистики в экологии» используются для математической обработки результатов исследования при подготовке выпускной квалификационной работы.

4. Содержание дисциплины

4.1 Тематический план дисциплины (табл. 2)

Очная форма обучения

	Всего часов	Аудиторные занятия	Контактная работа по видам учебных занятий			Самостоятельная работа	Формы текущего контроля	Итоговый контроль знаний по дисциплине
			Лекции	практические занятия	Лабораторные работы			
Раздел 1. Классификация признаков и способы группировки первичных данных.	48	18			18	30	Контроль-ная	
Тема 1: Понятие генеральной совокупности и выборки.	6	2	–	–	2	4	Опрос, практические задания	
Тема 2: Расчёт средних величин и средневзвешенной.	6	2	–	–	2	4	Опрос, практические задания	
Тема 3: Показатели вариации. Построение вариационных рядов. Мода, медиана.	6	2	–	–	2	4	Опрос, практические задания	
Тема 4: Основные типы распределения признаков.	6	2	–	–	2	4	Опрос, практические задания	
Тема 5: Оценка нормальности: асимметрия и эксцесс.	6	2	–	–	2	4	Опрос, практические задания	
Тема 6: Статистическая ошибка (ошибка репрезентативности выборочных параметров).	4	2	–	–	2	2	Опрос, практические задания	
Тема 7: Доверительный интервал. Определение точности опыта. Оптимальный объем выборки.	6	2	–	–	2	4	Опрос, практические задания	
Тема 8: Сравнение средних величин. Критерии сравнения.	8	4	–	–	4	4	Опрос, практические задания	
Раздел 2. Виды анализа и функциональные зависимости.	24	12	–	–	12	12	Контроль-ная	
Тема 9: Корреляционный анализ	8	4	–	–	4	4	Опрос, практические задания	
Тема 10: Регрессионный анализ.	8	4	–	–	4	4	Опрос, практические задания	
Тема 11: Метод наименьших квадратов.	8	4	–	–	4	4	Опрос, практические задания	
Зачет			–	–				+
Всего	72	30	–	–	30	42		

Заочная форма обучения

	Всего часов	Аудиторные занятия	Контактная работа по видам учебных занятий			Самостоятельная работа	Формы текущего контроля	Итоговый контроль знаний по дисциплине
			Лекции	практические занятия	Лабораторные работы			
Раздел 1. Классификация признаков и способы группировки первичных данных.	53	5	-	-	5	48	Контроль-ная	
Тема 1: Понятие генеральной совокупности и выборки.	7	1	-	-	1	6	Опрос, практические задания	
Тема 2: Расчёт средних величин и средневзвешенной.	7	1	-	-	1	6	Опрос, практические задания	
Тема 3: Показатели вариации. Построение вариационных рядов. Мода, медиана.	6	-	-	-	-	6	Опрос, практические задания	
Тема 4: Основные типы распределения признаков.	6	-	-	-	-	6	Опрос, практические задания	
Тема 5: Оценка нормальности: асимметрия и эксцесс.	7	1	-	-	1	6	Опрос, практические задания	
Тема 6: Статистическая ошибка (ошибка репрезентативности выборочных параметров).	7	1	-	-	1	6	Опрос, практические задания	
Тема 7: Доверительный интервал. Определение точности опыта. Оптимальный объем выборки.	7	1	-	-	1	6	Опрос, практические задания	
Тема 8: Сравнение средних величин. Критерии сравнения.	6	-	-	-	-	6	Опрос, практические задания	
Раздел 2. Виды анализа и функциональные зависимости.	15	3	-	-	3	12	Контроль-ная	
Тема 9: Корреляционный анализ	5	1	-	-	1	4	Опрос, практические задания	
Тема 10: Регрессионный анализ.	5	1	-	-	1	4	Опрос, практические задания	
Тема 11: Метод наименьших квадратов.	5	1	-	-	1	4	Опрос, практические задания	
Зачет	4		-	-				4
Всего	72		-	-	8	60		

4.2 Содержание дисциплины

Раздел 1. Классификация признаков и способы группировки первичных данных.

Лабораторная работа 1: Понятие генеральной совокупности и выборки.

Понятие генеральной совокупности и выборки; признаки и их свойства, классификация признаков; случайная величина, варьирование значений, достаточная точность измерений; принципы расчёта статистических показателей.

Основные понятия темы: признак, варианта, варьирование, выборка, генеральная совокупность.

Вопросы для обсуждения:

1. Понятие генеральной совокупности и выборки.
2. Признаки и их свойства.
3. Точность измерений.
4. Способы группировки первичных данных.

Задание 1: Распределение случайных величин на примере выпадения орла или решки при разном количестве подкидываний. Частные результаты сравниваются, затем объединяются в общую базу для получения вероятностного прогноза распределения случайных величин.

Для удобства частоту записывайте методом конверта:

· · · · · □ □ □ □ □ □
1 2 3 4 5 6 7 8 9 10

Таблица 1.1 — Частота выпадения орла (О) или решки (Р) в зависимости от числа повторений

1). Распределение случайных величин на примере выпадения орла или решки при разном количестве подкидываний. Частные результаты сравниваются, затем

	5 повторов	20 повторов	50 повторов
О			
Р			
Всего			

Сделать вывод.

Задание 2: определить зависимость средней арифметической от величины выборки.

Таблица 1.2 — Длина листа гибискуса, см

Блок 1 <i>n</i> =25	9,3	9,6	8,1	8,6	11
	9,3	8,5	10,1	9,9	11,1
	9,4	8,6	8,6	8,8	11,2
	9,4	11,6	10	9,4	11,2
	9,4	10	9,2	9,8	11,3
Блок 2 <i>n</i> =35	9,5	10,2	10,3	9,9	11,4
	8,7	7,3	9,2	9,3	7,2
	8,1	8,8	9,5	9,3	7,4
	9,7	9,7	9	9	11,1
	11,2	9,7	9,2	8,7	11,3
	8,3	9,8	8	9,9	11
	9,7	9,8	9,8	11,9	11,1

Блок 3 <i>n</i> =40	9,1	9,8	9,2	9,1	11,2
	9,1	9,9	9,3	9,5	11,2
	8,4	9,9	8,7	10	11,3
	9,6	9,2	9,8	9,6	11,4
	9,6	7,6	8,5	9,3	11
	9,6	8,1	10	9,4	11,1
	9,7	8,8	8,8	9,1	9,8
	9,2	11,5	9,6	7,9	8,6

В таблице 1.2 представлены данные 100 измерений длины листа гибискуса (китайская роза). По данным таблицы рассчитайте среднюю арифметическую:

- для выборки из 25 измерений (блок 1);
- для выборки из 35 измерений (блок 2);
- для выборки из 40 измерений (блок 3);
- рассчитайте среднюю арифметическую для всего массива данных из 100 измерений. Сравните результаты и сделайте вывод.

Расчеты проводятся в *Excel*. Для будущих расчетов сохраните в *Excel* данные ряда из 100 измерений.

Сделайте вывод о влиянии числа измерений на конечный результат.

Литература: [1], [2].

Лабораторная работа 2: Расчёт средних величин и средневзвешенной.

Средняя величина признака как одна из важнейших обобщающих характеристик вариационного ряда. Виды средних (средняя арифметическая — простая и взвешенная, средняя геометрическая), но в практике биологических исследований наибольшее значение имеет средняя арифметическая – величина, вокруг которой концентрируются варианты.

Практическое занятие: расчет средних величин

Задание 1: расчет средней геометрической.

Средняя геометрическая используется в анализе динамики для определения среднего коэффициента роста или если числа имеют тенденцию к большим колебаниям. На практике этот показатель используют не так часто, как среднее арифметическое, но все же встречается. Например, скорость роста численности населения. Она рассчитывается, как средняя геометрическая из нескольких индексов.

По формуле $x_{\text{геом}} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$ рассчитайте среднюю геометрическую по следующим данным:

скорость роста численности выросла за первый год в 1,2 раза, за второй — в 1,5 раза, в третий год — в 1,3 раза.

Рассчитайте среднюю скорость роста численности за 3 года.

Задание 2: расчет средневзвешенных величин.

а) Измеряли температуру глубокого озера по вертикали. Известно, что наиболее изменчива температура в поверхностном горизонте, с глубиной температура изменяется меньше, надо определить среднее значение температуры в слое 0–дно (таблица 2.1).

Таблица 2.1 — Расчет средней температуры воды в слое 0–дно

$h, \text{ м}$	T°	$(T_n+T_{n+1})/2$	$(T_n+T_{n+1})/2 \times dh$
0	12,3		
5	11,9		
10	11,2		
15	10,3		
25	7,0		
50	5,3		
100	4,5		
140	4,2		
		$\sum (T_n+T_{n+1})/2 \times dh$	
		T° средневзв.	
		T° среднеарифм.	

$h, \text{ м}$ — глубина, на которой измеряли температуру (T°)

$(T_n+T_{n+1})/2$ — средняя температура двух смежных горизонтов

$(T_n+T_{n+1})/2 \times dh$ — произведение средней температуры двух смежных горизонтов на величину слоя между ними.

Контрольные вопросы:

Почему различаются значения средневзвешенной и средней арифметической?

Какую величину корректнее рассчитывать с учетом неравномерности измерений по вертикали?

Задание 3: сравнение средней и средневзвешенной величин.

Необходимо объединить результаты расчетов по нескольким выборкам (табл. 2.2) и на этой основе найти общую среднюю, характеризующую весь материал. Рассчитайте взвешенную среднюю, которая учитывает объемы частных выборок:

$$\bar{M} = \frac{\sum n_j \times M_j}{\sum n_j}$$

M_j — значение частной средней

n_j — условные «веса» частного значения (объем выборки)

Таблица 2.2 — Содержание свинца в почве

Место взятия проб	мг/кг									
	Участок 1	10	19	13	20	21	25	26	28	29
Участок 2	24	23	26	10	14					
Участок 3	17	13	19	21	10	15				

Полученный результат сравните с обычной среднеарифметической и сделайте вывод.

Литература: [1], [2].

Задание для самостоятельной работы: расчет средневзвешенной концентрации кислорода

Таблица 2.3 — Вертикальное распределение концентрации кислорода

$h, \text{ м}$	0	5	10	15	25	50	75	100	150	200
$\text{O}_2, \text{ мг/л}$	12,3	11,9	11,2	10,3	7,0	5,3	5,3	4,5	4,2	4,2

Рассчитайте средневзвешенную величину, сравните со средней арифметической и сделайте вывод.

Ответьте на вопросы :

1. Что означает диапазон выборки?
2. Дайте определение средней арифметической.
3. Что является недостатком средней арифметической?
4. Напишите формулы расчета средней арифметической обычным и экспресс-

способом?

5. Как быстро рассчитать средние в Excel?

6. В каких случаях используют среднюю геометрическую?

7. Почему при разных выборках или при неравномерности взятия проб следую рассчитывать средневзвешенную, а не среднюю арифметическую?

Литература: [1], [2].

Лабораторная работа 3. Показатели вариации. Построение вариационных рядов. Мода, медиана.

Варьирование, размах изменчивости (лимиты), объем выборки, определение количества и интервалов классов, распределение частот, вариационный ряд. Дисперсия, среднее квадратичное отклонение (стандартное отклонение), коэффициент вариации. Мода, медиана.

Задание 1: построение вариационного ряда и графиков распределения.

Данные 100 измерений длины листа гибискуса, представленные в таблице 1.2, перенесите в *Excel* в виде единого ряда и ранжируйте по возрастанию. Определите крайние значения — лимиты (*lim*). Для установления характера распределения изучаемых признаков установите количество классов и длину классового интервала *dx*.

Оптимальное число классов (*k*) размерного ряда можно определить по эмпирическим формулам:

$$k = 1 + 1,44 \times \ln n$$
$$Lim = x_{max} - x_{min}$$
$$dx = lim/k$$

Установите границы классов и разнесите частоту встречаемости α (таблица 3.1).

Таблица 3.1 — Определение частот различных классов

Размерные классы	Центр классового интервала	Подсчёт частот методом конверта	Частоты, α
Σ			

Процесс разности можно значительно ускорить, если в ранжированном по возрастанию ряде выделить разным цветом соответствующие классовые размерные диапазоны и подсчитать в каждом количество вариантов.

Полученные данные представьте графически — или в виде полигона частот, в этом случае для построения графика используем в качестве данных центр классового промежутка и значения соответствующих частот или строим гистограмму — в этом случае для построения графика используем в качестве данных размерные классы и значения соответствующих частот.

Сделайте вывод.

Задание 2: Определение дисперсии, среднего квадратичного отклонения, коэффициента вариации.

1). По данным таблицы 1.2 рассчитать дисперсию по следующему алгоритму:

Рассчитать среднюю арифметическую \bar{X} . Затем определить величину отклонения x_i от средней арифметической \bar{X} . Сумма отклонений от средней по определению равна 0. Дальнейшие операции с 0 бессмысленны, поэтому величину отклонения возвести в квадрат $(x_i - \bar{X})^2$ и рассчитать сумму квадратов отклонений от средней для расчёта дисперсии. Далее по формуле определить дисперсию:

$$S^2 = \frac{\Sigma(x_i - \bar{X})^2}{n-1}$$

где

S^2 — дисперсия;

n — объем выборки;

x_i — варианты значения;

\bar{X} — средняя арифметическая;

$x_j - \bar{X}$ — отклонение варианты от средней арифметической.

$(x_j - \bar{X})^2$ — квадрат отклонения варианты от средней арифметической

$\sum(x_j - \bar{X})^2$ — сумма квадратов отклонений.

2). Рассчитать среднее квадратичное отклонение s путем извлечения квадратного корня из дисперсии.

3). Рассчитать коэффициент вариации CV

$$CV = \frac{s}{M} \cdot 100\%$$

4). Определить медиану

$$Me = \frac{x_{min} + x_{max}}{2}$$

5). Исходя из определения моды, по данным таблицы 3.1 укажите величину моды.

Литература: [1], [2], [3], [4], [6].

Задание для самостоятельной работы по теме 3: показатели вариации, построение вариационных рядов, мода, медиана.

Ответьте письменно на вопросы:

1. Каковы причины варьирования?
2. Почему разброс значений выборки не бесконечно широк, он ограничен неким диапазоном и тяготеет к определенному общему значению?
3. Что подразумевается под термином «случайная величина»?
4. Какую информацию несет термин «частота встречаемости»?
5. Что демонстрирует вариационный ряд?
6. Дайте определение дисперсии.
7. Что характеризует среднее квадратичное (стандартное) отклонение?
8. В каких случаях предпочтительно использовать коэффициент вариации?
9. Дайте определение медианы и моды.

Литература: [1], [2], [4], [6].

Лабораторная работа 4: Основные типы распределения признаков.

Основные характеристики нормального распределения. Симметричность распределения. Кривая Гаусса. Построение графиков нормального распределения. Сходство и отличие биномиального распределения в сравнении с нормальным. Распределение Пуассона. Описание редких событий законом Пуассона.

Задание 1: свойства распределения

Дать письменные ответы на следующие вопросы:

1. Перечислите свойства нормального распределения.
2. Опишите характеристики нормального распределения.
3. Во сколько раз среднее квадратичное отклонение укладывается в размах изменчивости признака при нормальном типе распределения?
4. Какими двумя параметрами определяется точная форма нормального распределения (характерная «колоколообразная кривая»)?
5. Дайте характеристику биномиального распределения.
6. В чем заключается основное отличие от нормального распределения?
7. Дайте характеристику распределения Пуассона.
8. Какие события описывает распределение Пуассона?
9. Как при сравнении дисперсии и средней арифметической можно диагностировать распределение Пуассона?

Задание 2: Определение типа распределения и визуализация распределения на графике

На основе данных таблицы 1.2 определите тип распределения, сопоставив среднюю арифметическую, стандартное отклонение, визуализируйте это на графике распределения.

Литература: [1], [2], [4], [6].

Лабораторная работа 5: Оценка нормальности: асимметрия и эксцесс.

Нормальное распределение как идеально симметричное относительно среднего; нулевая асимметрия нормального распределения, отрицательная асимметрия, положительная асимметрия; эксцесс как способ количественной оценки тенденции создавать значения, далекие от среднего. Диапазон ± 1 от нулевой асимметрии нормального распределения и эксцесс в пределах ± 1 от эксцесса нормального распределения как критерии достаточной нормальности для параметрических тестов. Коэффициент эксцесса (коэффициент островершинности Э). Нулевой эксцесс, положительный эксцесс, отрицательный эксцесс.

Задание 1: определение асимметрии и эксцесса.

Коэффициент асимметрии (As) в теории вероятностей — величина, характеризующая асимметрию распределения данной случайной величины. Английский статистик К. Пирсон на основе разности между средней величиной и модой предложил показатель асимметрии:

$$As_{\Pi} = \frac{\bar{x} - Mo}{\sigma},$$

где \bar{x} , — средняя арифметическая, Mo — мода, σ или s — стандартное отклонение.

При симметричном распределении относительно математического ожидания $As = 0$.

По данным таблицы 1.2 рассчитайте коэффициент асимметрии Пирсона и дайте характеристику симметричности рассчитанного распределения.

Коэффициент эксцесса (коэффициент островершинности Э) показывает, насколько необычно высок и заострен или, наоборот, низок и закруглен пик распределения. У распределения с положительным эксцессом высокий, заостренный пик; у распределения с отрицательным эксцессом низкий, пологий пик. У идеального нормального распределения нулевой эксцесс.

По данным таблицы 1.2 рассчитайте эксцесс, пользуясь статистическими функциями в *Excel*, и сделайте вывод.

На основании рассчитанных средней арифметической, стандартного отклонения, моды, показателей асимметрии и эксцесса дайте характеристику типа распределения.

Литература: [1], [4], [6].

Задание для самостоятельной работы по теме 5

Ответить письменно на вопросы:

1. В каком случае коэффициент асимметрии равен нулю?
2. Каков будет коэффициент асимметрии, если правый хвост распределения длиннее левого?
3. Каков будет коэффициент асимметрии, если левый хвост распределения длиннее правого?
4. Что означает наличие положительного эксцесса?
5. О чем свидетельствует нулевой эксцесс?
6. Имеется ли сходство фактического распределения с нормальным, если показатели асимметрии и эксцесса не превышают ± 1 от нулевых значений?

Литература: [1], [4], [6].

Лабораторная работа 6: Статистическая ошибка (ошибка репрезентативности выборочных параметров)

Выборочные оценки параметров генеральной совокупности. Статистическая ошибка параметра (ошибка репрезентативности выборочных параметров). Ошибка стандартного отклонения. Ошибка коэффициента вариации. Расчёт ошибок.

Отличия значений выборочных параметров от генеральных называются *ошибкой репрезентативности данного параметра*, или просто (статистической) ошибкой. При увеличении объема выборки ошибки репрезентативности стремятся к нулю (следствие закона больших чисел). Численно выраженные статистические ошибки служат мерой тех пределов, в которых выборочные параметры могут отклоняться от значений генеральных параметров.

Величина ошибки (m) тем больше, чем больше варьирование признака (S) и чем меньше выборка (n). Ошибку репрезентативности имеют все статистические параметры, рассчитанные по выборке. Для практики статистического оценивания разработаны специальные формулы.

Задание 1: Расчет ошибок средней арифметической, стандартного отклонения и коэффициента вариации.

По данным таблицы 1.2 (см. тему 1) рассчитайте среднюю и ошибку средней ($M \pm m$), стандартное отклонение и ошибку стандартного отклонения ($S \pm m_s$), коэффициент вариации и ошибку коэффициента вариации ($CV \pm m_{CV}$).

Ход работы:

1). Определяем среднюю арифметическую, стандартное отклонение и коэффициент вариации.

2). Определяем ошибку средней арифметической (\bar{X} или M):

Если $n \geq 30$, то делим на n

$$m_{\bar{x}} = \pm \frac{s}{\sqrt{n}}$$

Если $n \leq 30$, то делим на $n-1$

$$m_{\bar{x}} = \pm \frac{s}{\sqrt{n-1}}$$

3). Определяем ошибку среднего квадратичного отклонения s (стандартного отклонения):

$$m_s = \pm \frac{s}{\sqrt{2n}}$$

4). Определяем ошибку коэффициента вариации CV :

$$m_{CV} = \frac{CV}{\sqrt{2n}}$$

Вычисления можно ускорить, используя статистические функции в *Excel*.

Литература: [1], [2], [4], [6].

Задание для самостоятельной работы по теме 6

1) По результатам 15 промеров: 12, 14, 13, 16, 11, 9, 13, 15, 15, 11, 14, 14, 10, 12, 17 определить ошибку средней арифметической, ошибку стандартного отклонения и ошибку коэффициента вариации.

Сделать вывод и охарактеризовать степень однородности выборки.

Ответить письменно на следующие вопросы:

1. Что означает понятие «ошибка выборки»?

2. Что означает понятие «ошибка регистрации»?

3. От чего зависит величина ошибки репрезентативности?

4. Как называют отличие значений выборочных параметров от генеральных?

5. От каких основных показателей зависит величина ошибки выборочного наблюдения?

Литература: [1], [2], [4], [6].

Лабораторная работа 7: Доверительный интервал. Определение точности опыта. Оптимальный объем выборки.

Доверительный интервал, уровень значимости, определение точности опыта, оптимальный объем выборки.

Доверительный интервал. Параметры генеральной совокупности практически всегда остаются неизвестными, о них судят по выборочным оценкам, используя для этого значения ошибок репрезентативности. Теоретические исследования поведения выборочных средних (как случайных величин) показали, что они подчиняются нормальному закону, большинство из них (95%) находится поблизости от генеральной средней — в диапазоне $M_{ген.} \pm 1,96m$ (приблизительно $\pm 2m$).

Доверительный интервал — интервал значений изучаемого признака, в котором с той или иной вероятностью P находится значение генерального параметра.

Поэтому можно делать обратное заключение — генеральная средняя находится в диапазоне $M_{выбор.} \pm 1.96m$, т. е. предсказывать ширину интервала, в котором заключен генеральный параметр, давать интервальную оценку генеральному параметру. В соответствии с законом нормального распределения можно ожидать, что генеральный параметр (истинное значение) окажется в интервале от $M - tm$ до $M + tm$, где m — ошибка средней арифметической, t — квантиль распределения Стьюдента (табл. 6П) при данном числе степеней свободы (df) и уровне значимости (обычно $\alpha = 0,05$).

Следовательно, с вероятностью P можно ожидать, что генеральная средняя находится в доверительном интервале $M \pm tm$, построенном вокруг выборочной средней арифметической M .

Задание 1: расчет доверительного интервала.

По данным таблицы 7.1 рассчитать $M (\bar{X})$, m и s . Значение критерия Стьюдента определить по таблице 6П [2]: для уровня значимости $\alpha = 0,05$ найти значение t и дать интервальную оценку генеральному параметру: $M \pm tm$

Таблица 7.1 — Длина раковины моллюска, мм

n	x_i								
1	86	11	87,5	21	93,4	31	97,3	41	94
2	86,5	12	88,6	22	89,9	32	88	42	94,5
3	85	13	88,7	23	90	33	89,5	43	89,2
4	85,5	14	88,7	24	90	34	93,1	44	88,5
5	87	15	95,5	25	90,4	35	93,3	45	92
6	91,6	16	92,1	26	90,5	36	94,4		
7	88	17	92,3	27	89,7	37	89,9		
8	87,4	18	90,5	28	89,8	38	93,7		
9	96	19	90,5	29	91	39	94		
10	87,5	20	89,9	30	96,8	40	94,1		

Аналогичным образом рассчитайте доверительный интервал для:

- стандартного отклонения ($S \pm tmS$),
- коэффициента вариации ($CV \pm tmCV$).

Сделать вывод.

Литература: [1], [2].

В практике анализа используется относительная ошибка измерений — «показатель точности опыта» — отношение ошибки средней к самой средней арифметической, выраженное в процентах: $\varepsilon = \frac{m}{M} \times 100\%$.

Точность считается хорошей, если ε меньше 3%, и удовлетворительной при $3 < \varepsilon <$

5%. Иначе приходится собирать дополнительный материал.

Задание 2: Определение точности опыта.

По данным таблицы 7.1 определить существует ли необходимость дополнительного сбора материалов. Сделать вывод об оптимальности объема выборки.

Литература: [1], [2].

Задание 3: Определение оптимального объема выборки для обеспечения хорошей точности.

В исследованиях часто заранее требуется установить число наблюдений, достаточное для получения репрезентативных оценок генеральной совокупности. Для непрерывных признаков метод состоит в том, чтобы, используя известные соотношения между средней, стандартным отклонением, ошибкой средней, плотностью вероятности распределения Стьюдента, найти число степеней свободы, соответствующее доверительному интервалу для средней при уровне значимости $\alpha = 0,05$. Объем выборки, достаточной для получения результата заданной точности, находят по формуле: $n = \left(\frac{t \times CV}{\varepsilon}\right)^2$, где n — объем выборки; квантиль распределения Стьюдента (таблица 6П в [2]) при данном числе степеней свободы ($df=n-1$) и уровне значимости (обычно $\alpha = 0,05$); CV — приблизительное значение коэффициента вариации (%); ε — планируемая точность оценки (погрешности) (%).

По данным таблиц 1.2, 2.1 (см. темы 1 и 2) определить существует ли необходимость дополнительного сбора материалов. Сделать вывод об оптимальности объема выборки.

Задание для самостоятельной работы по теме 7

Ответить письменно на следующие вопросы:

1. Охарактеризуйте отношение выборочной средней к генеральной средней.
2. Что означает понятие «интервальная оценка генерального параметра»?
3. Как определяется доверительный интервал для выбранного уровня значимости?
4. Как рассчитывается относительная ошибка измерений — «показатель точности опыта»?
5. Как определить объем условной выборки для обеспечения хорошей точности?

Литература: [1], [2].

Лабораторная работа 8: Сравнение средних величин. Критерии сравнения (рассчитана на два занятия).

t -критерий Стьюдента, равенство дисперсий, критерий Фишера.

В терминах статистики отличия между средними могут иметь два противоположных источника:

1. обе выборки взяты из одной генеральной совокупности, но средние отличаются в силу ошибки репрезентативности.
2. выборки взяты из разных генеральных совокупностей, отличие средних вызвано в основном действием разных доминирующих факторов (а также и случайно).

Статистическая задача состоит в том, чтобы сделать обоснованный выбор. Для исключения чужеродных («выскакивающих») вариант применяют закон нормального распределения: в диапазоне четырех стандартных отклонений, $M \pm 1.96 \cdot S$, отклонение вариант от средней происходит по случайным причинам; за границами этого диапазона лежат чужеродные для данной выборки значения.

Сравнение различия средних. Для определения достоверности различия средних величин используют t -критерий Стьюдента при сравнении независимых выборок и сравнении связанных совокупностей (например, до и после воздействия), в том числе изменчивость коэффициентов вариации по формуле:

$$t = \frac{|M_1 - M_2|}{\sqrt{m_1^2 + m_2^2}} \sim t_{(\alpha, df)}$$

Разность средних берут по модулю без учета знака. Полученное этим способом значение критерия t Стьюдента сравнивают с табличным при выбранном уровне значимости (обычно для $\alpha = 0,05$) и числе степеней свободы (объемы выборок без числа ограничений, $df = n_1 + n_2 - 2$). Табличные данные представлены в приложении 6П в [2]. Результатом такого сравнения должен стать один из двух вариантов следующего статистического вывода. Если полученное значение (величина) критерия больше табличного, значит, различия между параметрами при заданном уровне значимости и установленном числе степеней свободы достоверны. Если же полученная величина критерия меньше табличной, то при данном уровне значимости и числе степеней свободы различия между параметрами недостоверны.

Задание 1: Расчет t-критерия Стьюдента

1) В процессе исследований было установлено, что у стариков до лечения инсулином среднее содержание белков в крови составляло $81,04 \pm 1,7$, а после лечения — $79,33 \pm 1,6$. Рассчитать t-критерий Стьюдента и определить влияние инсулина на содержание белков в крови.

Коэффициенты вариации также можно использовать для сравнения изменчивости разных показателей. Достоверность отличий коэффициентов оценивается с помощью критерия Стьюдента по формуле:

$$t = \frac{CV_1 - CV_2}{\sqrt{m_1^2 + m_2^2}} \sim t_{(0,05; n_1+n_2-2)}$$

где CV_1, CV_2 и m_1, m_2 – значения и ошибки коэффициентов вариации.

Вывод о достоверности отличий делается в том случае, если рассчитанное значение превысит табличное при заданном уровне значимости $\alpha = 0,05$ и числе степеней свободы $df = n_1 + n_2 - 2$ (приложение 6П в [2]).

2) Рассчитать изменчивость показаний радиоактивности почвы по коэффициентам вариации (табл. 8.1) и сделайте вывод о достоверности различия.

Таблица 8.1 — Радиоактивность образцов почвы

	мкР/час															<i>n</i>
Участок 1	25	21	32	28	31	20	22	26	25	27	21	32	25	24	27	14
Участок 2	25	30	32	28	32	23	22	26	25	26	24	32	25	24		13

Задание 2: Равенство дисперсий, критерий Фишера.

В случае применения двухвыборочного критерия для независимых выборок также необходимо соблюдение условия равенства дисперсий. Наиболее точным методом определения достоверности различий между выборочными дисперсиями служит критерий F Фишера в форме отношения дисперсий (большее значение должно стоять в числителе):

$$F = \frac{S_1^2}{S_2^2} \sim F_{(\alpha, df_1, df_2)}$$

где $S_1 > S_2$, $df_1 = n_1 - 1$, $df_2 = n_2 - 1$.

Если полученная величина F больше табличного значения при принятом уровне значимости (табл. 7П для $\alpha = 0,05$ и табл. 8П для $\alpha = 0,01$ в [2]) и числе степеней свободы (df_1 и df_2), то различие между дисперсиями признается достоверным; если она меньше, то расхождение между ними может считаться несущественным, случайным, т. е. нулевая гипотеза не отвергается.

1) Объясните последовательность действий, необходимых для решения этого задания.

2) По данным таблицы 8.1, определите степень достоверности различия между участками 1 и 2 по критерию F Фишера и подтвердите или опровергните вывод, полученный с использованием критерия Стьюдента.

Литература: [1], [2].

Задание для самостоятельной работы по теме 8

Задание 1. Ответить письменно на следующие вопросы:

1. Почему важно определить, что данное различие между выборками характерно для всей генеральной совокупности?
2. Какие наиболее распространенные методы разработаны для определения статистической значимости различий?
3. Какие статистические критерии достоверности или случайности отличий самые распространенные?
4. Какие характеристики используют при сравнении по уровню изменчивости показателей?
5. Как формулируется нулевая гипотеза?
6. При сравнении выборочных параметров нормального и биномиального распределений используется одна и та же формула?

Литература: [1], [2].

Задание 2. Сравнение двух популяций красной полевки по плодовитости

Первая (горная) популяция красной полевки: $M_1 = 5,8$; $S_1 = 1,82$; $n_1 = 27$.

Вторая (равнинная) популяция красной полевки: $M_2 = 5,4$; $S_2 = 0,52$; $n_2 = 12$.

Выяснить достоверность отличия популяций по величине плодовитости по критерию Стьюдента и по критерию Фишера и сделать вывод.

Какова последовательность действий при решении этой задачи?

Задание 3. Определение статистической значимости испытаний эффективности нового препарата железа

Для изучения эффективности нового препарата железа были выбраны две группы пациентов с анемией. В первой группе пациенты в течение двух недель получали новый препарат, а во второй группе — получали плацебо. После этого было проведено измерение уровня гемоглобина в периферической крови. Данные представлены в формате $M \pm m$

Опыт: $M_1 = 115,4 \pm 1,2$ г/л; $n=34$.

Контроль: $103,7 \pm 2,3$ г/л; $n=40$.

Необходимо сделать вывод о статистической значимости полученных различий и эффективности нового препарата железа.

Литература: [2], [3], [5].

Раздел 2. Виды анализа и функциональные зависимости.

Лабораторная работа 9: Корреляционный анализ (рассчитана на два занятия)

Корреляция, корреляционный анализ, поле корреляции, коэффициент корреляции, функциональная зависимость, ошибка коэффициента корреляции, коэффициент детерминации.

Установление направления и формы связи между признаками и измерение ее тесноты. Корреляция — это взаимная связь (взаимная зависимость) двух признаков при их изменчивости, т. е. сопряженность их вариации. Корреляция имеет место в тех случаях, когда признаки изменяются не автономно, а согласованно. Функциональной называется однозначная зависимость между переменными величинами, когда определенному значению одной (независимой) переменной x , называемой аргументом, соответствует определенное значение другой (зависимой) переменной y , называемой функцией.

Коэффициент корреляции рассчитывается только для *линейной* связи.

Коэффициент корреляции призван численно выражать долю сопряженной вариации двух признаков в общей их вариации:

$$r = \sqrt{\frac{\text{ковариация}}{\text{изменчивость}}} = \frac{C_{xy}}{\sqrt{C_x \times C_y}} = \frac{\sum(y - M_y) \times (x - M_x)}{\sqrt{\sum(y - M_y) \times \sum(x - M_x)}}$$

где C_{xy} — характеристика сопряженной изменчивости признаков,
 C_x и C_y — характеристика общей изменчивости признаков.

Положительная и отрицательная корреляция и графическое выражение их связи. Коэффициент корреляции — это численное выражение доли сопряженной вариации двух признаков в общей их вариации. *Корреляционный анализ используется только при линейной зависимости.*

Для небольших выборок его можно быстро найти и при ручном счете. Рабочая формула для расчетов имеет вид:

$$r = \frac{C_{xy}}{\sqrt{C_x \times C_y}} = \frac{\sum xy - (\sum x \times \sum y) / n}{\sqrt{(\sum x^2 - \frac{(\sum x)^2}{n}) \times (\sum y^2 - \frac{(\sum y)^2}{n})}}$$

Задание 1: расчет коэффициента корреляции вручную.

Составить таблицу для ручного счета, форма приведена ниже:

Таблица 9.1 — Алгоритм для расчета коэффициента корреляции вручную

№	x	y	x ²	y ²	xy
1	3,4	4			
2	4,2	5			
3	5,2	15			
4	4,0	12			
5	6,1	24			
6	7,0	41			
7	7,4	37			
8	8,2	50			
9	10,0	61			
∑					

а) По таблице рассчитать квадраты вариант и их произведения, а также суммы вариант, квадратов и произведений. По формуле 19 рассчитать коэффициент корреляции r .

В практической деятельности, когда число коррелируемых пар признаков X и Y невелико, то при оценке зависимости между показателями используется следующую градацию:

- *высокая* степень связи — значения коэффициента корреляции находится в пределах от 0,7 до 0,99;
- *средняя* степень связи — значения коэффициента корреляции находится в пределах от 0,5 до 0,69;
- *слабая* степень связи — значения коэффициента корреляции находится от 0,2 до 0,49.

Определить силу связи рассчитанного коэффициента r .

Литература: [2], [3], [5].

Задание 2: расчет коэффициента корреляции с использованием статистических функций в программе Excel

Задание 1: расчет коэффициента корреляции с использованием статистических функций в программе Excel.

а) Рассчитать по этим же данным величину коэффициента корреляции в программе Excel, используя функцию КОРРЕЛ и сравните результаты.

б) Методы определения коэффициента корреляции и формулы. Вычисление ошибки коэффициента корреляции:

$$\begin{aligned} \text{при } n < 10 & \quad m = \sqrt{\frac{1-r_{xy}^2}{n-2}} \\ \text{при } n > 10 & \quad m = \frac{1-r_{xy}^2}{\sqrt{n-1}} \\ \text{при } n > 100 & \quad m = \frac{1-r_{xy}^2}{\sqrt{n}} \end{aligned}$$

Выбрать нужную формулу и рассчитать ошибку.

г) Проверить значимость коэффициента критерием Стьюдента:

$$t_r = \frac{r}{m_r}$$

В нашем примере для уровня значимости $\alpha = 0,05$ и числа степеней свободы $df = n-2$ найти табличное значение критерия Стьюдента (таблица 6П в [2]), сравнить с рассчитанной величиной t и сделать вывод о достоверности связи.

Задание 3: определение ошибки аппроксимации.

Коэффициент детерминации рассматривают, как правило, в качестве основного показателя, отражающего меру качества регрессионной модели, описывающей связь между зависимой (y) и независимыми (x) переменными модели.

В регрессии коэффициент детерминации R^2 является статистической мерой того, насколько хорошо предсказания регрессии приближаются к реальным точкам данных. R^2 из 1 указывает, что предсказания регрессии идеально соответствуют данным. *Коэффициент детерминации указывает тесноту связи при нелинейной зависимости. При линейной связи коэффициент детерминации равен квадрату коэффициента корреляции ($r = \sqrt{R^2}$).*

Коэффициент детерминации изменяется в диапазоне от 0 до 1. Если он равен 0, это означает, что связь между переменными регрессионной модели отсутствует.

На практике, если коэффициент детерминации близок к 1, это указывает на то, что модель работает очень хорошо (имеет высокую значимость).

Для оценки качества однофакторной модели используют коэффициент детерминации и среднюю ошибку аппроксимации. Аппроксимация, или приближение — научный метод, состоящий в замене одних объектов другими, в том или ином смысле близкими к исходным, но более простыми. *Средняя ошибка аппроксимации (\bar{A}) определяется как среднее значение полученных значений от фактических:*

$$\bar{A} = \frac{1}{n} \sum \left| \frac{y_i - \bar{y}}{y_i} \right| \times 100\%$$

Допустимая ошибка аппроксимации не должна превышать 10%.

По данным таблицы 9.1 построить график зависимости, выбрать характер зависимости, показать линию тренда, величину коэффициента детерминации и рассчитать ошибку аппроксимации.

Контрольные вопросы:

2. Какой тип зависимости вы получили?
3. Насколько хорошо расположение точек соответствует линии тренда? Привести значение R^2 .
4. При вычислении средней ошибки аппроксимации какова должна быть последовательность действий?
5. Какова величина ошибки аппроксимации?

Сделать вывод.

Литература: [2], [3], [5].

Задание для самостоятельной работы: дать письменные ответы на следующие вопросы

1. Дайте определение корреляции.
2. Что означает согласованность изменения признаков?
3. Как соотносится корреляционная и функциональная зависимость?

4. Поясните, как изменяются признаки при положительной корреляции и как — при отрицательной корреляции?
5. Что означает коэффициент детерминации?
6. Каков диапазон изменения коэффициента корреляции?
7. Какая связь считается функциональной?
8. Назовите границы тесной (сильной), средней, слабой связи.
9. Как проводится проверка значимости коэффициента критерием Стьюдента?
10. Что представляет собой поле корреляции?
11. По виду корреляционного поля можно судить о наличии или отсутствии связи и ее типе?

Литература: [2], [3], [5].

Лабораторная работа 10: Регрессионный анализ (рассчитана на два занятия).

Регрессия, линейная регрессия, нелинейная (криволинейная) регрессия, графики рассеяния, степенная функция, экспоненциальная функция, логарифмическая функция, полиномиальная функция, коэффициент детерминации.

Регрессия в статистике применяется для обозначения корреляционной зависимости. Коэффициент корреляции указывает лишь на степень (тесноту) связи в изменчивости двух переменных величин, но не позволяет судить о том, как меняется одна величина по мере изменения другой. Ответ на этот вопрос дает вычисление коэффициента регрессии, показывающего, на какую величину в среднем изменяется один признак при изменении другого на единицу измерения. Регрессионный анализ, в отличие от корреляционного, изучает эффект влияния одного признака на другой, зависимость признака от фактора, характер влияния фактора на признак. Его основные результаты таковы:

1. таблица дисперсионного анализа, в которой показаны сила и достоверность влияния на признак изучаемого фактора или другого признака.
2. уравнение регрессии, выражающее пропорциональность сопряженного изменения признаков, тенденции их взаимосвязанной изменчивости или динамики.
3. оценки значимости коэффициентов уравнения регрессии.

Регрессионный анализ методически ориентирован односторонне — на изучение зависимости одного признака от другого (зависимость y от x или, напротив, зависимость x от y), хотя может применяться к случаям, когда фактически имеется взаимозависимость двух переменных.

Весь арсенал средств, применяемых для описания корреляционных связей, составляет содержание регрессионного анализа. Для выражения регрессии служат корреляционные уравнения или уравнения регрессии, эмпирические или теоретически вычисленные ряды регрессии, их графики, называемые линиями регрессии, а также коэффициенты линейной и нелинейной регрессии. Уравнение регрессии выражает пропорциональность сопряженного изменения признаков, тенденции их взаимосвязанной изменчивости или динамики.

Оценка значимости коэффициента регрессии по критерию t Стьюдента. Расчет ошибки коэффициента регрессии m_a . *Нелинейная* регрессия как способ описания связи признаков, изменяющихся либо с разной скоростью, либо в разных масштабах. Выбор функций для графиков рассеяния. Подбор формы зависимости: степенная, экспоненциальная, логарифмическая, полиномиальная. Коэффициент детерминации.

Литература: [1], [2], [4], [6].

1. Линейная регрессия

Регрессионный анализ, в отличие от корреляционного, изучает эффект влияния одного признака на другой, зависимость признака от фактора, характер влияния фактора на признак. Регрессия — изменение среднего уровня одного признака при изменении другого.

Судить о том, как меняется одна величина по мере изменения другой, позволяет

коэффициент регрессии (a), показывающий, на какую величину в среднем изменяется один признак (y) при изменении другого (x) на единицу измерения. Рассмотрим простейший случай, когда аргумент один. Уравнение линии линейной регрессии:

$$Y = a + bX$$

где a — свободный член, а параметр b (коэффициент регрессии) определяет наклон линии регрессии по отношению к осям координат.

Задание 1

Допустим, получили эмпирические ряды данных X и Y и требуется выявить форму зависимости. До начала вычислений построить в *Excel* по данным таблицы 10.1 график зависимости y от x и провести линию тренда, видно, что точки довольно хорошо выстраиваются в прямую, но из-за ошибок измерений имеются отклонения от прямой.

Таблица 10.1 — Результаты измерений длины (x) и ширины панциря крабов (y), см

x	25	26	31	32	34	38	38	16	22	12
y	35,2	37,6	30	40	48,4	52,8	55,5	25	26,5	9,3

Построить график распределения по первичным данным, провести линию тренда, вставить формулу зависимости и коэффициент детерминации. Проанализировать расположение точек относительно линии тренда.

Контрольные вопросы

1. Дайте определение регрессии.
2. В каких случаях рассчитывают линейную регрессию?
3. К какому виду относятся графики, описывающие зависимость одного признака от другого?
4. Какими функциями описываются зависимости в нелинейной регрессии?

Литература: [1], [2], [4], [6].

Нелинейная регрессия

Нелинейные связи между переменными величинами встречаются гораздо чаще, чем линейные. Они отражают те или иные биологические закономерности и могут быть выражены соответствующими формулами и распределением поля точек в системе прямоугольных координат.

Полиномиальная зависимость между X и Y выражается уравнением параболы второго порядка $y = a + bx + cx^2$; параболы третьего порядка $y = a + bx + cx^2 + dx^3$.

Зависимости гиперболического типа описываются уравнением гиперболы первого порядка $y = a + b/x$, гиперболы 2-го порядка $y = a + b/x^2$, гиперболы 3-го порядка $y = a + b/x^3$.

Уравнение экспоненциального типа описывается уравнением $y = ae^{xb}$.

Уравнение показательного типа описывается уравнением $y = ab^x$.

Уравнение степенного типа описывается уравнением $y = ax^b$, которое в результате логарифмирования превращается в уравнение прямой линии $lg y = lg a + b lg x$.

Уравнение логистической кривой (уравнение Ферхюльста) $y = N/(1+x^{a+bt}) + c$. Для биолога представляет значительный интерес логистическая зависимость между переменными величинами. Зависимость такого рода встречается во многих случаях, например, изменение численности популяции в замкнутой среде или скорость роста позвоночных.

Задание 1

По данным таблицы 10.2 постройте график, выявите форму зависимости, покажите коэффициент детерминации.

Таблица 10.2 — Сравнение продолжительности трудового стажа (x) и средней производительности труда (y)

x	11	12	13	13	14	14,5	16	17	18	20
y	5	7	8	8	9	9	8	8	6	6

Задание 2

По данным таблицы 10.3 постройте график, выявите форму зависимости, покажите коэффициент детерминации.

Таблица 10.3 — Сравнение плотности расположения предприятий на загазованность воздуха

x	1	2	2	3	3	4	5	5	6	7	7	8	8
y	2	4	4	5	10	16	22	32	64	128	163	192	256

Литература: [2], [3], [5].

Задание для самостоятельной работы по теме 10: регрессионный анализ

Ответить письменно на следующие вопросы:

1. Дайте определение регрессии.
2. В каких случаях рассчитывают коэффициент корреляции и в каких — коэффициент детерминации?
3. К какому виду относятся графики, описывающие зависимость одного признака от другого?
4. Какими функциями описываются зависимости в нелинейной регрессии?
5. Как рассчитывается ошибка коэффициента аппроксимации?

Литература: [1], [2], [4], [6].

Лабораторная работа 11: Метод наименьших квадратов (МНК). Выравнивание эмпирических рядов регрессии (рассчитана на два занятия).

Метод наименьших квадратов (МНК) как способ расчета коэффициентов уравнения *линейной* регрессии. Рассчитать коэффициенты уравнения регрессии позволяет метод наименьших квадратов, основная идея которого состоит в том, чтобы линия регрессии прошла на наименьшем удалении от каждой точки, т. е. чтобы сумма квадратов расстояний от всех точек до прямой линии была наименьшей. В математической статистике показано, что для случая двумерного нормального распределения лучшей (эффективной, несмещенной и пр.) линией, описывающей зависимость одного признака от другого, может быть только линия частных средних арифметических.

Эмпирические данные всегда содержат ошибку измерений и выравнивание рядов необходимо для расчёта теоретических значений функции. Выравнивание производится с помощью метода наименьших квадратов (МНК). Метод наименьших квадратов — метод оценки параметров модели на основании экспериментальных данных, содержащих случайные ошибки. В основе метода лежат следующие рассуждения: при замене точного (неизвестного) параметра модели приблизительным значением необходимо минимизировать разницу между экспериментальными данными и теоретическими (вычисленными при помощи предложенной модели). Это позволяет рассчитать параметры модели с помощью МНК с минимальной погрешностью.

Мерой разницы в МНК служит сумма квадратов отклонений действительных (экспериментальных) значений от теоретических. Полученные с помощью МНК параметры модели являются наиболее вероятными.

Задание 1: выравнивание рядов методом наименьших квадратов (МНК)

Выравнивание рядов производим по следующей схеме (табл. 11.1):

Таблица 11.1 — Алгоритм расчета вспомогательных величин для определения коэффициентов линейной регрессии a и b

n	x	y	x^2	y^2	xy	Y	$(y-Y)^2$	$t \times m_y$	$Min Y$	$Max Y$
1	25	35,2								
2	26	37,6								
3	31	30								
4	32	40								
5	34	48,4								
6	38	52,8								
7	38	55,5								
8	16	25								
9	22	26,5								
10	12	9,3								
Σ										
\bar{M}		?								

1). Найти вспомогательные величины:

$$C_x = \Sigma x^2 - (\Sigma x)^2 / n,$$

$$C_y = \Sigma y^2 - (\Sigma y)^2 / n,$$

$$C_{xy} = \Sigma(xy) - (\Sigma x \times \Sigma y / n),$$

$$M_y = \Sigma y / n,$$

$$M_x = \Sigma x / n.$$

2). Рассчитать коэффициенты a и b :

$$a = C_{xy} / C_x,$$

$$b = M_y - a \times M_x.$$

3). По формуле $y=a+bx$ рассчитать теоретические значения Y .

4). В программе *Excel*, используя функцию КОРРЕЛ, найти величину коэффициента корреляции r между x и y , рассчитать ошибку коэффициента r .

Вычисление ошибки коэффициента корреляции:

$$\text{при } n < 10 \quad m_r = \sqrt{\frac{1-r_{xy}^2}{n-2}}$$

$$\text{при } n > 10 \quad m = \frac{1-r_{xy}^2}{\sqrt{n-1}}$$

$$\text{при } n > 100 \quad m = \frac{1-r_{xy}^2}{\sqrt{n}}$$

5). В программе *Excel*, используя функцию СТАНДОТКЛОН.В найти величину стандартного отклонения для x и y и по формуле рассчитать ошибку коэффициента a .

$$m_a = \frac{S_y}{S_x} \times m_r$$

6). Для проверки значимости коэффициента регрессии использовать критерий t Стьюдента:

$$t_a = \frac{a}{m_a}$$

Для уровня значимости $\alpha = 0,05$ и числа степеней свободы $df = n - 2$ найти табличное значение критерия Стьюдента t (табл. 6П в [2]), сравнить с полученной величиной и сделать вывод о статистической значимости коэффициента регрессии (a) и достоверности его отличия от нуля.

7). Рассчитать доверительную зону (интервал), в которой с той или иной вероятностью заключены теоретические средние значения y . Для этого в таблице 6П [2] найти значения критерия Стьюдента (нормированное отклонение) для уровня значимости $\alpha = 0.05$, и числа степеней свободы $df = n - 1 = 9$ в нашем случае это составит 2,26.

Теоретические исследования поведения выборочных средних (как случайных величин) показали, что они подчиняются нормальному закону, большинство из них (95%) находится поблизости от генеральной средней — в диапазоне $M_{ген.} \pm 1,96m$ (приблизительно $\pm 2m$).

Это обстоятельство позволяет делать обратное заключение — генеральная средняя находится в диапазоне $M_{выбор.} \pm 1,96m$, т. е. предсказывать ширину интервала, в котором заключен генеральный параметр и дать интервальную оценку генеральному параметру.

В соответствии с законом нормального распределения можно ожидать, что генеральный параметр (истинное значение) окажется в интервале $M \pm tm$.

По данным таблицы 11.1 найти границу доверительного интервала:

верхняя: $maxY = Y_i + t \times m_y$

нижняя: $minY = Y_i - t \times m_y$.

Величины корреляции и регрессии считаются достоверными, если они превышают свои ошибки в определенное количество раз, зависящие от размера выборки. Критерии достоверности сравнивают со стандартами значений по таблице Стьюдента для установленного числа степеней свободы и порога вероятности безошибочных прогнозов.

Литература: [1], [2].

Задание для самостоятельной работы:

1) По данным таблицы 11.2 рассчитать коэффициенты a и b , написать уравнение регрессии.

Рассчитать коэффициент корреляции r

Построить график зависимости X и $Y_{теор.}$

Таблица 11.2 — Данные для расчета

X	2	3	4	5	6	7	8	9	10	11	12	13
Y	1,8	2,21	2,29	2,48	2,2	2,63	2,75	2,93	3,01	3,4	3,15	3,27

2) подготовка к контрольной работе, проработка пройденного материала.

Литература: [1], [2].

Контрольная работа по разделу 2.

1. При какой зависимости рассчитывают коэффициент корреляции?

2. Рассчитать коэффициенты уравнения методом наименьших квадратов зависимость между x и y , написать уравнение зависимости

x	13	24	35	46	57	67	77	88	98
y	40	50	150	120	240	410	370	500	610

Литература: [1], [2].

5 Учебно-методическое обеспечение для самостоятельной работы обучающихся

5.1. Внеаудиторная самостоятельная работа студентов

В целом внеаудиторная самостоятельная работа студента при изучении курса включает в себя следующие виды работ:

- проработка (изучение) теоретической части лабораторных работ;
- чтение и проработка рекомендованной основной и дополнительной литературы;
- подготовка к лабораторным занятиям;
- поиск и проработка материалов из Интернет-ресурсов, научных публикаций;

– выполнение домашних заданий (решение задач);
– подготовка к текущему и итоговому (промежуточная аттестация) контролю знаний по дисциплине.

Основная доля самостоятельной работы студентов приходится на подготовку к лабораторным занятиям, тематика которых полностью охватывает содержание курса.

Самостоятельная работа по разделу 1:

Работа с материалом лабораторных работ и рекомендованной литературой.

Подготовка материалов к контрольному опросу по изученным темам, лабораторным занятиям, диалогам с преподавателем и участниками проверки знаний первого дисциплинарного модуля.

Самостоятельная работа по разделу 2:

Работа с материалом лабораторных работ и рекомендованной литературой.

Подготовка материалов к контрольному опросу по изученным темам, лабораторным занятиям, диалогам с преподавателем и участниками проверки знаний второго дисциплинарного модуля.

6 Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине «Методы математической статистики в экологии» представлен в приложении к рабочей программе дисциплины и включает в себя:

– перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы;

– описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания;

– типовые контрольные задания или материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций;

– методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.

Вопросы для проведения промежуточной аттестации по дисциплине (зачет)

1. Понятие генеральной совокупности и выборки.
2. Расчет средней арифметической, средней взвешенной, средней геометрической.
3. Случаи использования средневзвешенной величины.
4. Недостаток средней арифметической, преимущество медианы.
5. Способы группировки первичных данных.
6. Показатели вариации.
7. Построение вариационных рядов. Мода. Медиана.
8. Понятия дисперсии и среднего квадратичного отклонения.
9. Коэффициент вариации.
10. Характеристика нормального распределения.
11. Отличие биномиального распределения от нормального.
12. Свойства нормального распределения.
13. Определения асимметрии и эксцесса.
14. Расчет статистических ошибок.
15. Доверительный интервал.

16. Определение оптимального объема выборки.
17. Способы сравнения показателей изменчивости.
18. Нормированное отклонение;
19. Сравнение теоретических частот с фактическими и расчёт нормированного отклонения.
20. Оценка статистической достоверности разности;
21. Критерии t Стьюдента (сравнение средних арифметических коэффициентов вариации) и F Фишера (сравнение долей и показателей изменчивости по дисперсии);
22. Однофакторный дисперсионный анализ количественных признаков;
23. Корреляционный анализ;
24. Регрессионный анализ;
25. Дисперсионный анализ;
26. Построение графиков зависимости.
27. Оценка формы связи;
28. Линейная и нелинейная регрессия, выбор уравнений регрессии.
29. Построение теоретической кривой по уравнению регрессии.

7. Рекомендуемая литература

7.1 Основная

1. Ефимова М.Р. Общая теория статистики. — М.: Инфра-М, 2005. — 416 с. (49 экз.)

7.2 Дополнительная

2. Кремер Н.Ш. Теория вероятностей и математическая статистика: учебник. — М.: Юнити-Дана, 2000. — 543 с. (7 экз.)
3. Пузаченко Ю.Г. Математические методы в экологических и географических исследованиях: учеб. пособие. — М.: Академия, 2004. — 416 с. (8 экз.)
4. Спирина М.С. Теория вероятностей и математическая статистика: учебник. — М.: Академия, 2007. — 352 с. (5 экз.)
5. Гмурман В.Е. Теория вероятностей и математическая статистика: учеб. пособие. — М.: Высшая школа, 2003. — 479 с. (5 экз.)

8 Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

Статистика. Урок 1. Введение [Электронный ресурс]. — <https://www.youtube.com/watch?v=MFM2ZYxgzWg>

Показатели формы распределения случайной величины [Электронный ресурс]. — URL: https://edu.tltsu.ru/sites/sites_content/site216/html/media96435/lec_3.pdf

Определение асимметрии и эксцесса [Электронный ресурс]. — http://www.mathprofi.ru/asimetriya_i_excess.html

Манита А.Д. Теория вероятностей и математическая статистика (Основные разделы содержания: События и их вероятности. Дискретные случайные величины и их распределения. Общие случайные величины. Совместное распределение общих случайных величин. Предельные законы теории вероятностей. Обзор методов математической статистики. Метод наименьших квадратов. Доверительные интервалы. Статистические гипотезы. Таблицы (стандартный нормальный закон, квантили хи-квадрат распределения, квантили распределения Стьюдента). Интернет-учебник, полный текст представлен на сайте [Электронный ресурс]. <http://teorver-online.narod.ru/teorver73.html>

Основы статистики (видеоурок по основам статистики от Анатолия Карпова). Три

части по часу каждая. Использование специальной программы для расчетов и визуализации. Темы: 1. Генеральная совокупность и выборка 2. Сравнение средних 3. Корреляция и регрессия. [Электронный ресурс]. — URL: https://www.youtube.com/watch?v=ksdrNa_g11M&t=2s

Статистический метод (критерий): как выбрать для анализа? (видеоурок по основам статистики от Кирилла Мильчакова). [Электронный ресурс]. — URL: <https://www.youtube.com/watch?v=ZaZYy0YUdY8>

9 Методические указания для обучающихся по освоению дисциплины

Методика преподавания данной дисциплины предполагает проведение лабораторных занятий, групповых и индивидуальных консультаций по отдельным вопросам дисциплины. Предусмотрена самостоятельная работа студентов, а также прохождение аттестационных испытаний промежуточной аттестации.

Лабораторные занятия посвящаются рассмотрению наиболее важных концептуальных вопросов — основным понятиям и методам математической статистики, применяемым в экологических исследованиях. В ходе занятий студентам следует вести записи по теоретической части: кратко, схематично, последовательно фиксировать основные положения, выводы, формулировки, обобщения; пометить важные мысли, выделять ключевые слова, термины; проверять термины, понятия с помощью справочников с выписыванием толкований в тетрадь; обозначить вопросы, термины, материал, который вызывает трудности, пометить и попытаться найти ответ в рекомендуемой литературе. Если самостоятельно не удастся разобраться в материале, необходимо сформулировать вопрос и задать преподавателю на консультации. Уделить внимание понятиям, которые обозначены обязательными для каждой темы дисциплины.

Целью проведения лабораторных занятий является знакомство студентов с методами математической статистики для обработки результатов наблюдений и исследований и выработка умения самостоятельной статистической обработки данных и осмысления полученных результатов. На лабораторных занятиях обсуждаются теоретические вопросы по теме, разбираются конкретные задания по изучаемой теме и решаются задачи. Для подготовки к занятиям студенты выполняют проработку рабочей программы, уделяя особое внимание целям и задачам, структуре и содержанию дисциплины; конспектирование источников; просмотр рекомендуемой литературы.

В ходе групповых и индивидуальных консультаций студенты имеют возможность получить квалифицированную консультацию по организации самостоятельного управления собственной деятельностью на основе анализа имеющегося у студента опыта обучения, используемых учебных стратегий, через обсуждение сильных сторон и ограничений стиля учения, а также поиск ресурсов, предоставляемых вузом для достижения намеченных результатов; решения учебных задач, для подготовки к интерактивным лабораторным занятиям, для подготовки к контрольным точкам, в том числе итоговой; детально прорабатывать возникающие проблемные ситуации, осуществлять поиск вариантов их решения, определять преимущества и ограничения используемых средств для решения поставленных учебных задач, обнаруживать необходимость изменения способов организации своей работы и др.

При изучении дисциплины используются интерактивные методы обучения, такие как:

1. лабораторное занятие с использованием средств визуализации (демонстрации слайдов) проводится с целью акцентирования внимания обучающихся на какой-либо актуальной теме или на наиболее важных и существенных ее аспектах.

2. Самостоятельная работа студентов углубляет знания, ориентирует их на активный поиск путей и способов решения конкретной задачи с использованием ресурсов информационно-телекоммуникационной сети «Интернет».

10 Курсовой проект (работа)

Выполнение курсового проекта (работы) не предусмотрено учебным планом.

11 Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационно-справочных систем

11.1 Перечень информационных технологий, используемых при осуществлении образовательного процесса

- электронные образовательные ресурсы, представленные в п. 8 рабочей программы;
- интерактивное общение с обучающимися и консультирование посредством электронной почты.

11.2 Перечень программного обеспечения, используемого при осуществлении образовательного процесса

При освоении дисциплины используется лицензионное программное обеспечение:

- текстовый редактор Microsoft Word;
- пакет Microsoft Office
- электронные таблицы Microsoft Excel (раздел «Статистические функции»);

11.3 Перечень информационно-справочных систем

- справочно-правовая система Консультант-плюс <http://www.consultant.ru/online>
- справочно-правовая система Гарант <http://www.garant.ru/online>

12 Материально-техническое обеспечение дисциплины

Для проведения занятий лабораторных занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации используются учебные аудитории 6-505 и 6-519 с комплектом учебной мебели.

Для самостоятельной работы обучающихся используются кабинеты 6-214 и 6-314; каждый оборудован комплектом учебной мебели, двумя компьютерами с доступом в информационно-телекоммуникационную сеть «Интернет» и в электронную информационно-образовательную среду организации, принтером и сканером.

Технические средства обучения для представления учебной информации включают аудиторную доску, мультимедийное оборудование, компьютеры.

При изучении дисциплины используется библиотечный фонд КамчатГТУ: учебники, учебные пособия, электронный ресурс; раздаточный материал (таблицы для расчетов и др.).

Дополнения и изменения в рабочей программе

Дополнения и изменения в рабочей программе за ____ / ____ учебный год

В рабочую программу по дисциплине «Методы математической статистики в экологии» для направления подготовки 05.03.06 «Экология и природопользование» вносятся следующие дополнения и изменения:

Дополнения и изменения внес _____
(должность, Ф.И.О., подпись)

Рабочая программа пересмотрена и одобрена на заседании кафедры _____

«__» _____ 202__ г.

Заведующий кафедрой _____ (подпись) _____ (Ф.И.О.)