

**МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ
ЗАБАЙКАЛЬСКИЙ АГРАРНЫЙ ИНСТИТУТ – филиал ФГБОУ ВО «ИРКУТСКИЙ
ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ**

Факультет Технологический
Кафедра Агрономии

**Методические указания и контрольные вопросы по прохождению
дисциплины
«Биометрия с основами вариационной статистики в агрономии»**

Направление подготовки 35.03.04 «Агрономия»
Форма обучения: заочная
Квалификация (степень) Бакалавр
Курс 5 курс

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель освоения дисциплины: Статистические методы необходимы при постановке экспериментов, так как только с их помощью можно установить, зависит ли наблюдаемое различие между опытными и контрольными группами от влияния изучаемого фактора или же оно чисто случайное, т.е. определяется многими другими факторами, не контролируемыми и не поддающимися учету.

Понимание и учет статистических закономерностей помогают экспериментатору составить методически обоснованный план опытов, правильно их провести и, наконец, сделать объективные выводы.

Целью дисциплины «Биометрия с основами вариационной статистики в агрономии» является более эффективное использование студентами основных вопросов по статистическому учету получаемого экспериментального материала при дальнейшем учете данных в постановке опыта.

Основные задачи освоения дисциплины:

- изучить основные статистические показатели, переменные в статистике, познакомиться с описательной статистикой и статистической гипотезой;
- рассмотреть алгоритм расчета качественных и количественных признаков;
- изучить понятия регрессионного и корреляционного анализов.

Результатом освоения дисциплины «Биометрия с основами вариационной статистики в агрономии» является овладение бакалаврами по направлению подготовки 35.03.04 Агрономия следующих видов профессиональной деятельности:

- производственно-технологическая;
- организационно-управленческая;
- научно-исследовательская.

в том числе компетенциями заданными ФГОС ВО.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина «Биометрия с основами вариационной статистики в агрономии» находится в Вариативной части блока Б1 учебного плана. Приступая к изучению данной дисциплины, студент должен иметь базовые знания по дисциплинам: математика, ботаника, введение в специальность, основы научных исследований, растениеводство.

Знания и умения, полученные в результате освоения дисциплины, являются необходимыми для успешного прохождения преддипломной и производственной практик, а также для написания выпускной квалификационной работы.

Дисциплина изучается на 4 курсе в 7 семестре.

3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫХ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

В результате освоения дисциплины обучающийся должен овладеть знаниями, умениями и навыками в целях приобретения следующих компетенций:

Трудовое действие ¹	Наименование компетенции, необходимой для выполнения трудового действия (планируемые результаты освоения ОП)	Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенции
профессиональные компетенции		
Трудовая функция Организация производства продукции растениеводства		
Ведение опытной работы по применению новых технологий, новейших сортов сельскохозяйственных культур	ПК – 2 - способностью применять современные методы научных исследований в агрономии согласно утвержденным планам и методикам	В области знания и понимания (A)
		Знать: современные методы научных исследований в агрономии
		В области интеллектуальных навыков (B)
		Уметь: применять современные методы научных исследований в агрономии согласно утвержденным планам и методикам
		В области практических умений (C)
Владеть: способностью применять современные методы научных исследований в агрономии согласно утвержденным планам и методикам		
Обобщение и статистическая обработка результатов	ПК – 4 - способностью к обобщению и статистической обработке результатов опытов, формулированию выводов	В области знания и понимания (A)
		Знать: методы обобщения и статистической обработки результатов опытов, формулирования выводов
		В области интеллектуальных навыков (B)
		Уметь: применять методы обобщения и статистической обработки результатов опытов, формулирования выводов
		В области практических умений (C)
Владеть: способностью к обобщению и статистической обработке результатов опытов, формулированию выводов		

¹ Указывается в соответствии с профессиональным стандартом (при наличии) или квалификационными требованиями. Трудовые действия указываются, как правило, для профессиональных компетенций в соответствии с видом профессиональной деятельности. Для общекультурных и общепрофессиональных компетенций трудовые действия указываются в случае соответствия.

Ведение агрономической документации с использованием современных технологий	ПК – 5 - способностью использовать современные информационные технологии, в том числе базы данных и пакеты программ	В области знания и понимания (А)
		Знать: современные информационные технологии, в том числе базы данных и пакеты программ
		В области интеллектуальных навыков (В)
		Уметь: использовать современные информационные технологии, в том числе базы данных и пакеты программ
		В области практических умений (С)
		Владеть: способностью использовать современные информационные технологии, в том числе базы данных и пакеты программ

4. ОБЪЕМ ДИСЦИПЛИНЫ В ЗАЧЕТНЫХ ЕДИНИЦАХ С УКАЗАНИЕМ КОЛИЧЕСТВА ЧАСОВ, ВЫДЕЛЕННЫХ НА КОНТАКТНУЮ РАБОТУ ОБУЧАЮЩИХСЯ С ПРЕПОДАВАТЕЛЕМ (ПО ВИДАМ УЧЕБНЫХ ЗАНЯТИЙ) И НА САМОСТОЯТЕЛЬНУЮ РАБОТУ ОБУЧАЮЩИХСЯ

Общая трудоемкость дисциплины составляет 108 часов – 3 з.е.

4.1. Объем дисциплины и виды учебной работы:

4.1.1. Заочная форма обучения: курс – 5, семестр – 2, вид отчетности – диф. зачет.

Вид учебной работы	Объем часов / зачетных единиц	Объем часов / зачетных единиц	Объем часов / зачетных единиц
	всего	1 семестр	2 семестр
Общая трудоемкость дисциплины	108/3	108/3	
Контактная работа обучающихся с преподавателем (всего)	20	20	
в том числе:			
Лекции (Л)	8	8	
Семинарские занятия (СЗ)	12	12	
Лабораторные работы (ЛР)	-	-	
Самостоятельная работа:	84	84	
Курсовой проект (КП)	-	-	
Курсовая работа (КР)	-	-	
Расчетно-графическая работа (РГР)	-	-	
Реферат (Р)	-	-	
Эссе (Э)	-	-	
Контрольная работа/опрос	14	14	

Самостоятельное изучение разделов	10	10	
Самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий, подготовка к лабораторным и практическим занятиям, коллоквиумам, рубежному контролю и т.д.)	60	60	
Подготовка и сдача экзамена			
Подготовка и сдача зачета	4	4	

5. СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

5.1. Содержание дисциплины, структурированное по темам с указанием отведенного на них количества часов и видов учебных занятий:

5.1.1 Заочная форма обучения

№ п/п	Раздел дисциплины (тема)	Семестр	Неделя семестра	Виды учебных занятий, включая самостоятельную работу студентов и трудоемкость (в часах)				Формы текущего контроля успеваемости (по неделям семестра) Форма промежуточной аттестации (по семестрам)
				Лекции (Л)	Практ. (семинарские) занятия	Лабораторные работы (ЛР)	Самостоятельная работа (СРС)	
1	2	3	4	5	6	7	8	9
1	Введение в математическую статистику в биологии	2	1	2			6	опрос
2	Переменные в статистике	2	1		2		8	Опрос
3	Описательная статистика	2	2	2	2		8	Опрос
4	Статистическая гипотеза	2	2		2		8	Опрос
5	Доверительные интервалы	2	2		2		6	Опрос
6	Критерии значимости. Анализ количественных переменных	2	3	2			10	Опрос
7	Критерии значимости. Анализ качественных переменных	2	4	2			10	Опрос
8	Корреляционный анализ	2	4		2		14	Опрос
9	Регрессионный анализ	2	5		2		14	Реферат
10	итого	2	5	8	12		84	Диф. зачет

5.2. Тематическое содержание дисциплины

№ п/п	Раздел дисциплины	Тема и краткое содержание темы
I	II	III
1	Введение в математическую статистику в биологии	Биометрия как наука. Значение биометрии в исследовательской работе и профессиональной подготовке студентов. Роль работ У.Петти, Дж. Гранта, П.-С. Де Лапласа, П. Пуассона, П. Л.Чебышева, А. Кетле, К. Ф.Гаусса, Ф. Гальтона, К. Пирсона, У. Госсета, Р.Фишера и других ученых в развитии биометрии
2	Переменные в статистике	Понятие о наименьшей выборочной единице (единице наблюдения) и данных в биологии. Переменные (признаки). Генеральная совокупность и выборка. Количественные переменные: дискретные и непрерывные. Качественные переменные. Ранговая шкала измерений. Производные переменные: пропорции, индексы, интенсивности протекания процессов
3	Описательная статистика	Группировка данных в вариационный ряд. Способы графического изображения вариационного ряда: полигон (кривая) распределения, гистограмма. Теоретические распределения случайных величин и их свойства: биномиальное распределение, распределение Пуассона, нормальное распределение. Коэффициенты асимметрии и эксцесса. Средние величины: средняя арифметическая, взвешенная средняя, геометрическая средняя. Меры разброса единиц совокупности: дисперсия и стандартное отклонение. Коэффициент вариации. Мода. Медиана и процентиля. 25-й и 75-й процентиля (квартили). Расчет параметров описательной статистики при качественной изменчивости. Оценка репрезентативности выборочных показателей при помощи стандартной ошибки. Центральная предельная теорема. Закон больших чисел. Определение достаточного объема выборки. Доверительные интервалы для средней арифметической и для доли. Способы представления средних величин, мер разброса, стандартных ошибок и доверительных интервалов в научных публикациях
4	Статистическая гипотеза	Понятие о статистической гипотезе. Нулевая и альтернативная гипотезы. Статистические критерии (тесты). Вероятность справедливости нулевой гипотезы (уровень значимости). Статистические ошибки I и II типа. Мощность критерия (теста). Понятие о параметрических и непараметрических критериях (тестах). Способы трансформации данных для приведения их к нормальному распределению: логарифмирование, извлечение квадратного корня, преобразование Бокса-Кокса, угловое преобразование
5	Доверительные интервалы	Доверительные интервалы для средней и доли. Доверительные интервалы для разности средних и разности долей. Проверка значимости доверительных интервалов
6	Критерии значимости. Анализ	Назначение дисперсионного анализа (ANOVA). Нулевая гипотеза при дисперсионном анализе. Расчет внутри- и межгрупповой дисперсий при однофакторном анализе с равномерным дисперсионным

	количественных переменных	комплексом. F-критерий Фишера. Определение внутри- и межгруппового числа степеней свободы. Однофакторный дисперсионный анализ повторных измерений. Понятие о многофакторном дисперсионном анализе. Эффект множественных сравнений. Апостериорный (post-hoc) анализ и его методы: тесты Ньюмена-Кейлса, Даннета. Непараметрические аналоги однофакторного дисперсионного анализа: H-тест Крускала-Уоллиса и тест Фридмана. Сравнение двух групп. Тест Стьюдента как частный случай дисперсионного анализа. t-распределение. Тест Стьюдента для парных измерений. Использование доверительных интервалов для проверки гипотезы о равенстве двух средних. Введение поправки Бонферрони для t-критерия при проведении множественных сравнений средних. Непараметрические аналоги критерия Стьюдента: U-тест Манна-Уитни, тест Уилкоксона
7	Критерии значимости. Анализ качественных переменных	z-критерий для сравнения двух выборочных долей и условие его применимости. Анализ таблиц сопряженности при помощи χ^2 -критерия. Поправка Йетса на непрерывность. Использование критерия χ^2 для определения нормальности распределения данных. Определение числа степеней свободы при анализе таблиц сопряженности. Точный критерий Фишера. Одностороннее и двустороннее значения точного критерия Фишера.
8	Корреляционный анализ	Понятие о функциональной и корреляционной зависимостях. Степень и направление корреляционной зависимости. Коэффициент корреляции Пирсона и оценка его статистической значимости. Коэффициент ранговой корреляции Спирмена
9	Регрессионный анализ	Назначение регрессионного анализа. Общий вид регрессионного уравнения. Связь коэффициента регрессии с коэффициентом корреляции. Оценка параметров регрессионного уравнения по выборке с помощью метода наименьших квадратов. Статистическая значимость регрессии. Проверка нулевой гипотезы о равенстве коэффициента регрессии нулю. Стандартные ошибки параметров регрессионного уравнения. Коэффициент детерминации. Анализ остатков. Оценка величины остаточной дисперсии с помощью F-критерия. Нахождение доверительной области для линии регрессии. Понятие о нелинейной и множественной регрессионной зависимости

6. Организация и учебно-методическое обеспечение самостоятельной работы студентов

Литература:

Основная

1. Крюков, А.М. Биометрия : учеб.-метод. пособие / Т.В. Шишкина, А.М. Крюков .— Пенза : РИО ПГСХА, 2014
<http://rucont.ru/efd/280054>
2. Моисейкина, Л.Г. Практикум по генетике и биометрии / Б.М. Турдуматов, П.М. Кленовицкий, Л.Г. Моисейкина .— Элиста : Калмыцкий государственный университет, 2012
<http://rucont.ru/efd/297586>
3. Моисейкина, Л.Г. Пособие по биометрии и генетике / Б.М. Турдуматов, П.М. Кленовицкий, Л.Г. Моисейкина .— Элиста : Калмыцкий государственный университет, 2011
<http://rucont.ru/efd/297585>

Дополнительная литература:

1. Смиряев Л.В., Гохман М.В. "Биометрические методы в селекции растений", М.,Агропромиздат, 1985. 215с.
2. Снедкор Дж. "Статистические методы в применении к исследованиям в сельском хозяйстве и биологии". Перевод с английского, М., "Сельхозиздат", 1961.
3. Демьянов Ю. Э. Литвин Н.Ф. Применение математических методов и ЭВМ в биологии. Под ред. Селькова. М. Изд-во МГУ. 1981.
4. Боровиков В. STATISTICA: искусство анализа данных на компьютере. Для профессионалов. – СПб.: Питер, 2001. – 656 с.

Рекомендации по выполнению самостоятельной работы

Для оценки самоподготовки студенты выполняют самостоятельную работу в виде реферата. В пособии даны задания для реферата, включающие основные вопросы курса биометрии. При написании реферата необходимо указать суть данного вопроса, ответ можно сопровождать рисунками, схемами и т.п., обязательно необходимо приводить примеры для иллюстрации объясняемого явления. Структура реферата включает в себя титульный лист с указанием дисциплины, темы, номера шифра и специальности, ФИО студента и преподавателя, кроме того, в структуру входит оглавление, введение, основная часть реферата, заключение, список литературы.

Общий объем реферата составляет 12 листов формата А4. Оформляется реферат согласно Методическому пособию по оформлению рефератов, курсовых и выпускных квалификационных работ (420 ауд.).

Вопросы для реферата.

1. Понятие о статистических и динамических совокупностях.
2. Выборочный метод исследования, достоинства и недостатки.
3. Статистические комплексы (понятие).
4. Первичная группировка данных.
5. Понятие о вариационном ряде, варианте и её удельном весе в вариационном ряду.
6. Способ графического изображения вариационных рядов: кривая распределения признака, полигон.
7. Способ графического изображения интервальных вариационных рядов: гистограмма, кумулята и огива распределения.
8. Показатели положения (центральной тенденции) параметрические: простая и взвешенная средние арифметические, средняя квадратическая, средняя кубическая, средняя геометрическая, средняя гармоническая; и непараметрические: Мода, медиана.
9. Показатели изменчивости: лимиты; размах; среднее квадратическое отклонение от средней арифметической для количественных и качественных признаков; коэффициент вариации.
10. Свойства среднего квадратического отклонения (σ).
11. Понятие о квартильном отклонении (Q), области использования.
12. Ошибки выборочных показателей. Коэффициент вариации, его отличие от среднего квадратического отклонения.
13. Закономерности случайной вариации. Вероятность. Формулы для вычисления вероятности.
14. Нормальная вариационная кривая и ее характеристика. Нормированное отклонение.
15. Уровни значимости. Связь между уровнем значимости и вероятностью.
16. Доверительные вероятности или доверительный интервал.
17. Оценка достоверности статистических показателей. Выборочные и генеральные совокупности.
18. Средние ошибки, ошибки выборочности. Формулы вычисления.
19. Критерий Стьюдента, случаи и примеры его использования.
20. Нулевая гипотеза. Сущность нулевой гипотезы.
21. Формулы для определения необходимого объема выборочной совокупности. Охарактеризуйте основные предпосылки выборочного метода.
22. Измерение связи. Корреляция. Понятие о корреляции. Положительная и отрицательная корреляция.
23. Коэффициент корреляции. Формулы для его вычисления.
24. Выборочность коэффициента корреляции. Оценка его достоверности.
25. Понятие о регрессии. Односторонняя и двусторонняя регрессия.
26. Коэффициент регрессии. Ошибка коэффициента регрессии и его достоверность.

Вопросы для зачета.

1. Предмет и основные понятия биологической статистики. История биометрии.

2. Группировка данных, совокупность и вариационный ряд.
3. Совокупность, примеры различных совокупностей. Отличие выборочной совокупности от генеральной совокупности.
4. Принципы группировки данных при качественной дискретной и непрерывной изменчивости.
5. Вариационный ряд. Особенности распределения вариантов в вариационном ряду. Графическое изображение вариационного ряда.
6. Статистические показатели для характеристики совокупности.
7. Размах вариационного ряда и лимиты. Мода и медиана.
8. Средняя арифметическая и ее свойства. Формулы для вычисления.
9. Варианса и среднее квадратическое отклонение.
10. Понятие степень свободы.
11. Средняя геометрическая. Формулы для ее вычисления.
12. Коэффициент вариации, его отличие от среднего квадратического отклонения.
13. Закономерности случайной вариации. Вероятность. Формулы для вычисления вероятности.
14. Нормальная вариационная кривая и ее характеристика. Нормированное отклонение.
15. Уровни значимости. Связь между уровнем значимости и вероятностью.
16. Доверительные вероятности или доверительный интервал.
17. Оценка достоверности статистических показателей. Выборочные и генеральные совокупности.
18. Средние ошибки, ошибки выборочности. Формулы вычисления.
19. Критерий Стьюдента, случаи и примеры его использования.
20. Нулевая гипотеза. Сущность нулевой гипотезы.
21. Формулы для определения необходимого объема выборочной совокупности. Охарактеризуйте основные предпосылки выборочного метода.
22. Измерение связи. Корреляция. Понятие о корреляции. Положительная и отрицательная корреляция.
23. Коэффициент корреляции. Формулы для его вычисления.
24. Выборочность коэффициента корреляции. Оценка его достоверности.
25. Понятие о регрессии. Односторонняя и двусторонняя регрессия.
26. Коэффициент регрессии. Ошибка коэффициента регрессии и его достоверность.
27. Статистический анализ вариации по качественным признакам.
28. Альтернативная вариация. Средняя арифметическая и среднее квадратическое отклонение при альтернативной вариации.
29. Средняя ошибка при альтернативной вариации. Доверительные границы для доли.
30. Дисперсионный анализ. Сущность дисперсионного анализа.
31. Общая схема дисперсионного анализа при однофакторном опыте.
32. Установление достоверности влияния изучаемого фактора. Фактические и табличные значения F.

33. Изучение степени соответствия фактических данных теоретически ожидаемым.
34. Критерий соответствия хи-квадрат. Формулы для его вычисления.
35. Закономерности распределения χ^2 . Понятие вероятности и значимости в применении χ^2 .
36. Фактические данные и нулевая гипотеза. Области отбрасывания нулевой гипотезы.

Глоссарий по дисциплине.

Алгоритм - полностью определенный, конечный набор шагов, операций или процедур, которые приводят к конкретному результату.

Альтернативная вариация – простейший случай качественной вариации, когда совокупность состоит только из двух групп: одной, имеющей данный признак, а другой – его не имеющей.

Анализ выживаемости - (разведочный анализ данных и проверка гипотез) включает описательные методы для оценивания распределения выборочных времен жизни, сравнения выживаемости в двух или нескольких группах, а также опции подгонки линейных и нелинейных регрессионных моделей к данным о выживаемости. Характерным аспектом данных о выживаемости является наличие так называемых цензурированных наблюдений, например, наблюдаемых объектов, которые дожили до определенного момента времени, а после этого были исключены из наблюдения. Вместо удаления такого наблюдения из множества изучаемых данных (т.е. необязательной потери потенциально важной информации), методы анализа выживаемости позволяют собрать цензурированные наблюдения и использовать их при проверке статистической значимости и подгонке модели.

Анализ соответствий - это раздел статистики, разрабатывающий описательные/разведочные методы анализа двухвходовых и многовходовых таблиц, которые обуславливают некоторую степень соответствия между строками и столбцами. Результаты этих методов похожи по своей природе на методы факторного анализа и позволяют исследовать структуру группирующих переменных, включенных в таблицу.

Аппарат Гальтона – устройство, предназначенное для наглядной демонстрации распределения вариант в виде вариационного ряда, частоты в котором следуют коэффициентам разложения бинома Ньютона.

Апостериорные сравнения - Обычно, получив при проведении дисперсионного анализа статистически значимое значение F-критерия, мы хотели бы узнать, какая из групп вызвала этот эффект, т.е. какие из групп значительно отличаются от других. Конечно, мы могли бы вычислить последовательность обычных t-критериев для сравнения всех возможных пар средних. Однако такая процедура будет основана на случайности. Получаемые уровни вероятности будут завышать значимость различия между средними. Например, предположим, что мы получили 20 выборок по 10 случайно выбранных чисел каждая, а затем вычислили 20 средних. После этого возьмем группу (выборку) с наибольшим средним и сравним ее с выборкой с наименьшим средним. t-критерий для независимых выборок проверяет, являются ли два средних значимо отличающимися друг от друга, в предположении, что рассматриваются всего две выборки. Метод апостериорных сравнений, наоборот, предполагает наличие более чем двух выборок. Этот метод используется для проверки гипотез и разведочного анализа.

Априорные вероятности - задают пропорции классов в популяции (в задачах классификации), особенно в тех случаях, когда известно, что эти пропорции отличаются от пропорций в обучающем множестве. Используются для модификации обучения.

Асимметрия или **коэффициент асимметрии** - (термин был впервые введен Пирсоном, 1895) является мерой несимметричности распределения. Если этот коэффициент отчетливо отличается от 0, распределение является асимметричным. Плотность нормального распределения симметрична относительно среднего.

Байесовы сети - сети, чей принцип действия основан на теореме Байеса, позволяющей сделать выводы о распределении вероятностей на основании имеющихся данных.

Бимодальное распределение - распределение, имеющее две моды (т.е. два "пика"). Бимодальность распределения выборки часто является показателем того, что распределение не является нормальным. Б.Р. дает важную информацию о природе исследуемой переменной. Например, если переменная представляет собой предпочтение

или отношение к чему-то, то бимодальность может означать противоположность мнений. Тем не менее, бимодальность часто может показывать, что выборка не является однородной и наблюдения порождены двумя или более "наложенными" распределениями. Иногда бимодальность распределения означает, что выбранные инструменты не подходят для измерения.

Биномиальное распределение – распределение, при котором вероятности появления отдельных значений x_i выражаются величинами, соответствующие коэффициентам разложения бинома Ньютона.

Варианта – значение или мера признака для единицы совокупности.

Варианса (средний квадрат отклонений вариант от средней арифметической) σ^2 – это сумма квадратов отклонений отдельных значений данной переменной от средней арифметической, деленная на число вариант.

Вариация (дисперсия) - различие между единицами совокупности.

Вариационный ряд – ряд, в котором показано, как часто встречаются варианты каждого класса и как варьируют признаки от минимальной величины до максимальной.

Вероятность – возможность осуществления определенного события в некотором количестве случаев из общего числа возможных, или, иначе говоря, степень уверенности в том, что событие произойдет.

Вероятностный или стохастический процесс – процесс осуществления явления на основе известной его возможности или вероятности.

Вероятностные нейронные сети - вид нейронных сетей для задач классификации, где плотность вероятности принадлежности классам оценивается посредством ядерной аппроксимации.

Взаимодействия - эффект *взаимодействия* возникает, когда зависимость между двумя или более переменными изменяется под воздействием одной или нескольких других переменных. Другими словами, сила или знак (направление взаимодействия) зависимости между двумя или более переменными зависит от значения принимаемого некоторыми другими переменными. Термин *взаимодействие* был впервые использован в работе Фишера (Fisher, 1926). Отметим, что слово "зависит" в данном контексте не означает причинной зависимости, а просто отражает тот факт, что в зависимости от рассматриваемого подмножества наблюдений (от значения модифицирующей переменной или переменных) характер зависимости будет меняться (модифицироваться).

Внутриклассовый коэффициент корреляции - значение внутриклассового коэффициента корреляции для популяции является мерой однородности наблюдений внутри классов случайного фактора относительно изменчивости наблюдений между классами. Он равен нулю только в случае, когда оцениваемый эффект случайного фактора равен нулю, и достигает единицы только если оцениваемый эффект ошибки равен нулю, при условии, что общая дисперсия наблюдений отлична от нуля. *Внутриклассовый коэффициент корреляции* может быть измерен с помощью метода оценивания компонент дисперсии.

Временной ряд - это последовательность измерений в последовательные моменты времени. Анализ *временных рядов* включает широкий спектр разведочных процедур и исследовательских методов, которые ставят две основные цели: (а) определение природы временного ряда и (б) прогнозирование (предсказание будущих значений временного ряда по настоящим и прошлым значениям). Обе эти цели требуют, чтобы модель ряда была идентифицирована и, более или менее, формально описана. Как только модель определена, вы можете с ее помощью интерпретировать рассматриваемые данные (например, использовать в вашей теории для понимания сезонного изменения цен на товары, если занимаетесь экономикой). Не обращая внимания на глубину понимания и справедливость теории, вы можете экстраполировать затем ряд на основе найденной модели, т.е. предсказать его будущие значения.

Выбросы - это нетипичные или редкие значения, которые существенно отклоняются от распределения остальных выборочных данных. Эти данные могут отражать истинные свойства изучаемого явления (переменной), а могут быть связаны с ошибками измерения или аномальными явлениями, и поэтому не должны включаться в модель.

Выборочная совокупность – сравнительно небольшая по объему совокупность, входящая в состав генеральной.

Генеральная совокупность – теоретически бесконечно большая или приближающаяся к бесконечности совокупность.

Групповое программное обеспечение - это программное обеспечение, которое дает возможность группе пользователей, использующих компьютерную сеть, одновременно работать над конкретным проектом. Оно содержит средства для организации связи (электронную почту), для совместной обработки документов, проведения анализа, создания отчетов и статистической обработки данных, а также календарного планирования и наблюдения. При этом обрабатываемые документы могут содержать информацию любого типа: текст, картинки или мультимедийный формат.

Дискриминантный анализ - используется для принятия решения о том, какие переменные дискриминируют или разделяют объекты на две или более естественно возникающих групп (его используют как метод проверки гипотез или как метод разведочного анализа).

Дисперсионный анализ – позволяет оценивать значимость влияния отдельных факторов, а также их относительную роль в общей изменчивости. **Д. а.** был разработан английским математиком и биологом Р. Фишером.

Доверительные вероятности – вероятность, при достижении которой можно с большой степенью уверенности заключить определенный вывод. В биологии используются доверительные вероятности: 0,95 и 0,99. Понятие **Д.В.** было введено Р. Фишером.

Доверительные границы или **доверительный интервал** - используются для оценки той или иной величины, указывают те границы, в которых она может находиться при разных вероятностях.

Доля выборки – отношение n/N , где n – численность выборочной совокупности, а N – численность генеральной совокупности. Используется для получения более точного значения средней ошибки.

Желаемая точность – допустимое расхождение между средней арифметической (по данному признаку) выборки и средней арифметической генеральной совокупности.

Закон больших чисел – выражает связь между статистическими показателями выборочных и генеральных совокупностей, заключается в том, что чем больше число n некоторых случайных величин, тем их средняя арифметическая ближе к средней арифметической генеральной совокупности.

Интервальная шкала - эта шкала измерений позволяет не только упорядочить наблюдения, но и количественно выразить расстояния между ними (при этом на шкале не обязательно присутствует *абсолютная* нулевая отметка).

Интерполяция - восстановление значения функции в промежуточной точке по известным ее значениям в соседних точках.

Канонический анализ - каноническая корреляция позволяет исследовать зависимость между двумя наборами переменных (и применяется для проверки гипотез или как метод разведочного анализа).

Категоризация, группировка, разбиение на подмножества - одним из наиболее важных, общих, а также мощных аналитических методов заключается в разделении (разбиении) данных на несколько подмножеств и последующее сравнение структуры данных в полученных подмножествах. У этого общего метода имеется много различных названий (в том числе: разбиение, группировка, категоризация, расщепление,

разветвление и условный анализ), и он используется как для разведочного анализа данных, так и для проверки гипотез.

Качественная изменчивость – изменчивость, различия между вариантами которой выражаются в каких-либо качествах.

Классификация - отнесение наблюдения к одному из нескольких, заранее известных классов (представленных значениями номинальной выходной переменной).

Кластерный анализ - термин *кластерный анализ* (впервые ввел Тьюон, 1939) в действительности включает в себя набор различных алгоритмов классификации. Общий вопрос, задаваемый исследователями во многих областях, состоит в том, как организовать наблюдаемые данные в наглядные структуры, т.е. развернуть таксономии и определить кластеры схожих объектов. Например, биологи ставят цель разбить животных на различные виды, чтобы содержательно описать различия между ними. В соответствии с современной системой, принятой в биологии, человек принадлежит к приматам, млекопитающим, амниотам, позвоночным и животным. Заметьте, что в этой классификации, чем выше уровень агрегации, тем меньше сходства между членами в соответствующем классе. Человек имеет больше сходства с другими приматами (т.е. с обезьянами), чем с "отдаленными" членами семейства млекопитающих (например, собаками) и т.д.

Ковариация - показатель, являющийся связующим звеном между корреляционным и регрессионным методами анализа.

Коды (значения группирующих переменных) - представляют собой значения группирующей переменной (например, 1, 2, 3, ... или *Мужской, Женский*), которые определяют уровни группирующей переменной для анализа. Коды могут быть как целочисленными, так и текстовыми значениями.

Количественная дискретная (прерывная) изменчивость – изменчивость, при которой различия между вариантами отдельными значениями случайной переменной, выражаются целыми числами, между которыми нет и не может быть переходов.

Количественная непрерывная изменчивость – вариация, при которой значения вариант выражаются как целыми, так и дробными числами.

Комплексные числа - это множество чисел, которое включает все действительные и мнимые числа. Комплексное число представляется выражением вида $a + ib$, где a и b - действительные числа, i - мнимая единица,

Компоненты дисперсии (в смешанной модели дисперсионного анализа). Термин *компоненты дисперсии* используется в контексте дисперсионного анализа и планирования эксперимента, включающего случайные эффекты, для обозначения оценки (доли) дисперсии, которая связана с этими эффектами.

Корреляция - это мера связи между двумя переменными. Коэффициент корреляции может изменяться от -1.00 до +1.00. Значение -1.00 означает полностью отрицательную корреляцию, значение +1.00 означает полностью положительную корреляцию. Значение 0.00 означает отсутствие корреляции.

Корреляция Пирсона - наиболее часто используемый коэффициент корреляции Пирсона r (Pearson, 1896) называется также *линейной корреляцией* (термин корреляция впервые ввел Galton, 1888), т.к. измеряет степень линейных связей между переменными. Можно сказать, что корреляция определяет степень, с которой значения двух переменных пропорциональны друг другу. Важно, что значение коэффициента корреляции не зависит от масштаба измерения. Например, корреляция между ростом и весом будет одной и той же, независимо от того, проводились измерения в дюймах и фунтах или в сантиметрах и килограммах. Пропорциональность означает просто линейную зависимость. Корреляция высокая, если на графике зависимость можно представить прямой линией (с положительным или отрицательным углом наклона). Проведенная прямая называется прямой регрессии или прямой, построенной методом наименьших квадратов. Последний термин связан с тем, что сумма квадратов расстояний (вычисленная по оси Y) от

наблюдаемых точек до прямой является минимальной из всех возможных. Заметим, что использование квадратов расстояний приводит к тому, что на оценки параметров сильно влияют выбросы. Корреляция Пирсона предполагает, что две рассматриваемые переменные измерены, по крайней мере, в интервальной шкале.

Корреляционные или статистические связи – связи, при которых численному значению одной переменной соответствует много значений другой переменной.

Корреляционные плеяды – сложная сеть корреляционных связей между многими признаками.

Коэффициент вариации – применяется при сравнении вариации различных признаков, представляет собой отношение σ к x , выраженное в процентах.

Коэффициент детерминации - это квадрат корреляции Пирсона между двумя переменными. Он выражает количество дисперсии, общей между двумя переменными.

Коэффициент корреляции r – указывает на степень связи в вариации двух переменных величин или на меру тесноты этой связи.

Коэффициент регрессии - количественная мера регрессии, вычисляемая если известны сигмы обоих вариационных рядов по признакам x и y , и коэффициенты корреляции между ними.

Кривая распределения (вариационная кривая) – графическое изображение вариационного ряда.

Критерий соответствия хи-квадрат χ^2 – показатель, определяющий степень соответствия фактических данных теоретически ожидаемым, или согласие фактических данных с предложенной гипотезой.

Критерий Стьюдента t – применяется при малых выборках ($n \leq 30$), характеризует отклонение выборочных средних от генеральной средней. Устанавливает тот факт, что среднее квадратическое отклонение для малых выборок постоянно отличается от того, которое ожидалось бы при нормальном распределении.

Круговая диаграмма - последовательность значений переменной изображается в виде последовательных круговых секторов (термин "круговая диаграмма" был впервые использован Хаскеллом в 1922 г.); размер каждого сектора пропорционален соответствующему значению. Значения должны быть больше 0 (нулевое и отрицательные значения не могут быть представлены в виде круговых секторов). Круговая диаграмма интерпретирует данные самым непосредственным образом: одно наблюдение соответствует одному сектору.

Лимиты (пределы) – значения крайних классов, верхняя и нижняя граница вариационного ряда.

Метод регрессии – метод, позволяющий установить, как количественно меняется одна величина при изменении другой на единицу.

Медиана – значение варианты, находящееся точно в середине ряда.

Множественная корреляция – зависимость изменения величины x от одновременного изменения величин y, z и т.д.

Мода – значение модального класса, являющееся как бы типичной для всей совокупности.

Модальный класс – класс, обладающий наибольшей частотой.

Номинальные переменные - переменные, которые могут принимать конечное множество значений, например, $Пол = \{Муж, Жен\}$.

Нормальная вариационная кривая – симметричная плавная кривая, при которой верхние границы ломанной линии полигона сливаются в гладкую кривую линию.

Нормированное отклонение t – представляет собой отклонение тех или других вариантов от их средней арифметической, выраженное в долях среднего квадратического отклонения.

Нулевая гипотеза - согласно этой гипотезе, первоначально принимается, что между данными показателями (или группами, на основе которых они получены) достоверного различия нет, т.е. что обе группы вместе составляют один и тот же однородный материал, одну совокупность.

Общность - это доля дисперсии, которая является общей для данной и всех остальных переменных. Доля дисперсии, которая является характерной для данной переменной (иногда называется характерностью) получается после вычитанием общности из дисперсии переменной. Другими словами дисперсия переменной есть общность плюс характерность. Обычно вначале в качестве оценки общности используют коэффициент множественной корреляции выбранной переменной со всеми другими.

Объем совокупности – число единиц совокупности.

Отрицательная корреляция - обратная зависимость между признаками: увеличение одного признака соответственно связано с уменьшением другого.

Ошибка выборочности или **ошибка репрезентативности** - представляют собой среднюю величину расхождения между средними значениями изучаемых признаков в выборках и генеральной совокупности.

Ошибка выборочности коэффициента корреляции – мера расхождения между коэффициентами корреляции для выборочной и генеральной совокупности.

Полигон распределения – графическое изображение конкретных вариационных рядов, применяющееся при дискретной вариации.

Положительная корреляция – прямая зависимость между признаками: при увеличении одного увеличивается и другой.

Поправка на непрерывность Йейтса – применяется при вычислении χ^2 в случае если исследуются малочисленные группы.

Ранжировка – расположение всех вариантов по порядку от минимальных до максимальных значений.

Распределение Пуассона или **пуассоновое распределение** – в биологии применяется для анализа редко наблюдаемые явления.

Симметричное распределение - если вы разобьете распределение пополам в точке среднего (или медианы), то распределения значений с двух сторон от этой центральной точки будут "зеркальным отображением" друг друга.

Случайная переменная – величина, изменяющаяся под влиянием многих случайных причин, которая может принимать разные значения.

Совокупность - всякое множество отдельных отличающихся друг от друга и в то же время сходных в некоторых существенных отношениях объектов.

Среднее - показывает "центральное положение" (центр) переменной и рассматривается совместно с доверительным интервалом. Обычно интерес представляют показатели (например, среднее), дающие информацию о популяции в целом. Чем больше размер выборки, тем более надежна оценка среднего. Чем больше изменчивость данных (больше разброс), тем оценка менее надежна.

Средняя арифметическая \bar{x} – некоторая урavnенная величина, отражающая основные свойства всех членов совокупности.

Средняя геометрическая – статистический показатель, применяемый в случае, если возрастание данного признака происходит умножением пропорционально степени.

Стандартная ошибка - термин стандартная ошибка среднего был впервые введен Юлом (Yule, 1897). Эта величина характеризует стандартное отклонение выборочного среднего, рассчитанное по выборке размера n из генеральной совокупности, и зависит от дисперсии генеральной совокупности (сигма) и объема выборки (n).

Стандартное отклонение - (термин был впервые введен Пирсоном, 1894), это широко используемая мера разброса или вариабельности (изменчивости) данных.

Таблицы сопряженности – таблицы, в которых предусматривается распределение групп по признакам, сопряженность или связь между которыми нужно будет установить.

Теоретические (априорные) вероятности – вероятности, которые знают заранее до проведения опыта.

Уровень значимости – обозначает вероятность получения случайного отклонения от установленных с определенной вероятностью результатов. Вероятности 0,95 (95%) соответствует уровень значимости 0,05% (5%). При вероятности 0,99% (99%) уровень значимости 0,01 (1%).

Функциональная зависимость – зависимость, при которой, каждому значению одной переменной величины соответствует одно вполне определенное значение другой переменной.

Частная корреляция - корреляция между двумя переменными, вычисленная после устранения влияния всех других переменных, называется частной корреляцией

Число степеней свободы df – величина $n-1$.

Экстраполяция - прогнозирование неизвестных значений путем продолжения функций за границы области известных значений.

Эмпирические (апостериорные) вероятности – вероятности, которые получены после проведения опыта.

7. Средства (ФОС) текущей и итоговой оценки качества освоения дисциплины

Оценка результатов самостоятельной работы организуется как единство двух форм: самоконтроль и контроль со стороны преподавателей.

Оценка успеваемости студентов осуществляется по результатам:

- 1) Подготовки к лекциям (написания конспектов).
- 2) Устного опроса на лекциях и практических занятиях.
- 3) Выполнения и защиты практических работ, индивидуальных контрольных работ.
- 4) Сдаче зачета.

Материально-техническое обеспечение модуля (дисциплины):

Компьютерная проекционная техника. Демонстрационные плакаты. Раздаточный методический материал. Макеты. Программное обеспечение: Statistica 6.0, Past.

Методические рекомендации составлены на основе Государственного образовательного стандарта и программе учебной дисциплины по специальности 35.03.04 «Агрономия»

Автор (ы) к.б.н., доцент Борискин И.А.

Программа одобрена на заседании кафедры Агрономия ЗабАИ-филиала ФГБОУ ВПО «ИрГСХА» (протокол № 7 от «25» мая 2016 г.).

Заведующий кафедрой _____ к.б.н., доцент Борискин И.А.

Программа одобрена на заседании учебно-методической комиссии Технологического факультета протокол № 6 от «26» 06 2016

Председатель учебно-методической комиссии _____