

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Таскаев Сергей Валерьевич
Должность: Ректор
Дата подписания: 12.09.2025 09:55:52
Уникальный программный ключ:
04c19ed8bfb98f3b6cb77a486b9a8788b8322323

	МИНОБРАЗОВАНИЯ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)
	Фонд оценочных средств по дисциплине «Статистика в генетике» по направлению подготовки 06.04.01 «Биология» ФГБОУ ВО «ЧелГУ»

**Фонд оценочных средств
для промежуточной аттестации
по дисциплине (модулю)**

Статистика в генетике

Направление подготовки (специальность)
06.04.01 Биология

Направленность (профиль)
Генетика

Присваиваемая квалификация
Магистр

Форма обучения
очная

Год (ы) набора: 2025

Челябинск, 2025 г.

1. ПАСПОРТ ФОНДА ОЦЕНОЧНЫХ СРЕДСТВ

Направление подготовки: **06.04.01 Биология**

Направленность (профили): Генетика

Дисциплина: **Статистика в генетике**

Семестры изучения: 1

Форма промежуточной аттестации: зачет

2. ПЕРЕЧЕНЬ ФОРМИРУЕМЫХ КОМПЕТЕНЦИЙ

2.1. Компетенции, закреплённые за дисциплиной

Изучение дисциплины «Статистика в генетике» направлено на формирование следующих компетенций:

Коды компетенции (по ФГОС)	Содержание компетенций согласно ФГОС	Коды и содержание индикаторов	Перечень планируемых результатов обучения по дисциплине
1	2	3	4
УК-1	Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	<p>УК-1.1 Критически анализирует проблемную ситуацию с целью выработки стратегии действий, аргументировано формулирует собственные суждения и оценки</p> <p>УК-1.2 Использует критический анализ, систематизацию и обобщение информации для решения проблемной ситуации</p>	<p>Знать: Для достижения индикатора УК-1.1: основные законы теории вероятностей, комбинаторики и математической статистики. Для достижения индикатора УК-1.2: возможности современных программных средств</p> <p>Уметь: Для достижения индикатора УК-1.1: проводить анализ принципов структурной и функциональной организации биологических объектов и механизмов гомеостатической регуляции. Для достижения индикатора УК-1.2: работать в качестве пользователя в статистических пакетах; самостоятельно формулировать задачи математической статистики; правильно применять различные статистические критерии.</p> <p>Владеть: Для достижения индикатора УК-1.1: основными физиологическими методами анализа и оценки состояния живых систем; приемами</p>

			планирования эксперимента. Для достижения индикатора УК-1.2: параметрическими и непараметрическими методами оценки, сравнения и характеристики данных медико-биологических исследований.
ПК-2	Способен использовать в научной и производственно-технологической деятельности знания фундаментальных и прикладных разделов генетических дисциплин	ПК-2.1 Имеет представление об основных методах и молекулярной биологии ПК-2.3 Анализирует основные методы исследования, применяемые в современной генетике	Знать: Для достижения индикатора ПК-2.1: принципы структурной и функциональной организации биологических объектов Уметь: Для достижения индикатора ПК-2.3: выбирать адекватные поставленным задачам специализированные статистические пакеты для анализа генетических данных. Владеть: Для достижения индикатора ПК-2.3: методами количественной оценки наследуемости и повторяемости признаков человека, растений и животных.

3. СОДЕРЖАНИЕ ОЦЕНОЧНЫХ СРЕДСТВ ПО ДИСЦИПЛИНЕ

3.1 Виды оценочных средств

№ п/п	Код компетенции/планируемые результаты обучения	Контролируемые темы/разделы	Наименование оценочного средства для текущего контроля	Наименование оценочного средства на промежуточной аттестации № задания
	УК-1 Знать: Для достижения индикатора УК-1.1: основные законы теории вероятностей, комбинаторики и математической статистики. Для достижения индикатора УК-1.2: возможности современных	1. Понятие о качественных и количественных признаках. Основные генетико-статистические величины и их применение. 5. Показатели связи между признаками. Измерение связи. Корреляционный анализ.	Устный опрос, выполнение практических заданий, реферат	Вопросы к зачету № 1, 2, 9, 10, 19

	<p>программных средств</p> <p>Уметь: Для достижения индикатора УК-1.1: проводить анализ принципов структурной и функциональной организации биологических объектов и механизмов гомеостатической регуляции. Для достижения индикатора УК-1.2: работать в качестве пользователя в статистических пакетах; самостоятельно формулировать задачи математической статистики; правильно применять различные статистические критерии.</p> <p>Владеть: Для достижения индикатора УК-1.1: основными физиологическими методами анализа и оценки состояния живых систем; приемами планирования эксперимента. Для достижения индикатора УК-1.2: параметрическими и непараметрическими методами оценки, сравнения и характеристики данных медико-биологических исследований.</p>	<p>Регрессионный анализ. 9. Критерии эффективности отбора в животноводстве.</p>		
	<p>ПК-2 Знать: Для достижения индикатора ПК-2.1: принципы структурной и функциональной</p>	<p>2. Показатели изменчивости признаков. Основные характеристики вариационного ряда. Малые выборки и их</p>	<p>Устный опрос, выполнение практических заданий, реферат</p>	<p>Вопросы к зачету № 1, 2, 9, 10, 19</p>

<p>организации биологических объектов</p> <p>Уметь: Для достижения индикатора ПК-2.3: выбирать адекватные поставленным задачам специализированные статистические пакеты для анализа генетических данных.</p> <p>Владеть: Для достижения индикатора ПК-2.3: методами количественной оценки наследуемости и повторяемости признаков человека, растений и животных.</p>	<p>особенности.</p> <p>3. Анализ распределения. Нормальное распределение и его закономерности. Примеры типов распределения случайных величин.</p> <p>6. Дисперсионный анализ. Т-критерий Стьюдента.</p> <p>7. Непараметрические методы: критерий χ^2, точный критерий Фишера, критерий Манна-Уитни.</p> <p>8. Пакеты статистических программ и работа с ними</p>		
--	--	--	--

Примечание: типовые задания, критерии и показатели оценивания в рамках текущего контроля представлены в рабочей программе дисциплины (модуля). Полные комплекты оценочных средств и контрольно-измерительных материалов хранятся на кафедре.

3.2 Содержание оценочных средств

Оценочные средства промежуточной аттестации представлены перечнем вопросов для зачета.

Теоретические вопросы к зачету «Статистика в генетике»

1. Качественные признаки. Количественные признаки. Модели наследования количественных признаков.

Качественные признаки – категориальные данные; могут быть номинальными (категории альтернативные, неупорядоченные, такие как пол м/ж, группа крови А/В/АВ/0) или порядковыми, ранговыми (категории альтернативные, могут быть упорядоченными, такие как стадия заболевания легкая/средняя/тяжелая). Бинарный качественный признак – есть только две возможных категории, например, жив/мертв, болен/здоров. Измерение может принадлежать только к одной из категорий переменных.

Количественные признаки выражаются в числовых значениях. Дискретные данные – переменные могут принимать только определенные числовые значения (число особей в помете). Непрерывные данные – переменная может принимать любое значение (масса тела, рост, большинство хозяйственно-полезных признаков).

Количественным признакам невозможно дать точной качественной характеристики, между особями наблюдаются постепенные малозаметные переходы, а при расширении не образуются четко выделяемые фенотипические классы. Изменчивость подобных признаков в популяции является непрерывной, что связано с

действием многих генов, полигенов. Каждый из полигенов оказывает незначительное влияние на изменчивость количественного признака. Для характеристики популяции животных по количественным признакам невозможно использовать генетические параметры частоты и распределения генов. Закономерности наследования количественных признаков:

1) Средняя арифметическая признака у животного F₁, как правило, бывает промежуточной между средними арифметическими у родительских форм, а изменчивость признака, выражаемая величиной среднего квадратического отклонения, близка к изменчивости признака у родителей. Кривая изменчивости признака животных F₁ распределяется между кривыми родителей.

2) Между средними величинами признака у животных F₁ и F₂ различия недостоверны. Однако фенотипическая изменчивость животных F₂, выраженная в средних квадратичных отклонениях или вариантах, значительно выше, чем у животных F₁, что свидетельствует о расщеплении генотипов.

3) Кривые изменчивости животных, полученные от обратного скрещивания, сдвинуты ближе к кривым распределения тех родителей, с которыми было произведено скрещивание.

4) Кривая распределения особей третьего поколения, полученных из разных частей вариационного ряда второго поколения, располагается в зависимости от того, из какой части вариационного ряда взяты для скрещивания особи второго поколения.

2. Биометрия. Основные генетико-статистические величины и их применение.

Биометрия (от греч. *bios* — жизнь и *metreo* — измеряю) — направление биологических исследований, в которых находит широкое применение совокупность приемов математической обработки исследуемых данных (массовые измерения различных признаков организмов, популяций, биоценозов). Основными генетико-статистическими величинами можно назвать следующие параметры.

Средняя арифметическая – основной параметр, характеризующий совокупность изучаемого признака. Она показывает, какое значение признака наиболее характерно в целом для конкретной совокупности объектов. Средняя взвешенная представляет собой результат усреднения средних арифметических нескольких совокупностей.

При изучении изменчивости (вариабельности) признака особей данной совокупности применяют следующие параметры: лимит ($lim = V_{max} - V_{min}$), среднее квадратическое (стандартное) отклонение (δ), коэффициент вариации или изменчивости (CV, %). Наиболее простой показатель варьирования признака – величина лимита. Лимиты характеризуют минимальную и максимальную величины изучаемого признака в выборочной совокупности и указывает на амплитуду вариации. Чем больше разность между максимальной и минимальной вариантой, тем значительнее изменчивость признака. Чаще используются два других параметра. Среднее квадратическое отклонение или сигма (δ) позволяет судить о степени разнообразия признака в абсолютных величинах. Она выражается в тех же единицах, что и средняя арифметическая (в кг, см, л, % и т.д.). Сигма показывает, насколько однородна или разнообразна группа объектов по изучаемому признаку. Чем больше величина сигмы, тем выше изменчивость, и наоборот. Вся изменчивость признака укладывается от средней арифметической в пределах $\pm 3\delta$ (правило плюс-минус трех сигм). Для сопоставления изменчивости разноименных признаков и для выявления уровней изменчивости у одноименных признаков разных совокупностей при больших различиях средних арифметических величин сравниваемых групп удобнее пользоваться коэффициентом изменчивости (CV), который показывает изменчивость в относительных величинах – в процентах.

3. Основные характеристики вариационного ряда. Характеристика центра распределения (среднее арифметическое, мода, медиана).

Вариационный ряд - упорядоченная выборка, последовательность $X_1 \leq X_2 \leq \dots \leq X_n$ полученная в результате расположения в порядке неубывания исходной последовательности независимых одинаково распределённых случайных величин X_1, X_2, \dots, X_n . Центр распределения данных в вариационном ряде характеризуют такие меры положения, как средняя арифметическая величина, медиана, мода, средняя геометрическая величина, взвешенное среднее.

Среднюю арифметическую в многочисленной выборках ($n < 30$) вычисляют прямым способом, который заключается в суммировании всех величин признака у отдельных особей и деления полученной суммы на число особей. выборках (сложный способ). Для больших выборок ($n > 30$) предварительно строят вариационный ряд, в котором все члены распределяются по классам с учетом величины варьирующего признака. Дальнейшую обработку выборочной совокупности, оформленной в виде вариационного ряда, осуществляют различными методами, а именно: методом произведений, сумм или разностей. Недостатки использования средней величины – это значение легко искажается выбросами и ассиметричными данными.

Медиана характеризует усреднение в упорядоченном наборе данных, когда переменные упорядочены по величине от минимального значения к максимальному. Медиана делит ряд упорядоченных данных пополам, с равным числом этих значений как выше, так и ниже ее. Медиана близка к среднему значению, если распределение данных симметрично. Медиана не искажается выбросами и ассиметричными данными, но она игнорирует большую часть информации, не определяется алгебраически.

Мода – значение, встречающееся наиболее часто в наборе данных. Если данные непрерывны, для нахождения моды необходимо их сгруппировать и вычислить модальную группу. Некоторые наборы данных могут не иметь моды, если каждое значение встречается только один раз. В наборе данных может быть и более одной моды, если несколько значений встречаются одинаковое количество раз, превосходящее встречаемость других значений.

4. Характеристики вариации (среднее квадратическое отклонение, коэффициент вариации). Особенности обработки вариационных рядов в малых выборках.

Один из способов измерить рассеяние данных – определить степень отклонения каждого наблюдения от средней арифметической. Чем больше отклонение, тем больше вариабельность наблюдений. Среднее возведенных в квадрат отклонений каждого значения от средней арифметической называется дисперсией.

Стандартное (среднее квадратическое) отклонение δ – это положительный квадратный корень из дисперсии. Это тоже мера среднего отклонения наблюдений от среднего. Вычисляется по формуле:
$$\sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

Если разделить стандартное отклонение на среднее арифметическое и выразить этот показатель в процентах, получится коэффициент вариации (безразмерная мера рассеяния). Среднее квадратическое (стандартное) отклонение дает показатель изменчивости признака в именованных величинах (кг, см, л, %), а не в относительных. Вследствие этого сопоставление разноименных признаков по величине изменчивости с помощью лимита и стандартного отклонения произвести невозможно. В таком случае пользуются коэффициентом вариации. Но при сравнении двух групп особей по изменчивости одного признака следует пользоваться средними квадратическими отклонениями (δ), а не коэффициентами вариации (CV), зависящими от величины

средней арифметической.

Поскольку основные параметры совокупности (X , δ , CV и др.) вычисляют не по генеральной совокупности, а по выборочной, статистические параметры для выборки могут несколько отличаться от тех их значений, которые были бы получены для генеральной совокупности. Необходимо рассчитывать значение ошибки среднего, рассчитываемой по формулам: для малых выборок ($n < 30$) ошибка среднего $\frac{\delta}{\sqrt{n-1}}$; для больших выборок ($n \geq 30$) ошибка среднего $\frac{\delta}{\sqrt{n}}$. Ошибка среднего квадратического отклонения $\frac{\delta}{\sqrt{2n}}$; ошибка коэффициента изменчивости

$CV = \frac{\delta}{\sqrt{2CV}}$. Чем больше объем выборки приближается к объему генеральной

совокупности, и чем меньше варьирует признак, тем лучше выборка отражает свойства генеральной совокупности и тем меньше будут ошибки у всех выборочных коэффициентов.

При малом объеме выборки особенно важно решить вопрос о принадлежности крайних вариантов к генеральной совокупности. Исследование принадлежности всех элементов малой выборки к генеральной совокупности можно провести на основе использования распределения Стьюдента, сравнивая значение критерия t при данном числе степеней свободы f и данной величине вероятности P с табличными данными. Если нормированное наблюдаемое значение критерия t для испытуемой варианты превосходит по абсолютной величине соответствующее табличное значение при $P = 0,05$, то t считается значимым; в противном случае t не является значимым и соответствующее необходимо исключить из выборки.

5. Понятие о вероятности. Теоремы сложения и умножения вероятности.

Вероятность измеряет неопределенность, возможность появления данного события. Это положительное число в интервале от 0 до 1. При вероятности 0 событие произойти не может, при 1 – обязательно произойдет. Подходы к вычислению вероятности: частотная вероятность, выражающая соотношение количества событий, которые могли бы произойти, если бы эксперимент повторялся большое количество раз (например, вероятность «орал» или «решки» при подбрасывании монеты 1000 раз); априорная вероятность, требующая знания теоретической модели распределения вероятности, которая отображает вероятности всех возможных результатов эксперимента (например, вероятность распределения животных определенного окраса в потомстве родителей с известными генотипами).

Случайная переменная – это величина, которая может принимать любое из набора взаимоисключающих значений с определенной вероятностью. Распределение вероятности показывает вероятности всех возможных значений случайной переменной. Эти теоретические распределения выражены математически и имеют среднее и дисперсию, характеризующие данное распределение. Площадь под кривой распределения равна единице и отражает вероятность всех возможных событий.

Теорема сложения вероятности: если два события, A и B , взаимоисключающие, несовместимые, то вероятность того, что произойдет то или иное событие, равна сумме их вероятностей. $P_{(A \text{ или } B)} = P_A + P_B$.

Теорема умножения вероятности: если два события, A и B , независимы, возникновение одного события не влияет на возможность появления другого, то вероятность того, что оба события произойдут, равна произведению вероятности каждого. $P_{(A \text{ и } B)} = P_A \times P_B$.

6. Нормальное распределение и его параметры. Биноминальное

распределение.

Фонд оценочных средств промежуточной аттестации по дисциплине «Статистика в генетике» по направлению подготовки (специальности) "06.04.01 Биология" направленности (профилю) Генетика ФГБОУ ВО

стр. 9

Нормальное (гауссовское) распределение одно из важнейших в статистике. Его функция плотности полностью определяется двумя параметрами, средним и дисперсией. Распределение имеет колоколообразную форму, унимодально, симметрично относительно среднего; сдвигается вправо, если среднее увеличивается, и влево, если среднее уменьшается (при постоянной дисперсии); становится более широким, плоским, если дисперсия увеличивается, и более остроконечным, узким, если дисперсия уменьшается (при постоянном среднем).

Среднее и медиана нормального распределения равны. Вероятность того, что нормально распределенная случайная переменная x , со средним M и стандартным отклонением δ , находящаяся между:

$(M - \delta)$ и $(M + \delta)$, равна 0,68;

$(M - 1,96 \delta)$ и $(M + 1,96 \delta)$, равна 0,95;

$(M - 2,58 \delta)$ и $(M + 2,58 \delta)$, равна 0,99.

Существует множество нормальных распределений в зависимости от значений M и δ . Стандартное нормальное распределение имеет среднее, равное 0, и дисперсию, равную 1. Вероятности стандартного нормального распределения приведены в справочных таблицах. Если случайная переменная x имеет нормальное распределение со средним M и дисперсией δ^2 , тогда стандартизованное нормальное отклонение z будет случайной переменной, которая имеет стандартное нормальное распределение. $z = \frac{x - M}{\delta}$

Биномиальное распределение – это распределение вероятности дискретных переменных. Можно получить вероятности, соответствующие каждому возможному значению случайной переменной, сумма всех таких вероятностей равна 1. Это распределение количества «успехов» в последовательности из n независимых случайных экспериментов, таких, что вероятность «успеха» в каждом из них постоянна и равна p .

7. Доверительные интервалы. Средние ошибки средней арифметической, среднего квадратического отклонения и коэффициента вариации.

Анализируя выборку из популяции, исследователь получает точечную оценку параметров. Доверительный интервал — показатель интервальной оценки статистических параметров, более предпочтительной при небольшом объеме выборки, чем точечная. Доверительным называют интервал, который покрывает неизвестный параметр с заданной надёжностью (рассчитываются для заданного уровня значимости). Доверительные интервалы позволяют сделать утверждение, что истинное значение неизвестного статистического параметра генеральной совокупности находится в полученном диапазоне значений с вероятностью, которая задана выбранным уровнем статистической значимости. Когда известна вариация (δ^2) генеральной совокупности данных, для расчета доверительных пределов может быть использована z -оценка, основанная на нормальном распределении.

Выборочное среднее M имеет нормальное распределение, когда объем выборки большой, и можно считать распределение выборочного среднего нормальным. 95 % распределения выборочных средних находится в пределах $1,96 \delta$ стандартных отклонений среднего популяции. Когда есть только одна выборка, нижняя граница 95 % доверительного интервала для среднего рассчитывается как $M - (1,96 \times \delta)$, верхняя граница как $M + (1,96 \times \delta)$. В пределах этого интервала с доверительной вероятностью 95 % находится истинное среднее популяции (генеральной совокупности).

Значение ошибки среднего для малых выборок ($n < 30$) $\frac{\delta}{\sqrt{(n-1)}}$; для больших выборок ($n \geq 30$) ошибка среднего $\frac{\delta}{\sqrt{n}}$. Ошибка среднего квадратического отклонения $\delta \frac{\delta}{\sqrt{2n}}$; ошибка коэффициента изменчивости $\frac{\delta}{\sqrt{2n}}$.

$$m_{CV} = \frac{\delta}{\sqrt{2CV}}.$$

8. Сравнение средних арифметических и других характеристик вариационных рядов.

О сравнительном значении характеристик вариационного ряда говорят в тех случаях, когда рассматривают два или более ряда чисел. Пользуясь характеристиками вариационных рядов, сравнению подлежат средние арифметические между собой и факторы рассеивания между собой. В этом случае сравнение двух рядов чисел сводится к сравнению двух чисел, что всегда очевидно.

Часто требуется установить, есть ли отличия в двух или более сравниваемых вариационных рядах, когда один ряд – это результаты измерения некоторого параметра одной экспериментальной группы, а второй ряд – измерения второй экспериментальной группы. Для сравнения групп выдвигают нулевую и альтернативную гипотезы. Нулевая гипотеза H_0 отвергает наличие эффекта, говорит о том, что обе группы принадлежат к одной генеральной совокупности (разница в средних значениях для двух выборок равна нулю). Альтернативная гипотеза H_1 принимается, если нулевая гипотеза не верна (выборки принадлежат к разным генеральным совокупностям).

При сравнении средних может быть использован двусторонний критерий, который говорит о том, что средние могут отличаться друг от друга в обе стороны; и односторонний критерий, когда отличия могут быть только однонаправленными (только больше или только меньше).

Используют разные статистические критерии для сравнения средних и их дисперсии. Затем необходимо определить р-значение для рассчитанного критерия. Для этого необходимо сравнить значение критерия с уже известным распределением, которому он подчиняется, чтобы определить р-значение. Р-уровень значимости – это вероятность получить вычисленное значение критерия или еще большее значение, если нулевая гипотеза верна. Чем меньше р-значение, тем сильнее аргументы против H_0 . Традиционно H_0 отвергается при $p \leq 0,05$ (5 % уровень значимости). Уровень 5 % является в известной степени произвольным, можно установить меньшую границу отсечки (0,01 или 0,001), называемую критическим уровнем значимости критерия.

Какой именно статистический критерий использовать в конкретном случае зависит от типа переменных, плана эксперимента, распределения, которому отвечают переменные.

9. Функциональная связь и коррелятивная изменчивость. Коэффициент корреляции. Оценка коэффициента корреляции.

Явление связи между различными признаками широко распространено в природе. Связь признаков может быть подразделена на функциональную и коррелятивную. Для функциональной связи характерно то, что при изменении одного показателя на определенную величину другой, с ним сопряженный показатель принимает всегда одно определенное значение. При коррелятивной связи если один признак изменяется на какую-то определенную величину, то другой может принимать различные значения. Корреляцией называется взаимосвязь между признаками в организме или особями одного пола и вида.

Коэффициент корреляции r – основной биометрический показатель, позволяющий определить величину связи между двумя, тремя и большим числом признаков. Величина этого коэффициента имеет дробное значение в пределах от 0 до ± 1 . Чем ближе показатель к единице, тем больше связь между коррелирующими признаками.

По форме корреляция может быть прямолинейной и криволинейной, по

направлению – положительной (прямой) и отрицательной (обратной), на что указывает знак «плюс» или «минус». Прямой называется такая корреляция, когда с увеличением одного признака увеличивается и другой. Приняты три степени корреляции: $r = 0,1-0,3$ – связь слабая; $0,4-0,6$ – средняя; $0,7-1,0$ – высокая.

Коэффициент корреляции в малых выборках широко применяют при определении связи между признаками у родственных групп животных для решения вопросов связанных с выявлением закономерностей наследования и наследственности. Наиболее удобны для вычисления коэффициента фенотипической корреляции в малых выборках следующие формулы:

$$r = \frac{\sum x \cdot y - \frac{\sum x \cdot \sum y}{n}}{\sqrt{C_x \cdot C_y}}, \quad r = \frac{C_x + C_y - C_d}{2 \cdot \sqrt{C_x \cdot C_y}}$$

где n – число наблюдений в выборке; x и y – значение вариант первого и второго признака; C – сумма квадратов центральных отклонений, вычисляемая по формуле

$$C_x = \sum x^2 - \frac{(\sum x)^2}{n}, \quad C_y = \sum y^2 - \frac{(\sum y)^2}{n}, \quad C_d = \sum d^2 - \frac{(\sum d)^2}{n}$$

10. Понятие о регрессии. Коэффициент регрессии. Оценка коэффициента регрессии.

Линейная регрессия позволяет оценить соотношение между двумя непрерывными переменными x и y , измеренными для одних и тех же объектов. Если предполагается, что y зависит от x , причем именно изменения в x вызывают изменения в y , можно разместить все наблюдения на двумерной диаграмме рассеяния. Если данные аппроксимируются прямой линией, говорят о линейном соотношении, и возможно определить линию регрессии y на x , лучше всего описывающую прямолинейное соотношение между этими двумя переменными. Одномерный анализ учитывает изменение только одной переменной x , множественная регрессия может описывать изменение y с участием двух и более объясняющих переменных.

Математическое уравнение, описывающее линию простой линейной регрессии: $y = a + bx$. Здесь x – независимая, объясняющая переменная, или предиктор; y – зависимая переменная, переменная отклика. Для данной величины x , y – значение, которое можно ожидать в среднем при известной величине x (предсказанное значение y). a – свободный член линии регрессии, значение y при $x = 0$. b – угловой коэффициент, или коэффициент регрессии, величина, на которую y увеличивается в среднем при увеличении x на одну единицу.

a и b определяют методом наименьших квадратов таким образом, чтобы расстояние от каждой точки на двумерном графике рассеяния до линии регрессии было оптимальным (сумма квадратов этих расстояний была наименьшей). Вертикальное расстояние от каждой точки до линии регрессии называют остатком.

Необходимые допущения: между x и y существует линейное соотношение; наблюдения в выборке независимы друг от друга; для каждой величины x имеется распределение величин y в популяции, и это распределение нормальное.

11. Дисперсионный анализ и его сущность. Общая схема дисперсионного анализа.

Дисперсионный анализ — метод, направленный на поиск зависимостей в экспериментальных данных путём исследования значимости различий в средних значениях. В отличие от t -критерия, позволяет сравнивать средние значения трёх и более групп. Разработан Р. Фишером, встречается обозначение метода как ANOVA.

Суть дисперсионного анализа сводится к изучению влияния одной или

© ФГБОУ ВО «ЧелГУ»

нескольких независимых переменных, обычно именуемых факторами, на зависимую переменную. Зависимые переменные представлены значениями абсолютных шкал (шкала отношений). Независимые переменные являются номинативными (шкала наименований), то есть отражают групповую принадлежность, и могут иметь два или более значения (типа, градации или уровня). Примерами независимой переменной с двумя значениями могут служить пол (женский, мужской) или тип экспериментальной группы (контрольная, экспериментальная). Градации, соответствующие независимым выборкам объектов, называются межгрупповыми, а градации, соответствующие зависимым выборкам, — внутригрупповыми.

В зависимости от типа и количества переменных различают: однофакторный и многофакторный дисперсионный анализ (одна или несколько независимых переменных); одномерный и многомерный дисперсионный анализ (одна или несколько зависимых переменных); дисперсионный анализ с повторными измерениями (для зависимых выборок); дисперсионный анализ с постоянными факторами, случайными факторами, и смешанные модели с факторами обоих типов.

Процедура дисперсионного анализа состоит в определении соотношения систематической (межгрупповой) дисперсии к случайной (внутригрупповой) дисперсии в измеряемых данных. В качестве показателя изменчивости используется сумма квадратов отклонения значений параметра от среднего SS (от англ. Sum of Squares). Можно показать, что общая сумма квадратов SS раскладывается на межгрупповую сумму квадратов и внутригрупповую сумму квадратов. Соотношение межгрупповой и внутригрупповой дисперсий имеет F -распределение (распределение Фишера) и определяется при помощи F -критерия Фишера.

12. Дисперсионный анализ двух-и многофакторного опытов.

Простейшим случаем дисперсионного анализа является одномерный однофакторный анализ для двух или нескольких независимых групп, когда все группы объединены по одному признаку. В ходе анализа проверяется нулевая гипотеза о равенстве средних. При анализе двух групп дисперсионный анализ тождественен двухвыборочному t -критерию Стьюдента для независимых выборок, и величина F -статистики равна квадрату соответствующей t -статистики.

Для подтверждения положения о равенстве дисперсий обычно применяется критерий Ливена. В случае отвержения гипотезы о равенстве дисперсий основной анализ неприменим. Если дисперсии равны, то для оценки соотношения межгрупповой и внутригрупповой изменчивости применяется F -критерий Фишера. Если F -статистика превышает критическое значение, то нулевая гипотеза не может быть принята (отвергается) и делается вывод о неравенстве средних. При наличии трёх и более групп требуется попарное сравнение средних для выявления статистически значимых отличий между ними.

Многофакторный анализ позволяет проверить влияние нескольких факторов на зависимую переменную. Модель многофакторного анализа включает суммы квадратов для каждого фактора в отдельности и суммы квадратов всех взаимодействий между ними. Так, в двухфакторной модели межгрупповая сумма квадратов раскладывается на сумму квадратов фактора A , сумму квадратов фактора B и сумму квадратов взаимодействия факторов A и B .

В ходе анализа проверяются несколько нулевых гипотез: гипотеза о равенстве средних под влиянием фактора A ; гипотеза о равенстве средних под влиянием фактора B ; гипотеза об отсутствии взаимодействия факторов A и B . Каждая гипотеза проверяется с помощью критерия Фишера. При отвержении нулевой гипотезы о влиянии отдельного фактора принимается утверждение, что присутствует главный эффект фактора A (B и т. д.). При отвержении нулевой гипотезы о взаимодействии факторов принимается

утверждение о том, что влияние фактора А проявляется по-разному на разных уровнях фактора В. В таком случае результаты общего анализа признаются не имеющими силы, и влияние фактора А проверяется отдельно на каждом уровне фактора В с помощью однофакторного дисперсионного анализа или t-критерия.

13. Т-критерий Стьюдента.

t-критерий Стьюдента — общее название для класса методов статистической проверки гипотез (статистических критериев), основанных на распределении Стьюдента. Наиболее частые случаи применения t-критерия связаны с проверкой равенства средних значений в двух выборках.

t-статистика строится обычно по следующему общему принципу: в числителе — случайная величина с нулевым математическим ожиданием (при выполнении нулевой гипотезы), а в знаменателе — выборочное стандартное отклонение этой случайной величины, получаемое как квадратный корень из несмещённой оценки дисперсии.

Для применения данного критерия необходимо, чтобы исходные данные имели нормальное распределение. В случае применения двухвыборочного критерия для независимых выборок также необходимо соблюдение условия равенства дисперсий.

Одновыборочный t-критерий применяется для проверки нулевой гипотезы H_0 о равенстве математического ожидания некоторому известному значению m , формула для

$$t = \frac{\bar{X} - m}{s_X / \sqrt{n}}.$$

t-статистики имеет вид $t (n - 1)$. При нулевой гипотезе распределение этой статистики $t (n - 1)$. Следовательно, при превышении значения статистики по абсолютной величине критического значения данного распределения (при заданном уровне значимости) нулевая гипотеза отвергается.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

Двухвыборочный t-критерий для независимых выборок имеет вид

Для вычисления эмпирического значения t-критерия в ситуации проверки гипотезы о различиях между двумя зависимыми выборками (например, двумя пробами одного и того же теста с временным интервалом) применяется следующая формула:

$$t = \frac{M_d}{s_d / \sqrt{n}},$$

где M_d — средняя разность значений, s_d — стандартное отклонение разностей, а n — количество наблюдений.

14. Непараметрические методы. Критерий χ^2 , точный критерий Фишера.

Критерий хи-квадрат — любая статистическая проверка гипотезы, в которой выборочное распределение критерия имеет распределение хи-квадрат при условии верности нулевой гипотезы. Распределение хи-квадрат — это распределение суммы квадратов k независимых стандартных нормальных случайных величин. Считается, что критерий хи-квадрат — это критерий, который асимптотически верен, то есть, выборочное распределение можно сделать как угодно близким к распределению хи-квадрат путём увеличения размера выборки. Некоторые критерии имеют распределение хи-квадрат только в приближении.

Критерий согласия Пирсона или критерий согласия χ^2 — если критерий хи-квадрат упоминается без каких-либо модификаций или без другого исправляющего контекста, используется критерий χ^2 Пирсона. Для точного теста, используемого вместо χ^2 , применяется точный тест Фишера.

Критерий согласия Пирсона, или критерий согласия χ^2 , — это непараметрический метод, который позволяет оценить значимость различий между фактическим

(выявленным в результате исследования) количеством исходов или качественных характеристик выборки, попадающих в каждую категорию, и теоретическим количеством, которое можно ожидать в изучаемых группах при справедливости нулевой гипотезы. Метод позволяет оценить статистическую значимость различий двух или нескольких относительных показателей (частот, долей).

Является наиболее часто употребляемым критерием для проверки гипотезы о принадлежности наблюдаемой выборки объемом n некоторому теоретическому закону распределения F . Процедура проверки гипотез с использованием критериев типа χ^2 предусматривает группирование наблюдений. Область определения случайной величины разбивают на k непересекающихся интервалов граничными точками. При использовании критериев согласия, как правило, не задают конкурирующих гипотез: рассматривается принадлежность выборки конкретному закону.

Точный тест Фишера — тест статистической значимости, используемый в анализе таблиц сопряженности для выборок маленьких размеров. Относится к точным тестам значимости, поскольку не использует приближения большой выборки (асимптотики при размере выборки стремящемся к бесконечности). Тест обычно используется, чтобы исследовать значимость взаимосвязи между двумя переменными в факторной таблице размерности 2×2 (таблице сопряженности признаков). Величина вероятности p теста вычисляется, как если бы значения на границах таблицы известны.

15. Непараметрические методы. Критерий Манна-Уитни. Условия применения параметрических и непараметрических методов.

Для применения параметрических критериев необходимо, чтобы исходные данные имели нормальное распределение, а часто также требуется соблюдение условия равенства дисперсий. При несоблюдении этих условий при сравнении выборочных средних должны использоваться аналогичные методы непараметрической статистики, среди которых наиболее известными являются U -критерий Манна-Уитни (в качестве двухвыборочного критерия для независимых выборок), а также критерий Уилкоксона (используются в случаях зависимых выборок).

U -критерий Манна-Уитни — статистический критерий, используемый для оценки различий между двумя независимыми выборками по уровню какого-либо признака, измеренного количественно. Позволяет выявлять различия в значении параметра между малыми выборками. Этот метод определяет, достаточно ли мала зона перекрещивающихся значений между двумя рядами (ранжированным рядом значений параметра в первой выборке и таким же во второй выборке). Чем меньше значение критерия, тем вероятнее, что различия между значениями параметра в выборках достоверны. В каждой из выборок должно быть не менее 3 значений признака. Допускается, чтобы в одной выборке было два значения, но во второй тогда не менее пяти. В выборочных данных не должно быть совпадающих значений (все числа — разные) или таких совпадений должно быть очень мало (до 10).

Рассчитывается U -критерий Если полученное значение U меньше табличного или равно ему, то признается наличие существенного различия между уровнем признака в рассматриваемых выборках (принимается альтернативная гипотеза). Если же полученное значение U больше табличного, принимается нулевая гипотеза.

16. Основные статистические пакеты: SPSS, Статистика, SigmaPlot.

SPSS Statistics — программа для статистической обработки данных, один из лидеров рынка в области коммерческих статистических продуктов, предназначенных для проведения прикладных исследований, главным образом в общественных науках. Программа позволяет осуществлять ввод данных любого типа, включая номинативные. Возможно редактирование свойств переменной. Сформированные матрицы SPSS в

дальнейшем пригодны для разных видов анализа, доступны для редактирования и слияния с другими матрицами данных. Частотный анализ, корреляционный анализ, дисперсионный анализ (однофакторный и многофакторный). Все типы регрессионного анализа и установление необходимых показателей: нелинейная регрессия, логистическая регрессия, регрессия Кокса. Дискриминантный анализ, факторный и ковариационный анализ, кластерный анализ.

Statistica — программный пакет для статистического анализа, реализующий функции анализа данных, управления данными, добычи данных, визуализации данных с привлечением статистических методов. Пакет обладает широкими графическими возможностями, позволяет выводить информацию в виде различных типов графиков (включая научные, деловые, трёхмерные и двумерные графики в различных системах координат, специализированные статистические графики — гистограммы, матричные, категоризованные графики и др.), все компоненты графиков настраиваются. Существуют различные варианты пакета в зависимости от целей и задач пользователя. Широкий набор основных статистик и графических инструментов, возможность подключить модули с продвинутыми функциями – мощные инструменты для построения линейных/нелинейных моделей, многомерные технологии анализа данных, инструменты для анализа мощности и вычисления объема выборки.

SigmaPlot — программа для анализа и визуализации научных и статистических данных. Для анализа математических и статистических данных в программе предусмотрены: ввод табличных данных, математическая обработка табличных данных, широкий выбор графиков и диаграмм, широкий выбор способа оформления графиков и диаграмм, регрессионный анализ с широким выбором вида уравнений, редактирование графиков. Интеграция с пакетом Office: доступ к электронным таблицам XLS и экспорт результатов в виде презентаций PowerPoint.

17. Статистические средства в табличных процессорах управления базами данных MSEXcel, QuattroPro, Lotus1-2-3.

Microsoft Excel — программа для работы с электронными таблицами, созданная корпорацией Microsoft. Она предоставляет возможности экономико-статистических расчетов, графические инструменты. Основными средствами анализа статистических данных в Excel являются статистические процедуры надстройки Пакет анализа (Analysis ToolPак) и статистические функции библиотеки встроенных функций. В Пакет анализа входят следующие статистические процедуры: описательная статистика; ранги перцентиль; двухвыборочный z-тест для средних; двухвыборочный t-тест для средних с одинаковыми и различными дисперсиями; парный двухвыборочный t-тест для средних; двухвыборочный F-тест для дисперсий; ковариация; корреляция; регрессия; однофакторный дисперсионный анализ; двухфакторный дисперсионный анализ; анализ Фурье.

Lotus 1-2-3 — табличный процессор. Программа получила название «1-2-3», так как состояла из трёх элементов: таблиц, графиков и основных функциональных возможностей базы данных. В настоящее время не выпускается.

Quattro Pro – редактор электронных таблиц, входит в офисный пакет Corel WordPerfect Office. Одной из отличительных особенностей процессора Quattro Pro являются аналитические графики, которые позволяют применять к исходным данным агрегирование, вычислять скользящее среднее и проводить регрессионный анализ; результаты перечисленных действий отражаются на графиках. Набор встроенных статистических функций в пакете Quattro Pro включает в себя все стандартные функции.

18. Специализированные статистические пакеты для генетических данных: SNPSStats.

SNPStats - это простое, готовое к использованию программное обеспечение, разработанное для анализа генетико-эпидемиологических исследований ассоциации с использованием SNP. При загрузке данных для каждого выбранного SNP возможно получить: частоты аллелей и генотипов; тест на равновесие Харди-Вайнберга, анализ связи с переменной ответа на основе линейной или логистической регрессии; модели множественного наследования (кодминантная, доминантная, рецессивная, сверхдоминантная и аддитивная); анализ взаимодействий (ген-ген или ген-среда).

При выборе нескольких SNP возможны следующие виды анализа: статистика неравновесия по сцеплению, оценка частоты гаплотипа, анализ ассоциации гаплотипов с ответом, анализ взаимодействий (гаплотипы-ковариаты).

19. Наследуемость признаков. Повторяемость признаков. Критерии достоверности и соответствия. Эффект селекции.

Селекция сельскохозяйственных животных направлена на изменение свойств популяции в желательном направлении. Разнообразие признака может быть обусловлено как наследственностью, так и факторами среды. Для улучшения следующих поколений нужно знать величину генетического разнообразия признака, называемого наследуемостью и обозначаемую h^2 . Наследуемость – это доля генетического разнообразия в общей изменчивости признака. Коэффициент наследуемости вычисляется различными методами: 1. С помощью коэффициента корреляции (r) между показателями дочерей и теми же показателями матерей по формуле $h^2 = 2 \times r_{\text{мд}}$, где h^2 – коэффициент наследуемости; $r_{\text{мд}}$ – коэффициент корреляции между признаками матерей и их дочерей (или сыновей и отцов). 2. При разной интенсивности отбора среди родителей и потомков показатели разнообразия признака резко различаются. В таких случаях целесообразно пользоваться формулой $h^2 = 2 \times R_{\text{мд}}$, где $2R$ – удвоенный коэффициент регрессии между фенотипами родственников. Однако, из-за неодинаковой наследуемости различных признаков, эффективность отбора по ним будет различной. Успешнее будет отбор при достаточной наследственной (генетической) неоднородности стада, т. е. при значительной величине коэффициента наследуемости (h^2), который может быть вычислен по следующей

$$h^2 = \frac{D_{\text{лучш.}} - D_{\text{худш.}}}{M_{\text{лучш.}} - M_{\text{худш.}}} \cdot 2$$

формуле: $h^2 = \frac{D_{\text{лучш.}} - D_{\text{худш.}}}{M_{\text{лучш.}} - M_{\text{худш.}}} \cdot 2$, где h^2 – коэффициент наследуемости; $M_{\text{лучш.}}$ – средние показатели группы лучших животных стада; $M_{\text{худш.}}$ – средние показатели группы худших животных стада; $D_{\text{лучш.}}$ – средний показатель потомства полученного от лучших животных стада; $D_{\text{худш.}}$ – средний показатель потомства от худших животных стада.

При проведении генетического анализа количественных признаков выявляют коэффициент повторяемости – r_w (within-внутри). Его свойства заключаются в следующем: является показателем генетического разнообразия; является мерой верхнего предела коэффициента наследуемости; определяет надежность вносимых поправок в варьирующий признак с учетом изменения средовых факторов; служит мерой определения ошибки измеряемого признака. Коэффициент повторяемости вычисляют путем определения корреляции между продуктивностью одних и тех же животных по одним и тем же признакам, но за различные возрастные периоды, а также с помощью дисперсионного анализа с использованием однофакторного или двухфакторного комплекса.

При любом типе искусственного отбора важно определить, насколько он будет эффективным. Ожидаемый эффект селекции (R) рассчитывают по формуле:

$$= \frac{R}{j} \text{ где } h^2 \text{ – коэффициент наследуемости признака; } d \text{ – селекционный}$$

дифференциал; j – интервал между поколениями. Селекционный дифференциал – это разница между отобранными особями по селекционному признаку и средним значением данного признака по стаду.

4. ПОРЯДОК ПРОВЕДЕНИЯ И КРИТЕРИИ ОЦЕНИВАНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

4.1. Порядок проведения промежуточной аттестации

В рамках **текущего контроля** в течение семестра для оценки знаний, умений, навыков, получаемых в ходе изучения дисциплины, учитываются ответы на устные поросы, заслушивание сообщений (реферат), успешность выполнения практических заданий.

Критерием успешности освоения учебного материала **по окончании учебного семестра** (промежуточная аттестация) является экспертная оценка преподавателя, учитывающая: текущую успеваемость в течение семестра (устный опрос, выполнение практических заданий, заслушивание сообщений (реферат)). Кроме того, экспертная оценка преподавателя может основываться на регулярности посещения обязательных учебных занятий, успешности выполнения установленных на данный семестр объемов рабочей программы.

Промежуточная аттестация проводится в форме зачета по билетам. Билет содержит два теоретических вопроса.

4.2. Критерии оценивания промежуточной аттестации по видам оценочных средств

4.2.1. Критерии оценивания теоретического вопроса

«зачтено» содержание материала раскрыто, требующий лишь незначительных уточнений и дополнений, которые студент может сделать самостоятельно после наводящих вопросов преподавателя. Допускаются такие незначительные недочеты в ответе студента как отсутствие самостоятельного вывода, нарушение последовательности в изложении, речевые ошибки и др.

«не зачтено» - студент не может изложить содержание материала, не знает основных понятий дисциплины, не отвечает на дополнительные и наводящие вопросы преподавателя

4.3. Результаты промежуточной аттестации и уровни сформированности компетенций

При подведении итогов учитываются результаты текущей аттестации.

Особенности проведения процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья обозначены в рабочей программе дисциплины (модуля).

«1 уровень» - ознакомление (иметь общее представление, узнавать);

«2 уровень» - понимание учебного материала, излагаемого в учебнике, методической разработке или преподавателем;

«3 уровень» - умение логично, последовательно, достаточно полно и точно излагать изученный материал;

«4 уровень» - творчески использовать полученные знания.

Для удовлетворительной (положительной) оценки знаний требуется минимум 3-й уровень усвоения учебного материала.

Требования (критериальные показатели) к уровню освоения дисциплины

Результат зачета	Требования к знаниям
Зачтено	содержание материала раскрыто, требующий лишь незначительных уточнений и дополнений, которые студент может сделать самостоятельно после наводящих вопросов преподавателя. Допускаются такие незначительные недочеты в ответе студента как отсутствие самостоятельного вывода, нарушение последовательности в изложении, речевые ошибки и др.
Не зачтено	студент не может изложить содержание материала, не знает основных понятий дисциплины, не отвечает на дополнительные и наводящие вопросы преподавателя.

06.04.01 Биология, ОПОП Генетика, ФОС РПД Статистика в генетике, год набора 2025, форма обучения очная

Проректор по учебной работе утверждено 24.02.2025 А.А. Саламатов

Ученым советом биологического факультета

Протокол заседания № 6 от 21.02.2025

Председатель Ученого совета

биологического факультета согласовано Д.С. Сташкевич

Заседанием кафедры радиационной биологии

Протокол заседания № 7 от 21.02.2025

Заведующий кафедрой согласовано А.В. Аклеев

Автор (составитель) В.С. Никифоров

Структура рабочей программы соответствует приказу ректора ФГБОУ ВО «ЧелГУ» от «13» апреля 2021 г. № 247-1