

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

Технологический факультет

Кафедра «Экология и природопользование»

УТВЕРЖДАЮ
Декан технологического
факультета

Л.М. Хорошман
«17» 03 2021 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Основы математической статистики в экологических исследованиях»

направление подготовки
05.04.06 Экология и природопользование
(уровень магистратуры)

профиль:
«Природопользование»

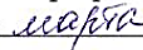

Петропавловск-Камчатский
2021

Рабочая программа по дисциплине «Основы математической статистики в экологических исследованиях» составлена на основании ФГОС ВО направления подготовки 05.04.06 «Природопользование»

Составитель рабочей программы

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

Рабочая программа рассмотрена на заседании кафедры ЭП

«16»  2021 г., протокол № 

Заведующий кафедрой

«17»  2021 г.,  Ступникова Н.А.

1. Цели и задачи учебной дисциплины

Целью дисциплины «Основы математической статистики в экологических исследованиях» является овладение методами количественной оценки получаемых результатов исследований.

Задачи дисциплины:

- овладение способами группировки первичных данных;
- изучение основных характеристик варьирующих объектов;
- изучение законов распределения;
- изучение критериев достоверности оценок;
- знакомство с корреляционным анализом;
- знакомство с регрессионным и дисперсионным анализом.

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

- умение правильно анализировать результаты наблюдений;
- способы расчёта средних величин и ошибки;
- построение кривых распределения;
- выбор уравнений регрессии, имеющих биологический смысл.
- построение графиков зависимости.
- оценка достоверности полученных результатов.

2. Требования к результатам освоения дисциплины

Процесс изучения дисциплины направлен на формирование профессиональной компетенции:

– Способность анализировать научные данные, результаты экспериментов и наблюдений (ПК-4).

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
(ПК-4)	Способность анализировать научные данные, результаты экспериментов и наблюдений	ИД-1 _{ПК-4} : Знает методы анализа научных данных	Знать: – свойства и классификацию признаков; – точность измерений; – способы группировки первичных данных; – способы расчёта средних величин; – способы расчёта показателей вариации; – способы расчёта дисперсии, среднего квадратичного отклонения; – способы расчёта коэффициента вариации; – законы распределения; – вероятность события; – закон больших чисел; – нормальное распределение; – асимметрию и эксцесс; – статистические гипотезы и их проверка;	З(ПК-4)1 З(ПК-4)2 З(ПК-4)3 З(ПК-4)4 З(ПК-4)5 З(ПК-4)6 З(ПК-4)7 З(ПК-4)8 З(ПК-4)9 З(ПК-4)10 З(ПК-4)11 З(ПК-4)12 З(ПК-4)13

Код компетенции	Наименование компетенции	Код и наименование индикатора достижения	Планируемый результат обучения по дисциплине	Код показателя освоения
		ИД-2 _{ПК-4} : Умеет оформлять результаты научно-исследовательских работ.	<ul style="list-style-type: none"> – способы расчёта критерия Стьюдента; – функциональную зависимость и корреляции; – оценку формы связи; – понятие регрессии, выбор уравнений регрессии. – дисперсионный анализ 	З(ПК-4)14 З(ПК-4)15 З(ПК-4)16 З(ПК-4)17 З(ПК-4)18
			<ul style="list-style-type: none"> – формулировать и решать задачи, возникающие в ходе научно-исследовательской и педагогической деятельности на базе углубленных профессиональных знаний; – умение формулировать и решать задачи, возникающие в ходе научно-исследовательской деятельности на базе углубленных профессиональных знаний; – применять на практике базовые общепрофессиональные знания теории и методов исследований; способность пользоваться современными методами обработки, анализа и синтеза информации; – демонстрировать владение методами статистической обработки первичного материала. – видеть диалектику связи между частью и целым, причиной и следствием. 	У(ПК-4)1 У(ПК-4)2 У(ПК-4)3 У(ПК-4)4 У(ПК-4)5

3 Место дисциплины в структуре образовательной программы

Учебная дисциплина «Основы математической статистики в экологических исследованиях» 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 «Основы математической статистики в экологических исследованиях» используются для математической обработки результатов исследования при подготовке выпускной квалификационной работы .

4. Содержание дисциплины

4.1 Тематический план дисциплины (табл. 2)

Наименование разделов и тем	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.	125	50	18	32	–	49	Контроль -ная	26
Тема 1: Понятие генеральной совокупности и выборки.	13	6	2	4	–	4	Опрос, практические задания	3
Тема 2: Расчёт средних величин и средневзвешенной.	13	6	2	4	–	4	Опрос, практические задания	3
Тема 3: Показатели вариации. Построение вариационных рядов. Мода, медиана.	15	6	2	4	–	6	Опрос, практические задания	3
Тема 4: Основные типы распределения признаков.	15	6	2	4	–	6	Опрос, практические задания	3
Тема 5: Оценка нормальности: асимметрия и эксцесс.	13	4	2	2	–	6	Опрос, практические задания	3
Тема 6: Статистическая ошибка (ошибка репрезентативности выборочных параметров).	15	6	2	4	–	6	Опрос, практические задания	3
Тема 7: Доверительный интервал. Определение точности опыта. Оптимальный объем выборки.	24	8	2	6	–	11	Опрос, практические задания	5
Тема 8: Сравнение средних величин. Критерии сравнения.	17	8	4	4	–	6	Опрос, практические задания	3
Раздел 2. Виды анализа и функциональные зависимости.	55	20	10	10	–	25	Контроль-ная	10
Тема 9: Корреляционный анализ	11	4	2	2	–	4	Опрос, практические задания	3

Тема 10: Регрессионный анализ. Метод наименьших квадратов.	22	12	6	6	–	11	Опрос, практические задания	3
Тема 11: Дисперсионный анализ	18	4	2	2	–	10	Опрос, практические задания	4
Экзамен					–			2
Всего	180	70	28	42	–	74		36

4.2 Содержание дисциплины

Раздел 1. Классификация признаков и способы группировки первичных данных.

Тема 1: Понятие генеральной совокупности и выборки.

Лекция

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

Генеральная совокупность – все варианты одного типа. Это понятие можно интерпретировать как мыслимое множество вариантов, сформированных при одинаковых (внешних и внутренних) условиях.

Теоретическая бесконечность генеральной совокупности означает, что ее никогда нельзя познать до конца, в действительности мы всегда имеем дело с выборками. Выборочная совокупность, выборка – это множество вариантов одного типа, ограниченное способом отбора (методами получения вариантов) из генеральной совокупности. Отличие выборок от генеральной совокупности состоит в том, что действующие в генеральной совокупности факторы не могут проявиться в полной мере в любой отдельной выборке. Каждая новая выборка обязательно будет отличаться от предыдущей в силу случайности, варианты новой выборки будут нести одинаковый отпечаток действия доминирующих факторов, но разные следы действия случайных факторов.

Интуитивно понятно, что чем меньше объем выборок, тем менее точным будут выборочные оценки генеральных параметров, и, напротив, чем больше выборка, тем ближе выборочные средние и дисперсии лежат к генеральным значениям. Это явление называется законом больших чисел — с ростом числа наблюдений значения выборочных параметров стремятся воспроизвести генеральные.

Основные понятия темы:

Признак, варианта, варьирование, выборка, генеральная совокупность.

Вопросы для самоконтроля:

1. Определение признака и способы его выражения.
2. Способы регистрации признаков биологических объектов.
3. Параметрические и непараметрические методы регистрации признаков.
4. Что подразумевается под понятием «отдельная варианта»?
5. Какие существуют методы классификации признаков?
6. Что означает понятие случайной величины?
7. Дайте определение выборки.
8. Охарактеризуйте понятие «генеральная совокупность».
9. Чем выборка отличается от генеральной совокупности?
10. Как с ростом числа наблюдений изменяются значения выборочных параметров по отношению к генеральной совокупности?

Литература: [1], [2].

Практическое занятие

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 (табл. 1.1).

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

	5 повторов	20 повторов	50 повторов
О			
Р			
Всего			

Сделать вывод.

Практическое занятие

Задание: определить зависимость средней арифметической от величины выборки.

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.

Таблица 1.2 — Длина листа гибискуса, см

Блок 1 n=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 n=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 n=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

Сделайте вывод о влиянии числа наблюдений на конечный результат.

Литература: [1], [2].

Тема 2: Расчёт средних величин и средневзвешенной.

Лекция

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.

Средняя арифметическая (M) — это одна из важнейших обобщающих характеристик вариационного ряда. Существует несколько видов средних (средняя арифметическая — простая и взвешенная, средняя гармоническая, средняя квадратичная), но в практике биологических исследований наибольшее значение имеет средняя арифметическая — величина, вокруг которой «концентрируются» варианты. Средняя арифметическая характеризует действие только систематических факторов, поскольку сумма случайных отклонений влево и вправо от средней в силу симметричности нормального распределения обращается в нуль. В случае нормального распределения средняя арифметическая находится точно по центру (совпадает со значением медианы).

Взвешенная средняя учитывает объемы частных выборок, когда выборки отличаются разным количеством измерений или пробы берутся не на одинаковом расстоянии. Расчет средней арифметической приводил бы к заниженным или завышенным результатам в зависимости от ситуации.

Средняя геометрическая рассчитывается для величин, изменяющихся в геометрической прогрессии.

Основные понятия темы:

Разброс значений выборки, расчёт средней арифметической, средней геометрической. Взвешенная средняя как способ определения общей средней.

Вопросы для самоконтроля:

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 следует рассчитывать средневзвешенную, а не среднюю арифметическую?

Литература: [1], [2].

Практическое занятие

Задание 1: расчет средних величин.

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.

Почему различаются значения средневзвешенной и средней арифметической?

Какую величину корректнее рассчитывать с учетом неравномерности измерений по вертикали?

Литература: [1], [2].

Практическое занятие

Задание 1: расчет средних 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:

$$\bar{M} = \frac{\sum n_j \times M_j}{\sum n_j} \quad 1$$

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	21
Участок 2	24	23	26	10	14					
Участок 3	17	13	19	21	10	15				

Полученный результат сравните с обычной среднеарифметической и сделайте вывод.

Задание 2: Расчет средневзвешенной концентрации кислорода

Таблица 2.3 — Вертикальное распределение концентрации кислорода

h, м	0	5	10	15	25	50	75	100	150	200
O ₂ , 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.

Литература: [1], [2].

Тема 3: Показатели вариации. Построение вариационных рядов. Мода, медиана

Лекция

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

Основная особенность выборки как множества значений случайной величины — это отличие отдельных вариантов друг от друга, явление изменчивости, варьирования, появления отличий между отдельными вариантами. Один из источников, эндогенный, — это индивидуальные отличия организмов по статусу и по состоянию. Другой источник отличий между вариантами — факторы внешней среды, т. е. условия проведения наблюдений, среда существования объекта, возможная причина, определяющая текущее состояние объекта. Факторы, влияющие на значения вариантов, различаются по своей природе. Если фактор влияет на все варианты выборки постоянно и примерно одинаково, он называется систематическим (или доминирующим). Если фактор непостоянен, влияет на варианты не одинаково, с разной силой, он определяется как случайный.

При самом широком варьировании признаков разброс значений выборки не бесконечно широк, он ограничен неким диапазоном и тяготеет к определенному общему значению. Эти свойства статистических совокупностей — варьирование, но в ограниченном диапазоне, — позволяют предложить для описания две группы величин: оценку центрального значения диапазона (среднюю, моду или медиану) и оценку размаха варьирования (лимит, дисперсию, стандартное отклонение). Определение этих значений выполняется после построения вариационного ряда.

Основные понятия темы:

Варьирование, лимиты, случайная величина, частота встречаемости, вариационный ряд, дисперсия, среднее квадратичное отклонение, коэффициент вариации, мода, медиана.

Вопросы для самоконтроля:

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].

Практическое занятие

Задание: построение вариационного ряда и графика 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 эмпирической формуле:

$$k = 1 + 1,44 \times \ln n \quad 2$$

$$Lim = x_{max} - x_{min} \quad 3$$

$$dx = lim/k \quad 4$$

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

Таблица 3.1 — Определение частот различных классов

Размерные классы	Центр классового интервала	Подсчёт частот	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.

Литература: [1], [2].

Практическое занятие

Задание: Вычисление параметров выборок.

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:

$$S^2 = \frac{\sum(x_i - \bar{X})^2}{n-1} \quad 5$$

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 .

3). Рассчитать коэффициент вариации CV

$$CV = \frac{s}{M} \cdot 100\% \quad 6$$

4). Определить медиану

$$Me = \frac{x_{min} + x_{max}}{2} \quad 7$$

5). Исходя из определения моды, по данным таблицы 3.1 укажите величину моды.
Литература: [1], [2].

Тема 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.

Лекция

Среди многих известных типов распределений рассмотрим 6 (нормальное, биномиальное, Пуассона, альтернативное, полиномиальное, равномерное).

Нормальное распределение. Наиболее характерный тип распределения *непрерывных* случайных величин, из него можно вывести (к нему сводятся) все остальные. Распределение симметрично, причем крайние значения (наибольшие и наименьшие) появляются редко, но чем ближе значения признака к центру (к средней арифметической), тем оно чаще встречается. Кривая нормального распределения имеет куполообразный вид (кривая Гаусса). Среднее квадратичное отклонение примерно 4 раза укладывается в размахе изменчивости признака и по величине значительно уступает средней. Геометрически стандартное отклонение равно расстоянию от центра кривой распределения до точки перегиба кривой.

Биномиальное распределение во многом близко к нормальному. Отличие состоит лишь в том, что оно характеризует поведение дискретных признаков, выраженных *целыми* числами. Как правило, для описания биологических признаков подходит симметричное биномиальное распределение, у которого дисперсия много меньше средней.

Распределение Пуассона. Это вариант описания стохастического поведения дискретных количественных признаков для случаев, когда вероятность элементарных альтернативных событий неодинакова, одно из них наблюдается заметно чаще другого ($p \ll q$). Закон Пуассона описывает редкие события, происходящие 1, 2, 3 и т. д. раз на сотни и тысячи обычных событий. Поведение биологических объектов, соответствующее закону Пуассона, наблюдается в том случае, когда по пробам случайно распределены редкие объекты. Распределение Пуассона резко асимметрично, *причем дисперсия равна средней арифметической, что может служить критерием для оценки характера распределения изучаемого признака.*

Альтернативное распределение. Распределение дискретной случайной величины, имеющей лишь два противоположных (разнокачественных) значения (два класса, $k = 2$). В одной пробе (в одном наблюдении) содержится одна варианта ($m = 1$), одно из двух возможных значений. Вычисления констант достаточно просты и не требуют построения вариационного ряда.

Полиномиальное распределение наблюдается для качественных признаков, имеющих не два альтернативных свойства, но несколько возможных проявлений качества. Примеры полиморфизма популяций – из этой области.

Равномерное распределение — это частный случай распределения альтернативного и полиномиального, характеризуется одинаковой частотой встречаемости всех значений дискретного признака ($p = q$ для двух классов или $p_1 = p_2 = \dots = p_j \dots = p_k$ для нескольких классов).

Зная тип распределения, можно воспользоваться разработанными специально для него приемами математической обработки и получить наиболее полную информацию о

явлении, точнее оценить различия между параметрами разных выборок.

Основные понятия темы:

Нормальное распределение, кривая Гаусса, биномиальное распределение, распределение Пуассона, альтернативное распределение, полиномиальное распределение, равномерное распределение.

Вопросы для самоконтроля:

1. Дайте характеристику нормального распределения.
2. Что описывает кривая Гаусса?
3. Дайте характеристику биномиального распределения. Чем оно отличается от нормального?
4. Какое распределение больше подходит для описания редких событий?
5. Как с помощью средней, медианы, дисперсии, стандартного отклонения, коэффициента асимметрии можно определить характер распределения экспресс-методом?
6. Какое распределение является наиболее характерным типом распределения непрерывных случайных величин, к которому сводятся все остальные типы?

Литература: [1], [2], [4].

Практическое занятие: свойства распределения

Вопросы для обсуждения:

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:

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

Литература: [1], [2], [4].

Практическое занятие: построение графика распределения и оценка характера распределения.

На основании данных таблицы 4.1 постройте график распределения в виде гистограммы и полигона, оцените характер распределения.

Таблица 4.1 — Радиус ствола сосен

№	R, см	№	R, см	№	R, см	№	R, см	№	R, см
1	18,6	21	19,2	41	16,4	61	17,2	81	22
2	18,6	22	17	42	20,2	62	19,8	82	22,2
3	18,8	23	17,2	43	17,2	63	17,6	83	22,4
4	18,8	24	23,2	44	20,6	64	18,8	84	22,4
5	18,8	25	20	45	18,4	65	19,6	85	22,6
6	19	26	20,4	46	21	66	19,8	86	22,8
7	17,4	27	14,6	47	18,4	67	18,6	87	14,4

8	16,2	28	17,6	48	19,2	68	18,6	88	14,8
9	19,4	29	19,4	49	18	69	18	89	22,2
10	22,4	30	19,4	50	18,4	70	17,4	90	22,6
11	16,6	31	19,6	51	16	71	19,8	91	22
12	19,4	32	19,6	52	19,6	72	23,8	92	22,2
13	18,2	33	19,6	53	18,8	73	18,2	93	22,4
14	18,2	34	19,8	54	18,6	74	19	94	22,4
15	16,8	35	19,8	55	17,6	75	20	95	22,6
16	19,2	36	18,4	56	19,6	76	19,2	96	22,8
17	19,2	37	15,2	57	17,2	77	18,6	97	22
18	19,2	38	16,2	58	20	78	18,8	98	22,2
19	19,4	39	17,6	59	17,6	79	18,2	99	19,6
20	18,4	40	23	60	19,4	80	15,8	100	17,2

Для построения графика в виде полигона частот, какие предварительные действия вы должны осуществить? Каким размером радиуса характеризуется модальная группа?

Литература: [1], [2], [4].

Тема 5. Оценка нормальности: асимметрия и эксцесс.

Лекция

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 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 (s) — стандартное отклонение.

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

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

Основные понятия темы: нулевая асимметрия, отрицательная асимметрия, положительная асимметрия; нулевой эксцесс, положительный эксцесс, отрицательный эксцесс.

Вопросы для самоконтроля:

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], [2], [3].

Практическое занятие: Характеристика асимметрии и эксцесса.

Задание 1: определение коэффициента асимметрии.

По данным таблицы 1.2 рассчитайте коэффициент асимметрии Пирсона

Дайте характеристику симметричности рассчитанного распределения.

Задание 2: определение эксцесса.

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], [2], [3].

Тема 6: Статистическая ошибка (ошибка репрезентативности выборочных параметров)

Лекция

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.

Для нормального распределения:

ошибка средней $m_M = \frac{S}{\sqrt{n}}$;

ошибка стандартного отклонения $m_s = \frac{s}{\sqrt{2n}}$;

ошибка коэффициента вариации $m_{CV} = \frac{CV}{\sqrt{2n}}$.

Основные понятия темы:

Ошибка средней, ошибка стандартного отклонения, ошибка коэффициента вариации.

Вопросы для самоконтроля:

Что означает понятие «ошибка выборки»?

Что означает понятие «ошибка регистрации»?

От чего зависит величина ошибки репрезентативности?

Как называют отличие значений выборочных параметров от генеральных?

От каких основных показателей зависит величина ошибки выборочного наблюдения?

Литература: [1], [2].

Практическое занятие; расчет ошибок

Вопросы для обсуждения:

1. Ошибка средней арифметической.
2. Ошибка стандартного отклонения.
3. Ошибка коэффициента вариации.

Задание:

По данным таблицы 1.2 (тема 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

$$m_{\bar{x}} = \pm \frac{s}{\sqrt{n}} \quad 9$$

Если $n \leq 30$, то делим на $n-1$

$$m_{\bar{x}} = \pm \frac{s}{\sqrt{n-1}} \quad 10$$

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

$$m_s = \pm \frac{s}{\sqrt{2n}} \quad 11$$

4). Определяем ошибку коэффициента вариации CV :

$$m_{CV} = \frac{CV}{\sqrt{2n}} \quad 12$$

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

Литература: [1], [2].

Практическое занятие

Задание:

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.

Задание 2

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].

Тема 7: Доверительный интервал. Определение точности опыта. Оптимальный объем выборки.

Лекция

Подчинение выборочных средних (как случайных величин) нормальному закону, большинство из них (95%) находится поблизости от генеральной средней — в диапазоне $M_{ген.} \pm 1,96m$ (приближенно $\pm 2m$). Интервальная оценка генерального параметра. Доверительный интервал для выбранного уровня значимости. Определение точности опыта. Относительная ошибка измерений — «показатель точности опыта». Объем условной выборки для обеспечения хорошей точности.

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 $M_{ген.} \pm 1,96m$ (приближенно $\pm 2m$).

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 .

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

Определение точности опыта. 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.

Оптимальный объем выборки. 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) (%).

Основные понятия темы:

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

Вопросы для самоконтроля

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].

Практическое занятие: доверительный интервал.

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	n	x_i	n	x_i	n	x_i	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		

Задание

Аналогичным образом рассчитайте доверительный интервал для:

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

Сделать вывод.

Литература: [1], [2].

Практическое занятие: определение точности опыта.

Относительная ошибка измерений — «показатель точности +опыта» — отношение ошибки средней к самой средней арифметической, выраженное в процентах:

$$\varepsilon = \frac{m}{M} \cdot 100\% \quad (14)$$

Чем точнее определена средняя, тем меньше будет ε , и наоборот. Точность считается хорошей, если ε меньше 3%, и удовлетворительной при $3 < \varepsilon < 5\%$. Иначе придется собирать дополнительный материал.

По данным таблиц 1.2, 2.1 и 7.1 определить существует ли необходимость дополнительного сбора материалов.

Литература: [1], [2].

Практическое занятие: оптимальный объем выборки для обеспечения хорошей точности.

Задание

По данным таблиц 1.2, 2.1 и 7.1 рассчитать объем условной выборки, необходимый для обеспечения хорошей точности $\varepsilon = 3\%$ для уровня значимости $\alpha = 0,05$ и для коэффициента вариации $CV \leq 10\%$ (значения коэффициента вариации, не выходящие за пределы 10%, принято считать нормальными).

Делаем вывод об оптимальности объема выборки.

Литература: [1], [2].

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

Лекция

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, $df = n_1 + n_2 - 2$). 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.

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 $df = n_1 + n_2 - 2$.

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.

Основные понятия

t -критерий Стьюдента, равенство дисперсий, критерий Фишера.

Вопросы для самоконтроля

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].

Практическое занятие: расчёт критерия Стьюдента.

Вопросы для обсуждения

1. Как можно отличить закономерное от случайного?
2. Как можно исключить чужеродные («выскакивающие») варианты?
3. Если полученная величина t-критерия меньше табличной, подтверждается нулевая гипотеза?

Задание

Отличить закономерное от случайного можно на основе законов поведения случайной величины. Для исключения чужеродных («выскакивающих») вариант применяется закон нормального распределения ($M \pm 1.96 \cdot S$), за пределами этого диапазона лежат чужеродные для данной выборки значения. Выборочные средние имеют нормальное распределение, критерий отличия двух выборочных средних также базируется на свойствах нормального распределения: в границах $M_{общ.} \pm 1.96 \cdot m$ (или приблизительно $M_{общ.} \pm 2 \cdot m$) выборочные средние арифметические отличаются от общей (генеральной) средней по случайным причинам. При соблюдении этих условий рабочая формула для t критерия отличия средних будет:

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

Разность средних брать по модулю, т. е. без учета знака. Значение критерия t Стьюдента сравнить с табличным при выбранном уровне значимости (обычно для $\alpha = 0,05$) и числе степеней свободы (объемы выборок без числа ограничений, $df = n_1 + n_2 - 2$).

Если полученное значение (величина) критерия больше табличного (таблица БП в [2]), значит, различия между параметрами при заданном уровне значимости и установленном числе степеней свободы достоверны, если нет — то различие недостоверно.

Задача 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.

Задача 2. 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. В первой группе средний уровень гемоглобина составил $115,4 \pm 1,2$ г/л, а во второй — $103,7 \pm 2,3$ г/л (данные представлены в формате $M \pm m$), сравниваемые совокупности имеют нормальное распределение. При этом численность первой группы составила 34, а второй — 40 пациентов. Необходимо сделать вывод о статистической значимости полученных различий и эффективности нового препарата железа.

Задача 3. Исследовалась радиоактивность образцов почвы на двух участках (табл. 8.1).

Таблица 8.1 — Радиоактивность образцов почвы

	мкР/час															n
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

Определите, существует ли достоверное отличие в уровне радиоактивности почвы на этих участках?

Литература: [1], [2].

Практическое занятие: расчёт критерия Фишера.

Вопросы для обсуждения

Наиболее точным методом для определения достоверности различий между выборочными дисперсиями является критерий F Фишера в форме отношения дисперсий (большее значение должно стоять в числителе):

$$F = \frac{s_1^2}{s_2^2} \sim F_{(\alpha, df_1, df_2)} \quad (16)$$

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 $\alpha = 0,01$) 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.

Пример. При сравнении по показателю плодовитости (число эмбрионов на самку) двух популяций красной полевки с разным уровнем численности (у первой, горной, популяции плотность населения в два раза выше, чем у равнинной) оказалось, что при очень близких средних арифметических (соответственно $M_1 = 5,8$ и $M_2 = 5,4$), разница статистически недостоверна, стандартные отклонения значительно различаются: $S_1 = 1,82$; $S_2 = 0,52$ (при $n_1 = 27$, $n_2 = 12$). Отсюда

$$F = \frac{s_1^2}{s_2^2} = \frac{3,3124}{0,2704} = 12,25 \quad (17)$$

Полученное значение критерия ($F = 12,2$) больше табличного $F(0,05; 26, 11) = 2,6$, следовательно, нулевую гипотезу о случайности отличий можно отбросить, сделав вывод о том, что показатели изменчивости плодовитости в разных по численности популяциях достоверно отличаются.

Задание

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].

Раздел 2. Виды анализа и функциональные зависимости.

Тема 9: Корреляционный анализ

Лекция

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 выразить 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 как численное выражение доли сопряженной вариации двух признаков в общей их вариации. Корреляционный анализ, коэффициент корреляции, ошибка коэффициента корреляции, коэффициент детерминации. Корреляционный анализ используется при *линейной* зависимости.

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

$$r = \frac{c_{xy}}{\sqrt{c_x \times c_y}} = \frac{\sum xy - (\sum x \times \sum y) / n}{\sqrt{(\sum x^2 - \frac{(\sum x)^2}{n}) \times (\sum y^2 - \frac{(\sum y)^2}{n})}}$$

Основные понятия темы:

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

Вопросы для самоконтроля:

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?

Литература: [1], [2].

Практическое занятие: расчёт коэффициента корреляции.

Вопросы для обсуждения

1. Каков диапазон изменения коэффициента корреляции?
2. Какая связь считается функциональной?
3. Назовите границы тесной (сильной), средней, слабой связи.
4. Почему необходим расчёт ошибки коэффициента корреляции?
5. Проверка значимости коэффициента критерием Стьюдента.
6. Что представляет собой поле корреляции?
7. По виду корреляционного поля можно судить о наличии или отсутствии связи и ее типе?

Задание:

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			
Σ					

а) По таблице рассчитать квадраты вариантов 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 .

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

в) 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.

д) Построить график зависимости (график рассеяния), показать линию тренда и привести коэффициент детерминации (R^2).

Литература: [1], [2].

Тема 10: Регрессионный анализ. Метод наименьших квадратов (МНК) (лекция рассчитана на два занятия).

Лекция

Регрессия как метод описания изменения функции в зависимости от одного или нескольких аргументов. 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.

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.

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.

Основные понятия

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

Вопросы для самоконтроля

1. Дайте определение регрессии.
2. В каких случаях рассчитывают линейную регрессию?
3. Какую возможность предоставляет МНК?
4. Какой вид регрессии используют при описании количественной связи между признаками, изменяющимися либо с разной скоростью, либо в разных масштабах?
5. К какому виду относятся графики, описывающие зависимость одного признака от другого?
6. Какими функциями описываются зависимости в нелинейной регрессии?
7. Дайте определение коэффициента детерминации.

Литература: [1], [2].

Практическое занятие: 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

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 23 ($y=a+bx$) рассчитать теоретические значения Y .

4). В программе *Excel*, используя функцию КОРРЕЛ, найти величину коэффициента корреляции r между x и y , а по формуле 20 — ошибку коэффициента r .

5). В программе *Excel*, используя функцию СТАНДОТКЛОН.В найти величину стандартного отклонения для x и y и по формуле 24 рассчитать ошибку коэффициента a .

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

6). Для проверки значимости коэффициента регрессии использовать критерий t Стьюдента по формуле 25:

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

Для уровня значимости $\alpha = 0,05$ и числа степеней свободы $df = n - 2$ найти табличное значение критерия Стьюдента t , 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 $M_{ген.} \pm 1,96m$ (приближенно $\pm 2m$).

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 $M \pm tm$ (формула 13)

По данным таблицы 10.2 найти границу доверительного интервала:

$$\text{верхняя: } \max Y = Y_i + t \times m_y$$

$$\text{нижняя: } \min Y = Y_i - t \times m_y.$$

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].

Тема 11. Однофакторный дисперсионный анализ количественных признаков (лекция рассчитана на два занятия).

Лекция

В практической деятельности возникает необходимость установить влияние факторов на результаты воздействия. Рассмотрим дисперсионный анализ, предложенный Фишером в 1920-х годах. Метод дисперсионного анализа заключается в расчленении

общей вариации (общей дисперсии) изучаемого признака на составные части — факториальную дисперсию, оценивающую межгрупповую изменчивость и дисперсию, обусловленную остальными неорганизованными факторами (случайную дисперсию).

Сущность метода дисперсионного анализа заключается в измерении отдельных дисперсий (общая, факториальная, остаточная), и дальнейшем определении силы (доли) влияния изучаемых факторов (оценки роли каждого из факторов, либо их совместного влияния) на результативный признак.

Дисперсионный анализ — это статистический метод оценки связи между факторными и результативным признаками в различных группах, отобранный случайным образом, основанный на определении различий (разнообразия) значений признаков. В основе дисперсионного анализа лежит анализ отклонений всех единиц исследуемой совокупности от среднего арифметического. В качестве меры отклонений берется дисперсия. Отклонения, вызываемые воздействием факторного признака (фактора) сравниваются с величиной отклонений, вызываемых случайными обстоятельствами. Если отклонения, вызываемые факторным признаком, более существенны, чем случайные отклонения, то считается, что фактор оказывает существенное влияние на результативный признак.

Влияние изучаемого фактора отражает факториальная, или межгрупповая, дисперсия $S^2_{факт.}$, а влияние случайных неорганизованных в данном исследовании причин — случайная $S^2_{случ.}$, или внутригрупповая, остаточная дисперсия $S^2_{остат.}$:

Общая вариация (сумма квадратов) признака рассчитывается как сумма квадратов отклонений всех вариант (x_i) от общей средней (M):

$$C_{общ.} = \sum (x_i - M)^2$$

Факториальная (межгрупповая, межвыборочная) сумма квадратов рассчитывается как сумма квадратов отклонений частных средних (M_j) для каждой выборки (всего k выборок) от общей средней:

$$C_{факт.} = \sum (M_j - M)^2.$$

Остаточная (случайная, внутригрупповая) сумма квадратов есть сумма квадратов отклонений вариант каждой выборки (x_i) от своей средней (M_j):

$$C_{случ.} = \sum (x_i - M_j)^2.$$

Сила влияния фактора определяется как доля частной суммы квадратов в общем варьировании признака.

Показатель силы влияния изучаемого факториального фактора составляет:

$$\eta^2_{факт.} = C_{факт.} / C_{общ.},$$

неорганизованных (случайных): $\eta^2_{случ.} = C_{случ.} / C_{общ.}$; сумма этих показателей, естественно, равна единице: $\eta^2_{факт.} + \eta^2_{случ.} = 1$.

Расчет дисперсий:

Факториальная дисперсия $S^2_{факт.} = \frac{C_{факт.}}{df_{факт.}}$

Случайная дисперсия $S^2_{случ.} = \frac{C_{случ.}}{df_{случ.}}$

Достоверность оценок влияния факторов оценивают по критерию Фишера (см. тему 8):

$$F = \frac{S^2_{факт.}}{S^2_{случ.}} \sim F(\alpha, df1, df2)$$

где $df1 = k - 1$, $df2 = n - k$, k — число градаций, n — общий объем всех выборок.

Далее проверяем нулевую гипотезу: «влияние фактора на признак отсутствует». Влияние считается доказанным, если величина расчетного критерия равна или превышает свое табличное значение с принятым уровнем значимости (обычно $\alpha = 0,05$) (F определяется по табл. 7П [2]).

Основные понятия: дисперсионный анализ, общая дисперсия, факториальная дисперсия, остаточная дисперсия.

Вопросы для самоконтроля

1. В чем заключается сущность дисперсионного анализа?
2. Дайте определение общей дисперсии.
3. Дайте определение факториальной дисперсии.
4. Дайте определение остаточной дисперсии.

Литература: [1], [2].

Практическое занятие: расчет общей дисперсии

Задание 1

В ходе исследования были получены четыре выборки, четыре группы значений плодовитости дафний, выращенных в средах с разным содержанием стимулирующего вещества.

A1 (контроль, чистая вода) 6, 5, 5, 7

A2 (слабая концентрация) 8, 7, 6, 6

A3 (средняя концентрация) 8, 8, 7

A4 (сильная концентрация) 8, 7, 9

Вычислить общую дисперсию, расчет общей дисперсии проводим по алгоритму, изложенному в теме 3.

Литература: [1], [2].

Практическое занятие: расчет факториальной и остаточной дисперсии и определение силы влияния фактора и критерия достоверности Фишера.

Задание 1: определить факториальную и остаточную дисперсии по данным, представленным на предыдущем практическом занятии, руководствуясь алгоритмом (табл. 11.1).

Таблица 11.1 — Этапы расчета факториальной и остаточной дисперсии и определение силы влияния фактора и критерия достоверности Фишера.

Составляющие дисперсии	Суммы квадратов, C	Сила влияния, η^2	Степени свободы, df	Дисперсии, S^2	Критерий F
Факториальная					
Случайная					
Общая					

Сравнить полученное значение критерия F с табличным и сделать вывод о достоверности влияния стимулирующего фактора на плодовитость дафний.

Литература: [1], [2].

Практическое занятие: расчет общей, факториальной и остаточной дисперсии и определение силы влияния фактора и критерия достоверности Фишера

Задание: определить силу влияния обработки пестицидом на плотность популяции личинок совки на отдельных участках.

Контроль: 18, 15, 21, 18, 19, 20, 22, 17, 16, 29, 14, 16, 27, 15, 16.

Опыт: 12, 11, 9, 5, 4, 13, 4, 7, 9, 14, 5, 8, 12, 10, 9.

Сделать вывод.

Вопрос: Эту задачу можно решить, используя только критерий Фишера (по соотношению дисперсий)?

Практическое занятие: контрольная работа по разделу 2.

1. Используя возможности Excel (статистические функции) рассчитать коэффициент корреляции между живым весом коров и весом их приплода.
2. При какой зависимости рассчитывают коэффициент корреляции?
3. Рассчитать коэффициенты 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

4. Охарактеризовать понятия общей, факториальной и остаточной дисперсии. Почему необходимо при проведении экспериментов это выявлять?
Литература: [1], [2], [3], [4].

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) материалов лекций;
- чтение и проработка рекомендованной основной и дополнительной литературы;
- подготовка к практическим 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 практическим занятиям, тематика которых полностью охватывает содержание курса. Самостоятельная работа по подготовке к семинарским занятиям предполагает умение работать с первичной информацией.

Самостоятельная работа по разделу 1:

Работа с конспектом лекций и рекомендованной литературой (1 и дополнительная).

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

Самостоятельная работа по разделу 2:

Работа с конспектом лекций и рекомендованной литературой (1 и дополнительная).

Подготовка материалов к контрольному опросу по изученным темам, практическим занятиям, тестовым проверкам знаний, 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 «Основы математической статистики в экологических исследованиях» 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 (экзамен)

1. Отличие генеральной совокупности от выборки.
2. Классификация признаков.
3. Понятие точность измерений.
4. Способы группировки первичных данных.
5. Характеристика вариационного ряда.
6. Определение количества и интервалов классов.
7. Недостаток средней арифметической.
8. Случай использования средней геометрической.
9. Преимущество медианы.
10. Характеристика моды.
11. Случай использования расчёта средневзвешенной величины.
12. Понятия дисперсии и среднего квадратичного отклонения.
13. Коэффициент вариации.
14. Характеристика нормального распределения.
15. Отличие биномиального распределения от нормального.
16. Свойства нормального распределения.
17. Определения асимметрии и эксцесса.
18. Расчет статистической ошибки.
19. Расчет ошибки стандартного отклонения.
20. Расчет ошибки коэффициента вариации.
21. Доверительный интервал.
22. Определение оптимального объема выборки.
23. Способы сравнения показателей изменчивости.
24. Нормированное отклонение;
25. Сравнение теоретических частот с фактическими и расчёт нормированного отклонения.
26. Оценка статистической достоверности разности;
27. Критерии t Стьюдента (сравнение средних арифметических коэффициентов вариации) и F Фишера (сравнение долей и показателей изменчивости по дисперсии);
28. Однофакторный дисперсионный анализ количественных признаков;
29. Корреляционный анализ;
30. Регрессионный анализ;
31. Дисперсионный анализ;
32. Построение графиков зависимости.
33. Оценка формы связи;
34. Линейная и нелинейная регрессия, выбор уравнений регрессии.
35. Построение теоретической кривой по уравнению регрессии.

7. Рекомендуемая литература

7.1 Основная

1. Пузаченко Ю.Г. Математические методы в экологических и географических исследованиях: учеб. пособие. — М.: Академия, 2004. — 416 с.
2. Ивантер Э.В., Коросов А.В. Элементарная биометрия : учеб. пособие / Э.В. Ивантер, А.В. Коросов. — Петрозаводск: Изд-во ПетрГУ, 2010. — 104 с.
3. Крючков А.В. Сборник задач по дисциплине «Биометрия» / А.В. Крючков. —

Киров: Изд-во ВятГУ, 2011. – 28 с.

7.2 Дополнительная

Васильева Л.А. Статистические методы в биологии, медицине и сельском хозяйстве. — Новосибирск. НГУ, 2007. —127 с.

Терентьев П.В., Ростова Н.С. Практикум по биометрии. Учебное пособие. — Л.: Изд-во ЛГУ, 1977. – 152 с.

Сиделев, С. И. Математические методы в биологии и экологии: введение в элементарную биометрию: учебное пособие / С. И. Сиделев; Яросл. гос. ун-т им. П. Г. Демидова. — Ярославль : ЯрГУ, 2012. — 140 с.

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/asimmetriya_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 чтение лекций, проведение семинарских занятий, групповых и индивидуальных консультаций по отдельным вопросам дисциплины. Предусмотрена самостоятельная работа студентов, а также прохождение аттестационных испытаний промежуточной аттестации.

Лекции посвящаются рассмотрению наиболее важных концептуальных вопросов: основным понятиям и методам математической статистики, применяемым в экологических исследованиях. В ходе лекций студентам следует подготовить конспекты лекций: кратко, схематично, последовательно фиксировать основные положения, выводы, формулировки, обобщения; пометать 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 практических занятий является закрепление знаний студентов, полученных ими в ходе изучения дисциплины на лекциях и самостоятельно. Практические занятия проводятся в форме семинаров; на них обсуждаются вопросы по теме, разбираются конкретные задания по изучаемой теме, решаются задачи. Для подготовки к занятиям семинарского типа студенты выполняют проработку рабочей программы, уделяя особое внимание целям и задачам, структуре и содержанию дисциплины; конспектирование источников; работу с конспектом лекций, 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 занятиям семинарского типа, 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. Лекция:

– лекция-визуализация — подача материала осуществляется средствами технических средств обучения с кратким комментированием демонстрируемых визуальных материалов (презентаций).

2. Практическое занятие:

– проводится с целью акцентирования внимания обучающихся на какой-либо актуальной теме или на наиболее важных и существенных ее аспектах. Тематический семинар углубляет знания студентов, ориентирует их на активный поиск путей и способов решения затрагиваемой проблемы.

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 лекционного типа, практических (семинарских) 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.

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 «Основы математической статистики в экологических исследованиях» для направления подготовки 05.04.06 «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.)