МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ
ИРКУТСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ
имени А.А. ЕЖЕВСКОГО

А.К. Гордеева, Н.Б. Сверлова

ПРАКТИКУМ ПО БИОМЕТРИИ

Иркутск, 2020

УДК 636.082 БК

Составили: доцент, к. с-х. н Гордеева А.К., доцент, к. с-х. н Сверлова Н.Б.

Биометрия: методические указания/А.К. Гордеева, Н.Б. Сверлова- Иркутск: РИЦ Иркутский ГАУ, 2020.- 36 с.

В методических указаниях дается материал для лабораторно-практических занятий по основам биометрии, биометрических методов анализа качественных и количественных признаков.

Методические указания предназначены для подготовки бакалавров, магистров и аспирантов по направлению подготовки Зоотехния. Можно использовать на курсах повышения квалификации зоотехников, зоотехников-селекционеров.

Утверждена и рекомендована к изданию методической комиссией факультета биотехнологии и ветеринарной медицины (протокол № 9 от)

Рецензент:

Буквенные обозначения

P – вероятность.
n – число наблюдений в выборке.
N – число наблюдений в генеральной совокупности.
V – варьирующий признак или варианты.
X – средняя арифметическая.
Хвзв – средневзвешенная арифметическая.
К – величина класса (классового промежутка).
l – число классов в вариационном ряду.
D – разность между максимальным и минимальным значением варьирующего признака.
A – условная средняя.
a – условное отклонение классов от условной средней.
σ- среднее квадратическое отклонение.
C – дисперсия.
^{2} – варианса (девиата)
Cv – коэффициент изменчивости.
υ – число степеней свободы.
f – частоты.
r – коэффициент корреляции между количественными признаками.
m – статистическая ошибка.
td – критерий достоверности.
F – критерий достоверности Фишера,
n_i – частное число наблюдений по градациям.
κ – поправочный коэффициент, корректирующий дисперсию неортогонального комплекса.
n_x – число наблюдений по градациям комплекса.
X^2 - критерий совпадения наблюдаемых частот с теоретическими.
h^2 - коэффициент наследуемости,
r_w – коэффициент повторяемости.
S – селекционный дифференциал,
i -интенсивность селекции.
J- число лет между поколениями

ВВЕДЕНИЕ

Вариационная статистика представляет собой один из разделов высшей математики. Применение вариационно-статистического метода при анализе массовых данных в области биологии получило название ***биометрия.*** Биометрия является составной частью количественной биологии.

Основная задача биометрии – науки о применении статистических (математических) методов изучения живых организмов заключается в получении комплекса параметров и коэффициентов, характеризующих членов изучаемой группы по одному или нескольким признакам. В биометрии такой массовый материал называется генеральной совокупностью, которая и составляет цель изучения.

Биометрия позволяет систематизировать и обрабатывать числовые данные, получаемые при изучении биологических объектов в условиях экспериментов, а также при обобщении производственных первичных записей, проводимых в животноводстве или других отраслях сельского хозяйства. Особенно большие перспективы для использования биометрического метода даёт племенная документация, обобщение которой позволяет решать ряд вопросов племенного дела, селекции и генетического анализа.

Современная генетика и селекция сельскохозяйственных животных широко используют вариационно-статистический метод при генетическом анализе различных популяций (пород, стад, линий, семейств) в отношении количественных признаков. Пользуясь этим методом, устанавливают степень наследуемости признаков, определяют эффект селекции и интенсивность отбора.

Если генетический анализ наследования качественных признаков основывается на скрещивании и выявлении наследования потомством признаков родителей, то генетический анализ наследования количественных признаков опирается на биометрический метод. Большое значение приобретает этот метод, когда требуется сопоставить эмпирическое (основное на опыте), расщепление признаков у потомства с теоретически ожидаемым и установить достоверность получаемых различий в теоретическом и эмпирическом соотношении фенотипов и генотипов потомства.

Математический анализ массовых данных, исходящий из биометрического метода, находит широкое применение при решении ряда теоретических и практических вопросов генетики, селекции и племенного дела.

1. Использование методов биометрии в генетике и зоотехнии

Занятие 1. **Основные понятия терминов и положений биометрии. Методы изучения изменчивости.**

Цель занятия. Ознакомление с различными терминами и положениями биометрии, усвоение методов вычисления средних величин в зависимости от поставленной задачи и количества животных в выборке.

Методические указания. Применение вариационно-статистического метода при анализе массовых показателей в области биологии получило название **биометрия.**

Биометрия является составной частью количественной биологии и смежных с ней прикладных наук (зоотехния, ветеринария, агрономия, медицина).

Биометрия – это математический раздел науки, изучающий изменчивость количественных признаков живых организмов одного вида.

В основе современной селекции лежит генетический анализ продуктивных признаков животных. Сельскохозяйственные животные отличаются разнообразием хозяйственно полезных (продуктивных), морфологических, физиологических признаков. Однако лишь некоторые из них служат объектом практической селекции. Все хозяйственно полезные признаки животных подразделяют на качественные и количественные.

К качественным признакам относят пол (мужской и женский), окраску оперения птиц и шерстного покрова животных (альбинизм, пигментированность, пятнистость и др.), тип шерстного покрова (грубая, тонкая шерсть овец, смушки, овчины), рогатость, тип телосложения (конституция грубая, крепкая, рыхлая, нежная, плотная) форму гребня у кур, окраску скорлупы яиц и т.д.

Качественные признаки контролируются одним или несколькими генами, на действие которых не влияют ненаследственные факторы. Наследование качественных признаков подчиняется закономерностям, определённым Г. Менделем. Каждой паре качественных признаков (рогатость – комолость) соответствует пара аллельных генов (*кк – КК, Кк*), контролирующих развитие данных признаков.

Многие качественные признаки имеют только два альтернативных состояния, например, пол - мужской или женский, скорлупа яиц - окрашенная или неокрашенная и т.д. Некоторые качественные признаки могут иметь 3 – 5 состояний, например, тип конституции, интенсивность окраски скорлупы яиц, зёрен цветной кукурузы и др.

Большинство хозяйственно полезных признаков, по которым ведут селекцию животных, относят к **количественным,** или мерным. Эти признаки получили такое название потому, что они могут быть измерены и выражены в кг, см, % и т.п. К ним относятся удой, живая масса, содержание жира и белка в молоке, яйценоскость, масса яйца и др.

Развитие каждого количественного признака обусловлено большим числом пар генов, поэтому изучать процесс наследования таких признаков значительно труднее, чем качественных. При изучении количественных признаков приходится сталкиваться с непрерывной изменчивостью, когда переход от одного количественного уровня признака к другому составляет непрерывный ряд величин. Такая изменчивость количественных признаков обусловлена как действием большого числа генов, так и влиянием факторов внешней среды.

Основная задача биометрии – изучение изменчивости количественных признаков живых организмов одного вида, породы и получение комплекса параметров и коэффициентов, характеризующих членов изучаемой группы по одному или нескольким признакам. В биометрии такой массовый материал называется генеральной совокупностью, которая и составляет цель изучения. Совокупностями являются породы, стада животных, линии, семейства, дочери определенного производителя, группа овец, на которых проводился опыт, количество эритроцитов в каком-то объеме крови животного и т.д. Совокупность состоит из членов или единиц. Для стада коров единицей будет являться каждая корова.

Число единиц, входящих в совокупность, называется **объектом** совокупности и обозначается буквой ***n***. Единица совокупности может характеризоваться определенными признаками. Например, коровы характеризуются удоями за некоторый отрезок времени, жирномолочностью, белковомолочностью, живой массой, промерами, мастью, рогатостью, комолостью, числом эритроцитов или процентом гемоглобина в крови и т.д. Сумма отдельных измерений или наблюдений также является совокупностью.

Величину изучаемого признака для какой-то единицы совокупности называют **вариантой** и обозначают $V_1; V_2; V_3 \dots V_n$, а в общем виде V_i, где i – порядковый номер варианты.

Например, при изучении удоя (ц) за первую лактацию у трех коров получены следующие данные 50, 45, 40. Эти величины и будут вариантами, т.е. $V_1=50$, $V_2=45$, $V_3=40$.

Различия между отдельными вариантами называется изменчивостью или вариацией. В таких случаях говорят, что удой варьирует или «признак варьирует».

Изучение генеральной совокупности при большой численности животных – сложное и дорогостоящее мероприятие, поэтому применяют так называемый метод выборочной совокупности, который позволяет оценить генеральную совокупность путём отбора меньшей численности обследованных животных. Но при этом выборочная совокупность должна правильно отражать качества и особенности животных, составляющих генеральную совокупность. Такое условие обеспечивается отбором части животных из генеральной совокупности по принципу случайной выборки.

Следовательно, выборка всегда меньше генеральной совокупности и составляет её часть. Выборочная совокупность должна правильно характеризовать генеральную совокупность и быть репрезентативной (представительной).

Метод случайного отбора членов выборки принято называть рендоминизацией. Она даёт равную возможность любому члену генеральной совокупности войти в состав рендоминизированной выборки. Например, если требуется дать характеристику стада, включающего 500 коров, то делают случайную выборку, численность членов которой (объём) определяют по специальным формулам.

Для ряда признаков изучение их варьирования у всех особей генеральной совокупности невозможно ещё и потому, что это может привести к уничтожению такой совокупности. В этих случаях выборочная совокупность (или выборочная проба) – единственный способ, позволяющий изучать тот или иной признак. Например, средние пробы молока, крови, зерна – для оценки всхожести и др. Объектом биометрии служит варьирующий признак, полученный на достаточной по численности группе особей однородной по основным признакам.

Выборка может быть малой, когда в неё отобрано меньше 30 членов (вариантов $n<30$), или большой, если она составляет $n>30$ вариантов. Техника обработки малых и больших выборок различна. Число особей в выборке обозначается буквой *«n»*, а в генеральной совокупности *«N»*. Величина признака у отдельной особи называется вариантной и обозначается буквой *«V»*. Вариант – это численное выражение признака, полученное при его измерении.

Основные статистические параметры и коэффициенты, используемые в зоотехнии

Основными источниками статистической информации в практической работе с животными служат данные: первичного зоотехнического и ветеринарного учёта; научно-производственного опыта; специального научного эксперимента в условиях лаборатории; получаемые в экспедициях при обследовании больших массивов домашних или диких животных с точки зрения экологии, этологии или по другому поводу.

Полученные данные подвергают статистическому анализу и используют при планировании долгосрочных программ развития животноводческой отрасли. В редких случаях статистической обработке подвергают всю генеральную совокупность. Так, перепись населения, или какого-либо вида животных – примеры сплошного изучения генеральной совокупности, что делается очень редко, так как это требует больших материальных и трудовых затрат.

Отобранные в выборочную совокупность объекты далее подвергают статистической обработке, позволяющей получать различные статистические коэффициенты. Выборку для этого оформляют в виде вариационного ряда.

Вариационный ряд – это ряд вариантов, расположенных в определённой последовательности. Он показывает, как изменяется признак от минимальной до максимальной величины, какая частота вариант в каждом классе.

Класс, в котором встречается наибольшее число вариант, называется **модальным.** Избранную форму организации членов в выборке определяют тем, какие статистические параметры или коэффициенты должны быть вычислены.

После того, как члены в выборке организованы в виде вариационного ряда приступают к обработке материала для получения необходимых статистических показателей. Биометрическим методом получают:

- среднюю арифметическую (X);
- среднее квадратическое отклонение (±(σ);
- коэффициент вариации $(Cv\%)$ и корреляции ($\pm r$);
- ошибку средней арифметической ($\pm m$);
- критерий достоверности средней арифметической (0;

критерий достоверности разности между двумя средними арифметическими показателями *(td);*

- необходимый объем выборки для получения достоверных результатов при проведении научных исследований *(n)*.

Занятие 2. Построение вариационного ряда.

Цель занятия. Освоение метода построения вариационного ряда.

Методические указания. В группе особей, взятых для изучения, различные вариации признаков встречаются неодинаковое число раз. Частота проявления определенных значений признака в совокупности называется распределением. Распределение признака можно изобразить в виде вариационного ряда, вариационной кривой, вариационной гистограммой.

При анализе совокупности часто нужно полученные данные сгруппировать и представить в виде таблицы или ряде. При характеристике количественных признаков и большом числе вариант производят группировку данных и их разноску по классам, т.е. строят вариационный ряд.

Вариационный ряд – это есть распределение животных по классам или вариациям состоит из 2 граф (1-я – классы, 2-я частоты, т.е. количество животных в каждом классе.

Задача. Построить вариационный ряд по количеству лейкоцитов в крови у сухостойных коров черно-пестрой породы (тыс. в 1 мм^3 крови): 5,2 6,8 5,8 4,0 6,7 7,0 6,0 7,5 5,7 8,7 5,8 6,4 5,8 6,3 4,8 9,5 5,5 6,2 6,8 5,3 6,8 7,9 6,9 8,5 5,3 3,2 6,1 6,3 6,2 8,1 10,1 5,6 6,9 7,6 6,7 5,2 6,6 8,4 6,1 11,1 6,7 7,3 7,4 8,3 6,0 6,9 5,4 6,0 5,1 6,7 5,6 9Д8,4 10,3 10,7 7,8 10,2 5,1 8,6 6,4 6,5 5,0 6,6 8,5

n=64

1. Для построения вариационного ряда находим минимальное и максимальное значение вариант.

V_{max}=11.1 тыс. в 1 мм^3

V_{min}=3.2 тыс. в 1 мм^3

2. Находим величину классового промежутка *К*, которая определяется следующим образом:

$$K = \frac{V\max - V\min}{числоклассов}$$

Число классов берется в зависимости от объема выборки.

Число вариант	25 – 40	40 – 60	60 – 100	100 – 200
Число классов	5 – 6	6 – 8	7 – 10	8 – 12

В данном примере возьмем 8 классов.

Для упрощения оформления границ классов, классовый промежуток округляют до удобной величины. С непрерывными числами вариант, округление проводят до десятичного знака (7,9/8=0,9875≈1,0), с прерывным – до целого числа (25:8=3,125≈3,0).

3. Определяем границы классов (табл.2). Граница бывает нижней и верхней. За нижнюю границу первого класса принимается минимальная варианта V_{min}=3,2тыс. в 1 мм^3. Нижняя граница второго класса и последующих определяется путем прибавления к V_{min} классового промежутка *(К)* с нарастающим итогом.

4. Верхняя граница первого класса рассчитывается следующим образом: к V_{min} прибавляют классовый промежуток за минусом единицы – если числовые значения вариант в выборке целые, т.е. прерывные и 0,1 – если числа дробные, непрерывные. В нашем примере числа дробные, значит будем вычитать 0,1 и получим 3,2+1-0,1 =4,1.

Верхняя граница второго класса и последующих верхних определяется путем прибавления классового промежутка с нарастающим итогом.

Таблица 2 – Вариационный ряд

№	Границы классов	Частоты (число коров) *f*
1	3.2 – 4.1	2
2	4.2 – 5.1	4
3	5.2 – 6.1	17
4	6.2 – 7.1	20
5	7.2 – 8.1	7
6	8.2 – 9.1	7
7	9.2 – 10.1	3
8	10.2 – 11.1	4

5. Разносим по классам показатели крови у коров в виде точек методом конверта ⊠ по классам.
6. Вычерчиваем вариационную кривую – графическое изображение вариационного ряда. Графическое изображение вариационного ряда дает наглядное представление о характере распределения признака в изучаемой совокупности. Вариационный ряд можно представить в виде гистограммы. Для этого на горизонтальной линии (оси абсцисс) наносятся классы, а на вертикальной (ось ординат) – частоты. Основанием каждого столбика является соответствующее значение класса, а высота – число особей в нем. Если соединить прямыми линиями середины всех столбиков, получится вариационная кривая или полигон распределения. Полигон распределения должен своими ветвями касаться на оси абсцисс середины соседних классов.

Закономерности вариационного ряда

1. Большинство вариант располагается в середине вариационного ряда. Класс, в котором наибольшее количество признаков, называется **модальным.**
2. Распределение вариант в обе стороны от модального класса более- менее равномерно и симметрично.
3. Частота вариант постепенно убывает к краям вариационного ряда.

Вышеназванные закономерности характерны для большинства вариационных рядов. Для сравнения на одном графике нескольких распределений удобно пользоваться не гистограммой, а полигоном распределения. Если взято малое количество животных, то в некоторых классах вариационного ряда варианты могут отсутствовать, тогда вариационная кривая бывает разорванной. При малом числе особей и растянутых вариационных рядах часто наблюдается двухвершинность или многовершинность. Если выборка взята достаточно большой, то двухвершинность указывает на смешение двух различных совокупностей (двух пород, линий и т.д.) или выращивание и содержание животных в разных условиях и т.д.

Встречаются асимметричные вариационные кривые со смещением вершин влево или вправо, т.е. положительная или отрицательная асимметрия. Это объясняется неоднородностью условий развития животных данной совокупности, наличием в изучаемой группе большого количества особей с лучшими и худшими наследственными задатками, отбором. Если в средних классах вариационного ряда наблюдается преобладание вариант, то получается островершинная кривая, называемая положительным эксцессом.

Вопросы для самоконтроля.

1. Что такое генеральная совокупность и выборка?
2. Что такое вариационный ряд и как его построить?
3. Что такое вариационная статистика?
4. Какие признаки называются количественными и качественными?
5. Какими могут быть выборки по количественному составу?
6. Что такое вариант и как обозначается?
7. Как определить величину классового промежутка, назовите формулу.
8. Как определить границы классов?
9. Закономерности вариационного ряда?
10. Что такое полигон распределения вариантов признака?

Задания для самостоятельной работы:

1. Построить вариационный ряд и вычертить вариационную кривую:

По длине туловища хряков крупной белой породы, см:

152	164	173	166	172	183	170	171	170	167	170	171	173	176	163
183	181	160	169	170	163	164	170	176	169	163	180	174	167	160

2. Построить вариационный ряд и вычертить вариационную кривую:

По содержанию белка в мясе, %:22,0 17,0 25,6 20,7 17,3 21,1 20,2 23,2 20,4 22,8 20,2 23,4 19,6 20,6 21,4 21,5 21,0 24,1 22,4 22,1 21,5 20,5 22,6 21,3 22,4 17,4 20,3 22,5 20,6 21,4 20,5 21,8 22,1 19,8 17,5 18,6 19,2 18,7 20,1 22,3 21,5

Занятие 3. Вычисление средней арифметической в многочисленных выборках

Цель занятия. Освоить методы вычисления средней арифметической в многочисленных выборках.

Методические указания. Для больших выборок ($n>30$) применяют непрямой способ вычисления средней арифметической и других статистических показателей – способ произведений.

Задача. Вычислить среднюю арифметическую методом произведений (отклонения от условной средней) по количеству лейкоцитов в крови коров черно-пестрой породы.

Для этого строится вариационный ряд. Возьмем вариационный ряд, построенный в предыдущем занятии и дополним его еще 3 графами: W – середина класса, a – отклонение от условной средней и $f \times a$ – произведение частоты на отклонение.

Таблица 3 – Вычисление средней арифметической количества лейкоцитов в крови коров черно-пестрой породы

№	Границы классов	Частоты (число коров) f	Середина класса *(W)*	Отклонение a	$f \times a$
1	3.2 – 4.1	2	3.65	-3	-6
2	4.2 – 5.1	4	4.65	-2	-8
3	5.2 – 6.1	17	5.65	-1	-17
А	6.2 – 7.1	20	6.65	0	0
5	7.2 – 8.1	7	7.65	+1	7
6	8.2 – 9.1	7	8.65	+2	14
7	9.2 – 10.1	3	9.65	+3	9
8	10.2 – 11.1	4	10.65	+4	16
		n=64			$\Sigma f \times a$

1. Вычисляем середину каждого класса *W,* которая равна полусумме значений нижней и верхней границы.

$$W = \frac{3.2 + 4.1}{2} = 3.65$$

2. Выбираем условную среднюю *А*. В качестве таковой обычно берут значение середины того класса, в который входит наибольшее число вариант (6,2+7,1):2=6,65. Этот класс называется модальным классом и в данном примере *А=6,65* кг. Обводим жирной чертой четвертый класс вариационного ряда и ставим букву *«А»* в его начале.

3. Вычисляем отклонения (ранжирование) середины каждого класса от условной средней по формуле:

$$a = \frac{W - A}{K}$$

где *W* – середина класса, *А* – у словная средняя, *К* – классовый промежуток. Отклонения рассчитываются от класса, середина которого равна 6,65.
Записываем в графу отклонения.

$$a_{4кл} = \frac{6,65 - 6,65}{1} = 0$$

$$a_{1кл} = \frac{3,65 - 6,65}{1} = -3$$

$$a_{2кл} = \frac{4,65 - 6,65}{1} = -2$$

Так продолжаем по всем классам.
Середины класса, которые идут *от«А»* в сторону увеличения, ранжируют по порядку со знаком «плюс» (+1,+2,+3 и т.д.), а в сторону уменьшения – со знаком «минус» (-1,-2,-3 и т.д.).

4. Записав отклонения каждого класса с их знаками в третью графу таблицы, умножаем отклонения каждого класса на соответствующую частоту и вычисляем

произведение *«f×a»,* записываем в четвертую графу таблицы с последующим их суммированием $\sum f \times a = 16$.

5. Вычисляем поправку к условной средней по формуле:

$$b = \frac{\sum f \times a}{n} = \frac{16}{64} = 0.25$$

Когда поправка со знаком «+», её прибавляют к условной средней *«A»,* а когда со знаком «минус» - вычитают.

6. Вычисляем среднюю арифметическую $\overline{X}$ по формуле:

$$\overline{X} = A + K \times b$$

$$\overline{X} = 6.65 + 1 \times 0.25 = 6.90 \text{ тыс. в 1 мм}^3$$

Занятие 4. Вычисление среднего квадратичного отклонения σ и коэффициента вариации *Cv%* в многочисленных выборках. Нормированное отклонение *t* .

Цель занятия. Освоение методов вычисления показателей разнообразия признаков и практическое использование их в селекции.

Методические указания. Установление степени разнообразия признака в популяциях имеет важное значение в генетическом анализе популяций и в селекции. Именно величиной изменчивости определяется возможность улучшения путем отбора лучших животных в племенных стадах.

В зависимости от величины изменчивости все хозяйственно полезные признаки животных, по которым ведется селекция, подразделяют на признаки с низкой изменчивостью (коэффициент изменчивости находится в пределах 1 – 15 %), средней (16 – 25 %) и высокой изменчивостью (26 % и более). Для селекционера важно знать не только средние показатели по стаду, но и изменчивость признака. Чем выше изменчивость признака, тем эффективнее вести селекцию и наоборот.

При высокой изменчивости какого-либо признака лучшие и худшие показатели будут существенно отличаться от средней арифметической, что даст возможность постоянного повышения среднего уровня признака по стаду за счет отбора для воспроизводства лучших особей. В то же время возможность селекции на улучшение признака, характеризующегося низкой изменчивостью, практически исключается. Это связано с тем, что показатели селекционного признака будут очень близкими к средней, и поэтому отобрать лучших особей из стада весьма сложно.

При изучении изменчивости (вариабельности) признака особей данной совокупности применяют следующие параметры: лимит $(lim = V_{max} - V_{min})$, среднее квадратическое (стандартное) отклонение (σ), 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. Однако, эти показатели не отражают различий внутри выборки и животные с такими показателями могут быть не характерны для данного стада.

Например, при одинаковой средней величине животных двух групп живой массой $\overline{X}_1$=450 кг, $\overline{X}_2$ =450 кг лимиты составили в первой группе 400 – 500, а во второй – 430 – 580 кг. Размах колебаний в первой группе был 100 кг, во второй – 150 кг.

Таким образом, при одной и той же средней величине, группы неоднородны. Следовательно, более точным показателем изменчивости является среднее квадратическое отклонение, которое учитывает отклонение каждой варианты от средней арифметической.

Среднее квадратическое отклонение или сигма (σ) величина именная она выражается в тех же единицах, что и средняя арифметическая (в кг, см, %, тыс. в 1 мм^3 и т.д.)

Наиболее часто употребляемыми в практической селекции показателями изменчивости (вариабильности) признака является среднее квадратическое (стандартное) отклонение (σ) и коэффициент вариации или изменчивости (*Cv%*).

Среднее квадратическое отклонение или сигма (σ) 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 животных по изучаемому признаку. Чем больше величина сигмы, тем выше изменчивость, и наоборот.

Вся изменчивость признака укладывается от средней арифметической в пределах $\pm 3\sigma$ (правило плюс-минус трех сигм). Поэтому средняя арифметическая, уменьшенная или увеличенная на 3σ, дает крайние варианты признака.

Так, при нормальном распределении вариант генеральной совокупности, в пределы $\overline{X} \pm 3\sigma$ входит 99,7 % особей. Если какая-либо варианта отклоняется от $\overline{X}$ больше, чем на $\pm 3\sigma$, то она не может принадлежать данной генеральной совокупности. Около 95 % особей (вариант) входит в пределы $\overline{X} \pm 2\sigma$ и приблизительно 68 % особей – в пределы выборки $\overline{X} \pm 1\sigma$.

Например, если живая масса бычков выборки составляет $\overline{X}$ =450 кг, $\sigma=\pm 2$ кг, используя свойство $\pm 3\sigma$ можно определить максимальную и минимальную варианту в генеральной совокупности (*lim*=450 кг±3×2 кг=444 – 456 кг) и в выборке (*lim*=450 кг±1×2 кг =448 – 452 кг).

Наличие знаков «+» и «-» - указывает на то, что величина этого параметра характеризует изменчивость признака особей в сторону уменьшения, так и в сторону их увеличения.

Среднее квадратическое отклонение – это абсолютный показатель изменчивости. При вычислении сигмы, ее определяют с точностью на один десятичный знак больше, чем точность, которую применяют в отношении средней арифметической для того же ряда.

Вычисление среднего квадратического отклонения в многочисленных выборках *(n>30)*

При вычислении среднего квадратического отклонения будем использовать метод произведений.

Рассчитаем сигму на примере задачи по количеству лейкоцитов в крови коров черно-пестрой породы по ранее приведенным данным. Д ля этого дополним вариационный ряд (таблица 2) еще одной графой *($f\times a\times a=f\times a^2$).*

Таблица 2 – Вычисление среднего квадратического отклонения по количеству лейкоцитов в крови коров черно-пестрой породы

№	Границы классов	Частоты (число коров) f	Середина класса (W)	Отклонение a	$f\times a$	$f\times a^2$
1	3,2 – 4,1	2	3.65	-3	-6	18
2	4,2 – 5,1	4	4.65	-2	-8	16
3	5,2 – 6,1	17	5.65	-1	-17	17
А4	6,2 – 7,1	20	6.65	0	0	0
5	7,2 – 8,1	7	7.65	+1	7	7
6	8,2 – 9,1	7	8.65	+2	14	28
7	9,2 – 10,1	3	9.65	+3	9	27
8	10,2 – 11,1	4	10.65	+4	16	64
		$n=64$			$\Sigma f\times a=15$	$\Sigma f\times a^2=177$

1. Находим сумму $f\times a^2$, которая равняется 177.
2. Сигму вычисляем по формуле:

$$\sigma = \pm K\sqrt{\frac{\sum f\times a^2}{n} - b^2}$$

$$b = \frac{\sum fa}{n}, \text{ тогда}$$

$$\sigma = \pm K\sqrt{\frac{\sum fa^2}{n} - \left(\frac{\sum fa}{n}\right)^2}$$

где: K – величина классового промежутка; f - частота; a - отклонение от условно среднего класса, n - количество животных, b - поправка к условной средней.

Подставив вычисленные величины в формулу, получим:

$$\sigma = \pm K\sqrt{\frac{\sum fa^2}{n} - \left(\frac{\sum fa}{n}\right)^2} = \pm 1\sqrt{\frac{177}{64} - \left(\frac{16}{64}\right)^2} = \pm 1\sqrt{2.70} = \pm 1\times 1.64 = \pm 1.64$$

Сигма имеет два знака («+» и «-»), так как варианты могут отклоняться от средней арифметической как в положительную, так и отрицательную сторону. Крайние значения (лимиты) в генеральной совокупности будут находиться в пределах ±3σ, в данном примере ($\overline{X}$ =6,90+3×1,64=11,82 тыс.мм^3) и ($\overline{X}$ =6,90-3×1,64=1,98 тыс. мм^3), а в выборочной совокупности *±1σ (lim 6,90* кг+1×1,64=5,26 - 8,54 тыс. мм^3).

Вычисление коэффициента вариации или изменчивости (*Cv*).

С помощью среднего квадратического отклонения можно сравнить изменчивость двух и более групп животных в отношении одинаковых признаков. Однако им нельзя воспользоваться для сравнения изменчивости двух и более признаков, выраженных в разных единицах (в кг, см, тыс. мм^3 и т.д.), например, молочности, жирномолочности, живой массы, количества лейкоцитов в крови и т.д.

Поэтому 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, а именно – в процентах.

Для вычисления коэффициента изменчивости определенного признака, величину его сигмы делят на среднюю арифметическую, а полученный результат умножают на 100 %. Формула коэффициента изменчивости следующая:

$$Cv = \frac{\sigma}{\overline{X}} \times 100$$

Например, для коров первой лактации получены следующие данные:
-жирномолочность $\overline{X}$ = 3,8 % и σ=0,17 %
-удой $\overline{X}$ = 4240 кг и σ=748 кг.
-высота в холке $\overline{X}$ = 160 см и σ=10 см.

Необходимо сравнить разнообразие различных признаков в группе по этим показателям.
Как видно сравнивать величины 0,17 % и 748 кг невозможно.
Для сравненья используем коэффициент вариации, подставляя данные в формулу.

$$Cv = \frac{0.17}{3.8} \times 100 = 4.47\%$$

$$Cv = \frac{748}{4240} \times 100 = 17.64\%$$

$$Cv = \frac{10}{160} \times 100 = 6.2\%$$

При сравнении коэффициентов изменчивости видно, что наибольшее разнообразие признака наблюдается по удою молока (17,64 %), наименьшее – по высоте в холке коров (6,2 %).

Коэффициент вариации имеет три уровня: 1-й уровень *Cv* до 15 % - это невысокая изменчивость, 2-ой уровень *Cv* от 16 до 25 % - изменчивость признака средняя; 3-ий уровень *Cv* с 26 % и более – очень большая изменчивость, группе животные разнородные.

В нашем примере коэффициент вариации составил:

$Cv = \frac{1,64}{6,90} \times 100 = 23,7\%$ это указывает на среднюю изменчивость количества лейкоцитов в крови коров.

При сравнении двух групп животных по изменчивости одного признака следует пользоваться средними квадратическими отклонениями (σ), а не коэффициентами вариации (*Cv*), зависящими от величины средней арифметической ($\overline{X}$). Поэтому даже при одинаковой величине сигмы коэффициенты вариации могут быть различны, если неодинаковые средние арифметические, что может привести к неправильным выводам.

Например, результаты биометрической обработки двух групп животных показывают, что живая масса 1-ой группы $\overline{X}_1$=l4,0 кг, (σ=±2,4 кг, а 2-ой группы $\overline{X}_2$ =18,0 кг, *σ=±2,4* кг. Наблюдаемая разница в 4,0 кг (18,0-14,0=4,0 кг) показывает межгрупповое различие средних арифметических 2-х выборок. Однако, внутригрупповая изменчивость живой массы животных как 1-ой так и 2-ой группы находится на одном уровне (σ=±2,4 кг). Тогда как, рассматривая вариабельность признака через призму коэффициента изменчивости, находим что в 1-ой группе он будет составлять 17,0 % (средняя степень изменчивости), а во 2-ой группе – 13,3 % (низкая степень изменчивости).

Генетическая и селекционная науки выявили разную степень изменчивости основных селекционных признаков. За последние годы установлены коэффициенты изменчивости для одних и тех же признаков у разных пород. Использование коэффициента изменчивости в таком многообразном плане значительно расширяет информацию для решения практических задач и углубленного теоретического анализа особенностей популяции животных.

Особенности коэффициента изменчивости.

1. Величина коэффициента изменчивости не должна рассматриваться отдельно от средней арифметической и сигмы. Например, при близких значениях коэффициента изменчивости (Cv_1;=16,0 % и Cv_2=16,0 %), полученных на 2-х выборках, абсолютные величины $\overline{X}_1$ =15,0 кг и $\overline{X}_2$=18,0 кг; σ_1=±*2,40* кг и σ_2=2,88 кг) для этих совокупностей могут быть совершенно на разных уровнях. Следовательно, одинаковая или близкая величина *Cv* двух выборок еще не означает, что они качественно близки.
2. Одинаковые величины *Cv* двух выборок могут быть результатом разных причин, а именно: *Cv* может увеличиваться или за счет повышенного числителя, т.е. более высокой величины *σ* одной выборки, или за счет уменьшенного знаменателя, т.е. $\overline{X}$. Эти особенности необходимо учитывать при анализе материала, чтобы не впасть в ошибку и не сделать неправильны выводы, беря величину *Cv* вне связи с величиной $\overline{X}$ и *σ.*
3. Коэффициент изменчивости целесообразно использовать при изучении показателей, характеризующих возрастные особенности животных. Использование *Cv* совместно с $\overline{X}$ и *σ* дает ясное представление о динамике и закономерностях онтогенеза по тому или иному признаку.
4. Коэффициент изменчивости имеет большое значение при планировании объема опыта. Так как правильно установленный объем опыта позволяет получать достоверные биометрические параметры.

Нормированное отклонение *t*. Используется для изучения изменчивости при нормальном распределении. Кроме характеристики вариационного ряда по величине сигмы, бывает необходимость оценки отдельных вариант по отношению их к средней арифметической величине совокупности.

Показатель нормированного отклонения определяется по разности между вариантой (V) и средней арифметической $\overline{X}$, отнесенной к величине среднего квадратического отклонения *(σ)*.

$$t = \frac{V - \overline{X}}{\sigma}$$

Каждая варианта характеризуется значением *t*. Если показатель нормированного отклонения какой-либо варианты +1, значит, эта варианта больше $\overline{X}$ на одну сигму. Если другой вариант равен -2, то это означает, что он меньше $\overline{X}$ на две сигмы.

Показатель нормированного отклонения удобен как для оценки отдельных вариант, так и для характеристики сравниваемых групп.

С помощью таблицы 4, зная с $\overline{X}$ и σ, можно построить теоретическую кривую распределения и определить долю особей с определенной величиной признака.

Таблица 4 – Значение нормального интеграла вероятностей

$t = \frac{V - \overline{X}}{\sigma}$	*P*	$t = \frac{V - \overline{X}}{\sigma}$	*P*	$t = \frac{V - \overline{X}}{\sigma}$	*P*	$t = \frac{V - \overline{X}}{\sigma}$	*P*
0,10	0,0398	1,0	0,3413	1,9	*0,4713*	2,8	0,4974
0,20	0,0793	1,1	0,3643	2,0	*0,4773*	2,9	0,4981
0,30	0,1179	1,2	0,3849	2,1	*0,4821*	3,0	0,4987
0,40	0,1554	1,3	0,4032	2,2	*0,4861*	3,5	0,4998
0,50	0,1915	1,4	0,41,92	2,3	*0,4893*		
0,60	0,2258	1,5	0,4332	2,4	*0,4918*		
0,70	0,2580	1,6	0,4452	2,5	*0,4938*		
0,80	0,2131	1,7	0,4554	2,6	*0,4953*		
0,90	0,3159	1,8	0,4641	2,7	*0,4965*		

Если известны $\overline{X}$ и σ, то используя таблицу ***4,*** можно определить долю (процент) коров с количеством лейкоцитов, например, свыше 8 тысяч в мм^3.

Определяем $t = \frac{V - \overline{X}}{\sigma} = \frac{8 - 6{,}90}{1{,}64}$ =0,67≈0,7

По таблице находим, что в пределах от $\overline{X}$ до $\overline{X}$ **+**0,7 содержится 0,2580 или 25,80 % вариант. Поэтому коров с количеством лейкоцитов свыше 8 тысяч в мм^3 будет 50 % - 25,80 **%**=24,2 %.

Например, при отборе сравниваются две разновозрастные коровы стада. От одной коровы за 305 дней первой лактации получено 2500 кг молока (V_1=2500 кг), от второй за такой же период пятой лактации получено 4500 кг (V_2=4500 кг). Простое сравнение их удоев для выбора лучшей коровы привело бы к ошибочному выбору. При сравнении двух коров следует учитывать величину удоев в связи с их возрастом и вычислить показатель нормированного отклонения. На ферме средний удой коров по первой лактации составляет 2000 кг ($\overline{X}_1$ =2000 кг), а удой коров по пятой лактации - $\overline{X}_2$ =4000 кг. Для коров первотелок сигма составила σ_1=220 кг, а коров старшего возраста σ_2=270 кг.

Нормированное отклонение для сравниваемых коров будет составлять:

$$t = \frac{V - \overline{X}}{\sigma} = \frac{2500 - 2000}{220} = +2{,}3$$

$$t = \frac{4500 - 4000}{270} = +1,8$$

Полученная величина *t* для первой коровы свидетельствует о значительном отклонении от ее средней величины удоя в группе. Однако, к пятой лактации эта корова раздоиться и будет более молочной, чем вторая корова.

Занятие 5. Вычисление средней арифметической ($\overline{X}$) , среднего квадратического отклонения (σ) и коэффициента вариации *(Cv%)* в малочисленных выборках *(n<30)*

Цель занятия. Освоение методов вычисления показателей разнообразия признаков и практическое применение их в селекции.

Методические указания. Средняя арифметическая является основным показателем, характеризующим совокупность по величине изучаемого признака. Установление степени разнообразия признака в популяциях имеет определенное значение в селекции. Таким показателем разнообразия признака является среднее квадратическое отклонение *σ,* которое учитывает отклонение каждой варианты от средней арифметической. При изучении разнообразия признаков, выраженных в различных единицах измерения, используют другой показатель – коэффициент вариации *(Cv%).*

Известны два метода вычисления $\overline{X}$ и *σ* в малых выборках.

1-й метод – вычисление *σ* при малой выборке с использованием условной средней (табл. 5). Применяется он в основном тогда, когда цифры дробные и более чем трехзначные.

Задание. Вычислить $\overline{X}$ *, σ* и *Cv%* по количеству жира в молоке:

4,4 5,1 4,8 5,2 4,6 3,6 3,8 4,5 4,7 3,8

При статистической обработке малых выборок рекомендуют работать с двухзначными числами. Это можно сделать с любыми числами, используя одно из свойств средней арифметической (то есть, если число четырехзначное, то его делят на 100, а если однозначное, то умножают на 10 и т.д.).

Таблица 5 – Вычисление *σ* и использованием условной средней *A*

№ п/п	Вариант, *V*	Отклонение *D=V-A*	Квадрат отклонения D^2
1	*A 4,4*	0	0
2	5,1	+0,7	0,49
3	4,8	+0,4	0,16
4	5,2	+0,8	0,64
5	4,6	+0,2	0,04
6	3,6	-0,8	0,64
7	3,8	-0,6	0,36
8	4,5	+0,1	0,01
9	4,7	+0,3	0,09
10	3,8	-0,6	0,36
	n=10	*ΣD*=+0,5	ΣD^2=2,79

Статистическую обработку ведут следующим образом. За условную среднюю *(A)* можно брать любое число, но обычно для удобства берут целое число избегая крайние варианты *(V_{min}=3,6* и *Vmax=5,2).* В данном примере удобнее взять *V*=4,4 % (3,6+5,2=8,8:2=4,4).

Вычисляем отклонения вариант от условной средней по формуле: *D=V-A,* записываем в графу *D* и суммируем *ΣD=+0,5*. Далее отклонения возводят в квадрат и результат записывают в третью графу и суммируем ΣD^2=2,79

Средняя арифметическая рассчитывается по формуле:

$$\overline{X} = A + \frac{\sum D}{n} = 4.4 + \frac{0.5}{10} = 4.4 + 0.05 = 4.45\text{кг}$$

Можно и по формуле:

$$\overline{X} = \frac{V_1 + V_2 + V_3 + V_n}{n} = \frac{\sum V_n}{n}$$

Получим такую же $\overline{X}$ =4,45 кг.

Среднее квадратическое отклонение вычисляется по формуле:

$$\sigma = \pm\sqrt{\frac{\sum D^2 - b}{n-1}}$$

Где *n* – количество вариантов, *b* – поправка к условной средней, которая вычисляется по формуле:

$$b = \frac{\left(\sum D\right)^2}{n} = \frac{(+0.5)^2}{10} = 0.025$$

$$\sigma = \pm\sqrt{\frac{\sum D^2 - b}{n-1}} = \pm\sqrt{\frac{2.79 - 0.025}{10-1}} = \pm\sqrt{\frac{2.765}{9}} = \pm\sqrt{0.31} = \pm 0.55$$

Следовательно, коэффициент изменчивости будет равен:

$$Cv = \frac{\sigma}{\overline{X}} \times 100 = \frac{0{,}55}{4{,}45} \times 100 = 12{,}36\%$$

2-й метод – вычисление среднего квадратического отклонения в малых выборках с использованием средней арифметической, где *ΣD=0* (табл.6).

Вычислить $\overline{X}$, *σ* и *Cv%* по данным приведенным в 1-ом методе.

Таблица 6 – Вычисление σ с использованием средней арифметической $\overline{X}$

№ п/п	Вариант, *V*	Отклонение $D = V - \overline{X}$	Квадрат отклонения D^2
1	4,4	-0,05	0,0025
2	5,1	+0,65	0,4225
3	4,8	+0,35	0,1225
4	5,2	+0,75	0,5625
5	4,6	+0,15	0,0225
6	3,6	-0,85	0,7225
7	3,8	-0,65	0,4225
8	4,5	+0,05	0,0025
9	4,7	+0,25	0,0625
10	3,8	-0,65	0,4225
	n=10 ΣV_n=44,5	*ΣD*=0	ΣD^2=2,76

1. Среднюю арифметическую вычисляем по формуле:

$$\overline{X} = \frac{\sum V_n}{n} = \frac{44.5}{10} = 4.45\%$$

2. Отклонение вариант от средней арифметической вычисляем по

формуле:

$$D=V-\overline{X},$$

записываем разность в графу D и суммируем, где $\Sigma D=0$. В третью графу вносим показатели отклонения, возведенного в квадрат и суммируем:

$$\Sigma D^2=2{,}76$$

3. Среднее квадратическое отклонение вычисляем по формуле:

$$\sigma = \pm\sqrt{\frac{\sum D^2}{n-1}} = \pm\sqrt{\frac{2.76}{10-1}} = \pm\sqrt{0.31} = \pm 0.55\%$$

$$Cv = \frac{0{,}55}{4{,}45}\times 100 = 12{,}36\%$$

Показатель изменчивости жирности молока низкий при Cv=12,36 %

Занятие 6. Вычисление ошибки средней арифметической и оценки достоверности выборочных показателей

Цель занятия. Освоение методов вычисления показателей ошибок для различных выборочных параметров при малом и большом числе наблюдений в выборке.

Методические указания. При проведении экспериментальных работ в животноводстве, а также в практической селекционной работе основные параметры совокупности ($\overline{X}$, σ и $Cv\%$ и др.) вычисляют, как правило, не по генеральной совокупности, а по выборочной. Вследствие этого возникают ошибки, называемые ошибками выборочности (ошибки репрезентативности).

Поэтому статистические параметры для выборки могут несколько отличаться от тех их значений, которые были бы получены для генеральной совокупности. Следовательно, необходимо добиваться того, чтобы статистические ошибки были полностью устранены, а если их нельзя избежать, то следует свести к минимуму.

Статистические ошибки обусловлены самим выборочным методом, при котором из генеральной совокупности отбирается по принципу случайности часть ее членов (случайная выборка).

Таким образом, случайная выборка, составляя часть генеральной совокупности, должна достаточно правильно отражать свойства генеральной совокупности. Но как часть чего-либо не может полноценно отражать свойства целого, так и выборка не может полностью отражать генеральной совокупности, в результате чего и возникает статистическая ошибка. Поэтому все статистические параметры, вычисленные для выборочной совокупности, могут в той или иной мере не совпадать, а отличаться по своей величине от аналогичных параметров для всей генеральной совокупности. Чем меньше ошибка, тем лучше выборочные параметры ($\overline{X}$, σ и $Cv\%$ и т.д.) характеризуют генеральную совокупность.

В биометрии разработаны приемы вычисления величины ошибок. По величине ошибки и соотношению ее с той выборочной характеристикой, для которой она вычислена, можно судить о том, достаточно ли точно выборочные данные отражают параметры присущие генеральной совокупности. Если в обработку включены все члены генеральной совокупности, то никаких ошибок не возникает и вычислять их не нужно.

Для того чтобы вычисленные параметры выборки более точно соответствовали аналогичным параметрам генеральной совокупности, необходимо добиваться уменьшения ошибки. Следовательно, 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.

Так как степень изменчивости признака объективно существует у членов совокупности и изменять его по усмотрению экспериментатора нельзя, то влиять на величину ошибки можно путем определения необходимого объема выборки, численность которой обеспечит достаточное совпадение выборочных коэффициентов с коэффициентами генеральной совокупности

Увеличение объема выборки дает уменьшение статистических ошибок. Это особенно необходимо учитывать при изучении объектов, имеющих сильное варьирование признаков. Однако, увеличение объема выборки связано с увеличением затрат труда, времени и средств на получение и обработку цифрового материала. Следовательно, необходимо заранее определить такой объем выборки, который обеспечит достаточную точность опыта, т.е. достаточное совпадение выборочных статистических параметров с параметрами генеральной совокупности, и в то же время не потребует больших затрат для опыта и обработки материалов, но гарантирует получение достоверных статистических выборочных параметров.

В структуру формул статистических ошибок включают показатель изменчивости признака и объем выборки. В зоотехнической и ветеринарной литературе статистическую ошибку принято обозначать буквой *«m»* с подстрочным значком того параметра, для которого его вычисляют.

Формулы статистических ошибок для основных генетико-статистических параметров следующие:

- ошибка средней арифметической для выборки *n>30*

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

-ошибка средней арифметической для выборки *n<30*

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

-ошибка среднего квадратического отклонения

$$m_{\sigma} = \pm \frac{\sigma}{\sqrt{2n}}$$

-ошибка коэффициента вариации

$$m_{C_v} = \pm \frac{C_v}{\sqrt{2n}}$$

Ошибка средней арифметической есть неточность или разность между средней арифметической выборки и средней арифметической генеральной совокупности. Ошибка средней арифметической может быть выражена не только в именованных величинах, но и в относительных, т.е. в процентах. В этом случае ее называют показателем точности *(E)* и вычисляют путем определения процентного выражения ошибки от средней арифметической:

$$E = \pm \frac{m_x}{\overline{X}} \times 100\%$$

Чем меньше величина *E,* тем достовернее, надежнее полученная средняя арифметическая. Использование ошибки прежде всего необходимо для определения достоверности полученной выборочной средней арифметической. Это осуществляется путем вычисления так называемого критерия достоверности *(t).*

Для средней арифметической критерий достоверности определяется по формуле:

$$t_{\bar{x}} = \frac{\overline{X}}{m_x}$$

По величине *t* судят о достоверности данного статистического параметра, основываясь на связи этой величины с уровнем вероятности (*P*).

Так, при стандартном t_1=1,96, вероятность $P \geq 0{,}95$,

t_2=2,58 вероятность $P \geq 0{,}99$,

t_3=3,30 вероятность $P \geq 0{,}999$.

Эти данные показывают, какова вероятность того, что вычисленный параметр *t* (критерий достоверности средней арифметической) достоверно отражает уровень такого же параметра генеральной совокупности. Если, в конкретном примере *t*=1,96, а P≥0,95, то это значит, что из 100 вариантов выборки в 95 будет получено такое же значение параметра, какое получено в данной выборке, где *t* =1,96.

Величину $t_{0.95}$=1.96 называют первым порогом достоверности. Она дает возможность считать данные, полученные в выборке, достоверными, то есть правильно отражающими параметр генеральной совокупности. Этот порог считается минимальным для работ, имеющих поисковый характер, для научно-исследовательских опытов на производстве.

Второй порог достоверности принято брать на уровне *P*≥0,99, когда *t* =2,58. Этот показатель используют в том случае, когда требуется детализация различных явлений и закономерностей, например, для генетических исследований.

Третий порог принято брать на уровне *P*≥0,999, то есть при *t* =3,3.

В этом случае вероятность правильности выборочного параметра подтверждалось бы в 999 опытах из 1000. И только в одном случае параметры в выборках могли быть другими по величине. Этот порог достоверности принято использовать при изучении действия дозировок опасных препаратов.

Если в конкретном материале критерий достоверности *t* больше трех или четырех, то это значит, что достоверность вычисленных параметров высоковероятна.

В научной литературе иногда выражают показатель вероятности в величинах значимости *P,* которая отмечает уровень риска и ошибочности вывода. Следовательно, при *P*≥0,95 величина значимости *P*≤0,05, что соответствует значимости риска и ошибочности вывода. При *P*≥0,99, значимость меньше или равна ≤0,01, при P≥0,999, значимость меньше или равна ≤0,001.

Рассчитаем критерий достоверности по примеру из 64 коров (по ранее приведенным данным). Определено среднее количество лейкоцитов в крови коров $\overline{X}$ =6,90 тыс. в мм^3, а σ=±1,67 тыс. в мм^3.

Вычисляем ошибку средней арифметической в многочисленных выборках

$$m_x = \pm\frac{\sigma}{\sqrt{n}} = \pm\frac{1.67}{\sqrt{64}} = \pm\frac{1.67}{8} = \pm 0.21 \text{ тыс. мм}^3$$

Это означает, что средняя ошибка выборки составляет 0,21 тыс. в мм^3. Следовательно среднее количество лейкоцитов в крови коров в изучаемой выборке характеризуется:

$\overline{X}$ $\pm m$ = 6.90 ± 0.21 тыс. в мм^3

Критерий достоверности для средней арифметической количества лейкоцитов в крови коров будет следующий:

$$t_{\bar{x}} = \frac{\overline{X}}{m_x} = \frac{6.90}{0.21} = 32.8$$

При таком критерии достоверности ($t_{\bar{x}}$ =32.8)варианты генеральной совокупности средней арифметической будут находиться в пределах:

$$\tilde{X} = \overline{X} \pm t \times m$$

где $\tilde{X}$ - средняя арифметическая генеральная совокупность, $\overline{X}$ - средняя арифметическая выборки, *m* – ошибка средней арифметической выборки, *t* – критерий достоверности.

Эти границы возможного нахождения генерального параметра называются доверительными границами (или границами доверительного интервала). Верхняя граница $\tilde{X} = \overline{X} \pm t \times m$ =6,90+1,96×0,21=6,90+0,41=7,31, нижняя $\tilde{X} = \overline{X} \pm t \times m$ =6,90-l,96×0,19=6,90-0,41=6,49 тыс. мм^3).

Таким образом, в выборочной совокупности среднее количество лейкоцитов в крови коров составляете 6,90, а количество лейкоцитов в крови коров в генеральной совокупности находится в границах 6,49 – 7,31, что можно утверждать с вероятностью 0,95. Фактическая величина ($t_{\bar{x}}$ = 32.8) критерия достоверности для средней арифметической количества лейкоцитов в крови коров оказалась значительно выше нижней границы ($t_{0.95}$).

Следовательно, полученная средняя арифметическая имеет очень высокую достоверность.

Если вычисленное значение $t_{\bar{x}}$ будет меньше 1,96 (при самом низком уровне вероятности), то выборочный параметр (средняя арифметическая) недостоверен, т.е. он не может служить характеристикой генеральной совокупности, и в этом случае полученные в опыте данные не имеют ценности, так как выводы не могут быть распространены на генеральную совокупность, изучение которой служит основной целью опыта, построенного на выборочном методе.

Чаще всего недостоверность средней есть следствие недостаточного объема выборки, т.е. слишком мало число наблюдений. В таких случаях необходимо увеличить число животных в выборке и заново ставить опыт.

В зависимости от того, какой вопрос решается статистическим методом использованием выборочной совокупности, требование к порогу (уровню) достоверности выбирается различное.

Уровень достоверности		Характер изучаемого признака
Не<t=1,96	P=0,95	Биологические вопросы, научно-хозяйственные опыты, поисковые эксперименты
Не<t=1,96	P=0,99	Вопросы экономического и производственного характера, по которым будут даны рекомендации, детализация биологический явлений и закономерностей
Не<t=1,96	P=0,999 и выше	Изучение действия опасных для жизни препаратов и заключение о дозах безвредности

Для генетических и селекционных целей используют два первых порога достоверности. Высокий критерий достоверности (вероятности P=0,999) необходимо применять при работе с опасными веществами (дозы яда, облучения и т.д.).

Величина критерия достоверности тесно связана с величиной вероятности *(P)*. Понятие «значимость» является как бы обратным по смыслу понятию «вероятность». Так, если мы говорим, что полученная средняя арифметическая или разность между средними арифметическими двух выборок достоверна на уровне вероятности *P=0,99* (или 99 %), то это в то же время соответствует уровню значимости *P=0,01* (или 1 %). Уровень значимости показывает, что нахождение генерального параметра за пределами данных границ доверительного интервала может иметь место в результате случайности с вероятностью 0,01 (или 1 %), а с вероятностью 0,99.(или 99 %) утверждается, что он находится в принятых границах доверительного интервала. Использование уровня значимости удобно в том смысле, что он показывает – процент ошибочных случаев. Так, значимость 0,05 указывает на то, что в результате случайности ошибка в выводах будет наблюдаться в 5 % случаев.

Занятие 7. Оценка достоверности разности между средними арифметическими двух выборочных совокупностей.

Цель занятия. Освоение метода определения достоверности разности между средними арифметическими двух выборок.

Методические указания.

При сравнении средних арифметических двух генеральных совокупностей любая разность между ними будет достоверна.

Во многих исследованиях (в зоотехнии, в ветеринарии и т.д.) возникает необходимость сравнить средние арифметические двух групп животных (например, породы, линии, семейства, средний процент жира в молоке коров опытной и контрольной групп). Средние двух сравниваемых групп в некоторой степени отличаются друг от друга. Поэтому необходимо установить, достоверна ли разница между средними двух групп. Недостаточно, например, знать, что 20 дочерей какого-то производителя превосходят по удою своих матерей. Следует, кроме того, вычислить критерий достоверности разности, чтобы с определенной вероятностью судить о том, что следующие 100, 200 и т.д. дочерей этого производителя также будут превосходить по молочности своих матерей в аналогичных условиях.

При решении задач определяют разность *(d)* между двумя средними:

$$d = \overline{X}_1 - \overline{X}_2$$

А также среднюю ошибку разности:

$$m_d = \sqrt{m_1^2 + m_2^2}$$

Критерий достоверности разности между средними величинами групп определяют по формуле:

$$td = \frac{d}{m_d} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{m_1^2 + m_1^1}}$$

где $\overline{X}_1$ большая величина из двух средних, m_1 и m_2 – ошибки средней арифметической двух групп.

Например, нужно установить, различаются ли дочери быков по титру лизоцима в крови, если получены следующие показатели:

$n_1=30,\ \overline{X}_1 \pm m_1=1{,}612 \pm 0{,}02$

$n_2=28,\ \overline{X}_2 \pm m_2=1{,}538 \pm 0{,}03$

Вычисляем $td = \frac{1{,}612 - 1{,}538}{\sqrt{0{,}02^2 + 0{,}03^2}} = \frac{0{,}074}{\sqrt{0{,}0013}} = \frac{0{,}074}{0{,}036} = 2{,}06$

Достоверность разницы в 0,074 определяем по таблице Стьюдента (табл. 7) с учетом числа степеней свободы $\upsilon=n_1+n_2-2=30+28-2=56$.

Находим по таблице строку $\upsilon=56$. Из 3 стандартных значений t_{st} (нижняя строка), которые равны: при р=0,95 $t_{st}=1{,}96$; при р=0,99 $t_{st}=2{,}58$; при р=0,999 $t_{st}=3{,}29$. Сравниваем величину td с t_{st}.

Здесь возможно два вывода:

1. Если *td* равен или больше значения t_{st} для первого уровня вероятности ($td \geq t_{st}$), то разность между средними арифметическими двух групп статистически достоверна.

2. Если *td* меньше значения t_{st} для первого уровня вероятности, то разность между средними арифметических двух групп статистически недостоверна.

В нашем примере $td=2{,}06$ больше значения $t_{st}=1{,}96$ для первого уровня вероятности (0,95), но меньше $t_{st}=2{,}58$ для второго уровня вероятности (0,99). Поэтому можно сделать вывод: разность между потомством двух производителей по титру лизоцима в крови достоверна с вероятностью $p>0{,}95$. У дочерей первого быка более высокий титр лизоцима в крови.

Критерий достоверности зависит от числа наблюдений в выборке. Для устранения влияния числа наблюдений на величину критерия достоверности с учетом числа степеней свободы (υ) для трех уровней вероятности используют таблицу Стъюдента (табл. 7).

Таблица 7 – Стандартные значения критерия достоверности (по Стъюденту)

Число степеней	Уровень вероятности, Р			Число степеней	Уровень вероятности, Р		
	0,95	0,99	0,999		0,95	0,99	0,999

свободы, υ				свободы, υ			
1	12,71	63,66	637,00	21	2,08	2,83	2,82
2	4,30	9,93	31,60	22	2,07	2,82	3,79
3	3,18	5,84	12,94	23	2,07	2,81	3,77
4	2,78	4,60	8,61	24	2,06	2,80	3,75
5	2,57	4,03	6,86	25	2,06	2,76	3,73
6	2,45	3,71	5,96	26	2,06	2,78	3,71
7	2,37	3,50	5,41	27	2,05	2,77	3,69
8	2,31	3,36	5,04	28	2,05	2,76	3,67
9	2,26	3,25	4,78	29	2,05	2,76	3,66
10	2,23	3,17	4,59	30	2,04	2,75	3,65
11	2,20	3,11	4,44	35	2,03	2,72	3,59
12	2,18	3,06	4,32	40	2,02	2,70	3,55
13	2,16	3,01	4,22	45	2,01	2,69	3,52
14	2,15	2,98	4,14	50	2,01	2,68	3,50
15	2,13	2,95	4,07	60	2,00	2,66	3,46
16	2,12	2,92	4,02	70	1,99	2,65	3,43
17	2,11	2,90	3,97	80	1,99	2,64	3,42
18	2,10	2,88	3,92	90	1,98	2,63	3,40
19	2,09	2,86	3,88	100	1,98	2,62	3,37
20	2,09	2,85	3,85	120 и выше	1,96	2,56	3,29

Число степеней свободы – это число наблюдений, уменьшенное на число ограничений *(n-1, n_1+n_2-2* и т.д.). Таблица Стъюдента пригодна при определении критерия достоверности для средних арифметических, достоверности разности, коэффициентов корреляции.

Вопросы для самоконтроля:

1. Назовите основные свойства средней арифметической.
2. Методы вычисления средней арифметической.
3. Как определить величину условной средней и поправку к ней?
4. Для каких целей используется показатель сигмы, что он показывает и в каких единицах измеряется?
5. Что есть коэффициент вариации, в каких случаях он применяется?
6. Какие бывают уровни коэффициента вариации, их характеристика?
7. Что характеризует средняя арифметическая и как она определяется в многочисленных выборках?
8. Возможно ли определить максимальное и минимальное значение изучаемого признака, если известна величина средней арифметической и среднего квадратического отклонения.
9. Какие показатели характеризуют разнообразие признаков?
10. Как вычисляется среднее квадратическое отклонение в малых и больших выборках?
11. По какой формуле вычисляется $\overline{X}$, D и D^2 в малочисленных выборках?
12. Какова формула для определения σ и *Cv%* в малочисленных выборках?
13. Что такое средняя взвешенная? В каких случаях она применяется и как ее вычислить?
14. Что показывает нормированное отклонение и когда используется?
15. Что такое ошибка средней арифметической?
16. По какой формуле вычисляется критерий достоверности между средними

арифметическими двух выборок?
17. Что означает уровень вероятности или значимости?
18. С какой целью вычисляют критерий достоверности разности между средними величинами двух выборок?

Задания для самостоятельной работы:

1. Вычислить $\overline{X}$, *σ, Cv%, t* при *n>30* разными способами:

По суточному удою коров, кг:

24,923,8 30,2 28,3 29,9 19,9 12,8 17,0 12,5 21,5 24,8 16,7 24,5 26,3 25,4 16,5 25,0 23,4 22,5 22,9 22,7 22,8 19,1 19,6 27,3 25,6 23,7 14,8 20,6 25,7 23,0 28,4 26,3 24,9 25,6 24,7 16,8 19,6 20,5 22,4 22,6 22,8 23,4 23,1

2. Вычислить $\overline{X}$, *σ, Cv%, t* при *n>30* разными способами:

По среднесуточному приросту живой массы ремонтных свинок, г:

430 520 600 607 550 555 527 529 575 533 507 593 4621 521 480 476 442 509 560 516 549 447 484

3. Вычислить $\overline{X}_1$, $\overline{X}_2$, σ_1, σ_2, $Cv_1\%$, $Cv_2\%$, m_1, m_2, *td* и сделать вывод о достоверности разности между сравниваемыми группами.

По удою молока за лактацию у коров черно-пестрой породы, содержащихся на разных рационах кормления, кг

1 группа (ОР+добавка)

4002 4590 4300 5100 5009 4900 4592 5505 5000 5400 4506 4690 4505 5250 5350 4760 4000 4800

2 группа (ОР)

4590 4100 4100 5106 4700 4422 4305 5100 5100 4206 4580 4240 5000 5100 4260 3950 4400 4150

Занятие 8. Коррелятивная связь между признаками.

Цель занятия. Освоить методы вычисления корреляционной связи между признаками и использование этих показателей в селекционной работе и прогнозировании селекции.

Методические указания. Признаки и свойства животных находятся в определенной взаимосвязи. Например, существует связь между устойчивостью и восприимчивостью матерей и их дочерей к маститу, лейкозу и т.д., между уровнем кормления и продуктивностью, длиной туловища и живой массой, фагоцитарной активностью и резистентностью.

Такие связи называются корреляцией – взаимосвязь между признаками в организме или особями одного вида и пола по какому-либо признаку.

Особенность корреляции (связи) между признаками живых организмов состоит в том, что каждому значению одного признака соответствует не одно, а несколько значений другого. Так животные с одинаковой высотой могут иметь разную живую массу, но в среднем низкие животные имеют меньшую живую массу, чем более высокие животные.

Различают положительную (прямую) и отрицательную (обратную) корреляции. При положительной корреляции с увеличением одного признака увеличивается другой. Например, с увеличением живой массы коров-первотелок, возрастает и удой, чем выше процент жира в молоке, тем выше и процент белка в нем.

При отрицательной корреляции с увеличением одного признака второй уменьшается. Например, с увеличением удоя у коров снижается жирность молока, чем больше длину туловища у свиней, тем меньше толщина сала и т.д.

Корреляционная связь наблюдается не только между количественными признаками, но и между качественными. Например, известна связь между типом конституции романовских овец и их шубными качествами.

Связь между признаками позволяет осуществлять косвенную селекцию, когда отбор по одному из коррелирующих признаков будет приводить косвенно, на основе сопряженности, к отбору животных и по другому признаку. Например, отбирая коров с более высоким процентом жира в молоке, тем самым выделяют животных, имеющих в среднем более высокое содержание белка в молоке.

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. При положительной корреляции его величина может изменяться от 0 до +1, а при отрицательной от -1 до 0. Когда коэффициент корреляции равен 0, то изменение одного признака, происходит независимо от другого.

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».

Если коэффициент корреляции меньше 0,5 - связь слабая; 0,5 - 0,6 - средняя; 0.7 – 1.0 – высокая. Примеры положительной и отрицательной корреляции между различными признаками у разных видов животных приведены в таблице 8.

Таблица 8 – Корреляция между некоторыми признаками крупного рогатого скота, свиней, овец

Коррелирующие признаки	Коэффициент корреляции
Крупный рогатый скот: молочный	
Удой - % жира	от -0,001 до -0,376
Удой - % белка	от -0,10 до -0,36
Удой - количество молочного жира	0,88 – 0,98
Удой – живая масса	0,02 – 0,65
Удой – обхват вымени	0,43 – 0,72
Мясной	
Масса при рождении – прирост живой массы до отъема	0,46 – 0,50
Среднесуточный прирост – конечная масса	0,77 – 0,80
Масса телят при отъеме – молочность матерей	0,70 – 0,80
Свиньи:	
Средняя масса поросят при рождении и средняя масса поросят при отъеме	0,54 – 0,90
Количество поросят при отъеме – масса поросят при отъеме	0,82 – 0,90
Овцы:	
Густота шерсти – настриг шерсти	
Масса немытой шерсти – масса шерсти	
Длина шерсти – настриг шерсти	

Густота шерсти – длина шерсти	

Вычисление коэффициента корреляции (*r*) в многочисленных выборках (*n>30*)

Вычислить величину корреляции между количеством эритроцитов (млн. в 1 мм^3) и лейкоцитов (тыс. в 1 мм^3) у 80 коров черно-пестрой породы (табл. 9).

Таблица 9 – Количество эритроцитов и лейкоцитов в крови у коров (*n*=80) черно-пестрой породы.

№ п/п	Количество		№ п/п	Количество		№ п/п	Количество	
	эритроцитов, млн. в 1 мм^3	лейкоцитов, тыс. в 1 мм^3		эритроцитов, млн. в 1 мм^3	лейкоцитов, тыс. в 1 мм^3		эритроцитов, млн. в 1 мм^3	лейкоцитов, тыс. в 1 мм^3
1	6,3	8,7	28	7,6	8,0	55	5,7	7,4
2	6,2	7,4	29	6,9	7,7	56	7,6	8,4
3	7,7	7,8	30	5,7	7,1	57	5,5	6,4
4	5,7	12,8	31	7,6	7,7	58	6,7	7,3
5	6,0	6,2	32	6,6	12,3	59	5,6	7,3
6	6,6	8,1	33	5,8	6,8	60	6,4	8,8
7	6,4	6,8	34	6,8	8,7	61	6,2	7,4
8	5,9	7,8	35	5,4	8,2	62	7,5	7,6
9	5,0	8,8	36	6,1	6,2	63	6,3	9,7
10	8,0	7,5	37	6,7	7,8	64	5,9	6,7
11	7,5	8,8	38	7,3	6,3	65	6,7	8,3
12	6,6	7,4	39	6,3	7,7	66	6,8	6,7
13	6,5	6,0	40	5,7	8,5	67	7,3	7,1
14	6,7	8,0	41	7,3	6,2	68	7,4	6,9
15	7,3	6,8	42	5,8	6,2	69	5,6	8,4
16	6,3	9,1	43	5,9	8,9	70	7,7	5,4
17	6,5	8,9	44	5,8	9,7	71	6,5	7,3
18	8,6	8,8	45	6,9	6,8	72	8,1	6,4
19	6,3	6,7	46	6,5	6,9	73	6,8	7,7
20	5,4	9,0	47	5,9	11,7	74	6,7	5,8
21	7,4	6,9	48	6,2	4,4	75	8,0	8,4
22	7,9	8,9	49	6,4	9,8	76	6,3	5,6
23	6,4	6,6	50	7,5	6,9	77	6,0	9,5
24	7,0	6,0	51	7,3	6,8	78	6,3	4,9
25	5,4	5,8	52	8,3	8,8	79	7,3	7,7
26	7,4	8,5	53	6,6	4,7	80	6,4	7,4
27	7,1	7,8	54	7,4	7,7			

<u>Решение:</u>

1. Обозначим через «*x*» количество эритроцитов, а через «*y*» - количество лейкоцитов. Находим *min* и *max* значение вариант по двум признакам:

$minV_x$=5.0 $maxV_x$=8.6

$minV_y$=4.4 $maxV_y$=12.8

Находим классовый промежуток по формуле:

$$K = \frac{V_{\max} - V_{\min}}{числоклассов}$$

$$K_x = \frac{8,6-5,0}{9} = 0,4 \approx 0,5$$

$$K_x = \frac{12,8-4,4}{9} = 0,93 \approx 1,0$$

2. Начертить корреляционную решетку (таблица 10) и записывают в ней вверху по горизонтали значение классов по количеству эритроцитов (*x*) от минимального к максимальному, а по вертикали располагают классы по количеству лейкоцитов (*y*) от больших к меньшим.

3. Разносят данные о 80 животных в клетки корреляционной решетки с учетом одновременно значений двух признаков. Так, если у коровы содержится крови 6.6 млн. эритроцитов и 7.4 тыс. лейкоцитов, то она отмечается одной точкой в клетке на пересечении классов 6,5 – 6,9 по количеству эритроцитов и класса 7,0 – 7,9 по количеству лейкоцитов.

Таблица 10 – Корреляционная решетка распределения коров по количеству эритроцитов (x) и лейкоцитов (y) в 1 мм^3 крови.

x / y	5.0-5.4	5.5-5.9	6.0-6.4	6.5-6.9	7.0-7.4	7.5-7.9	8.0-8.4	8.5-8.9	f_x	a_x	$f_x \times a_x$	$f_x \times a_x^2$
12.0-12.9		1^{-10}		1					2	+5	10	50
11.0-11.9		1^{-8}							1	+4	4	16
10.0-10.9									0	+3	0	0
9.0-9.9		2^{-4}	4^{-2}						6	+2	12	24
8.0-8.9	2^{-3}	3^{-2}	2^{-1}	5	1^{1}	4^{2}	2^{3}	1^{4}	20	+1	20	20
7.0-7.9		4	4	7	4	3	1		23	0	0	0
6.0-6.9		4^{2}	5^{1}	4	6^{-1}	1^{-2}	1^{-3}		21	-1	-21	21
5.0-5.9	1^{6}		1^{2}	1		1^{-4}			4	-2	-8	16
4.0-4.9			2^{3}	1					3	-3	-9	27
f_y	3	15	18	19	11	9	4	1	80		∑=8	∑=174
a_y	-3	-2	-1	0	+1	+2	+3	+4				
$f_y \times a_y$	-9	-30	-18	0	11	18	12	4			∑=-12	
$f_y \times a_y^2$	27	60	18	0	11	36	36	16			∑=204	

4. При вычислении коэффициента корреляции в многочисленных выборках используют формулу:

$$r = \frac{\sum fa_x \times a_y - b_x \times b_y \times n}{\sigma_x \times \sigma_y \times n}$$

5. Подсчитывают частоту вариант f_x и f_y. Выделяют жирными линиями классы условной средней *(A_x)* 6.5 – 6.9 и *(A_y)* 7.0 – 7.9. В решетке образуется четыре квадранта (квадрант – четвертая часть от поделенной плоскости двумя взаимно перпендикулярными прямыми). Выделенные классы в дальнейших расчетах не участвуют.

6. Определяют ряд условных отклонений и ау. Находят ряды произведений $f_x \times a_x$, $f_y \times a_y$, $f_x \times a_x^2$, $f_y \times a_y^2$

7. Подсчитываем суммы: $\sum f_x \times a_x = 8$ $\sum = f_x \times a_x^2 174$

$\sum f_y \times a_y = -12$ $\sum f_y \times a_y^2 = 204$

8.Чтобы найти $f_x \times a_x, \times a_y$ (первая часть формулы), необходимо произведение в каждом классе записать сверху в виде степени над соответствующими частотами (с учетом знаков), например, во 2-ом классе одно животное, следовательно,

произведение $a_{x,} \times a_y = 5\times(-2)$ =-10 и записываем в решетке 1-го квадранта 1^{-10} и т.д.
8.Чтобы найти сумму $f_x\times a_{x,} \times a_y$ по 1-му квадранту: 1×(-10)+1× (-8)+2× (-4)+4× (-2)+2× (-3)+3× (-2)+2× (-1)=-10+(-8)+(-8)+(-8)+(-6)+(-6)+(-2)=-48
по 2-му квадранту: 1×1+4×2+2×3+1×4=1+8+6+4=19
по 3-му квадранту: 4×2+5×1+1×6+1×2+2×3=8+5+6+2+6=27
по 4-му квадранту: 6× (-1)+1× (-2)+1× (-3)+1× (-4)=-6+(-2)+(-3)+(-4)=-15
По четырем квадрантам $\sum f_x\times a_{x,} \times a_y$ =-47+l9+27+(-15)=-l7
9. Отклонение рассчитывают по формуле:

$$b = \frac{\sum fa}{n}$$

$$b_x = \frac{\sum f_x \times a_x}{n} = \frac{8}{80} = 0.1$$

$$b_y = \frac{\sum f_y \times a_y}{n} = \frac{-12}{80} = -0.15$$

10. Сигма рассчитывается по формуле (классовый промежуток не учитывается):

$$\sigma = \pm\sqrt{\frac{\sum fa^2}{n} - b^2}$$

$$\sigma_x = \pm\sqrt{\frac{\sum f_x \times a_x^2}{n} - b_x^2} = \pm\sqrt{\frac{174}{80} - 0.1} = \pm\sqrt{2.175 - 0.01} = \pm\sqrt{2.165} = \pm 1.47$$

$$\sigma_y = \pm\sqrt{\frac{\sum f_y \times a_y^2}{n} - b_y^2} = \pm\sqrt{\frac{204}{80} - (-0.15)^2} = \pm\sqrt{2.55 - 0.3} = \pm\sqrt{2.25} = \pm 1.50$$

11. Подставляем полученные значения в формулу:

$$r = \frac{\sum fa_x \times a_y - b_x \times b_y \times n}{\sigma_x \times \sigma_y \times n} = \frac{-17 - 0{,}1 \times (-0{,}15) \times 80}{1{,}47 \times 1{,}5 \times 80} = \frac{202.8}{176.4} = 1{,}15$$

Вывод: у исследуемых коров существует положительная корреляция между количеством эритроцитов и лейкоцитов.

Ошибка коэффициента корреляции: так как коэффициент корреляции высчитан не по генеральной совокупности, а по выборочной совокупности поэтому он имеет ошибку выборочности, которая рассчитывается по формуле:

$$m_r = \frac{1 - r^2}{\sqrt{n}} = \frac{1 - 1.15^2}{\sqrt{80}} = \frac{1 - 1.32}{8.94} = \frac{-0.32}{8.94} = -0.03$$

Достоверность коэффициента корреляции. Когда известна ошибка, можно определить степень достоверности г. При этом исходят из нулевой гипотезы, т.е. предполагают, что в генеральной совокупности связь между изучаемыми признаками отсутствует. Только при значении t, равном или больше табличного значения (при вероятности 0,95, 0,99 или 0,999), нулевая гипотеза отвергается и значение будет достоверно.
В нашем примере:

$$t = \frac{r}{m_r} = \frac{1.15}{0.036} = 31.9$$

С учетом степеней свободы, которое равно υ=n-2=80-2=78, находим значение tst (табл. 7). Они равны $t_{0,95}$=1,96; $t_{0,99}$=2,58; $t_{0,999}$=3,29. Так как наше значение t=31,9, то мы можем с достоверностью 99,9 % сделать вывод, что у исследованных коров действительно существует положительная корреляция между количеством эритроцитов и лейкоцитов.

Определение доверительных границ коэффициента корреляции.

После доказательства достоверности выборочного коэффициента корреляции нужно установить доверительные границы генерального значения коэффициента корреляции r, т.е. определить, в каких пределах может находиться коэффициент корреляции, вычисленный для всей генеральной совокупности.

Минимальная (r_1) и максимальная *(r_2)* граница доверительного интервала определяются по формулам:

$$r_1 = \frac{r-d}{1-r\times d}$$

$$r_2 = \frac{r+d}{1+r\times d}$$

где число *d* - число, соответствующее количеству пар n и доверительной вероятности *p* (табл. 11).

Найдем доверительные границы коэффициента корреляции ($\hat{r}$) при доверительной вероятности *p=0,95*

В нашем примере *r=1,15; n=80*

По таблице для *n=80 p=0,95 d=0,22* тогда

$$r_1 = \frac{r-d}{1-r\times d}$$

Вывод: $0<f<0,43$ (p=0,95)

Таблица 11 – Значения *d* для определения границ достоверности интервала коэффициента корреляции

n	*d*	*n*	*d*	*n*	*d*
4	0.96 0.99	18	0.47 0.58	80	0.22 0.29
5	0.88 0.95	20	0.44 0.56	90	0.21 0.27
6	0.81 0.90	25	0.40 0.50	100	0.20 0.26
7	0.75 0.86	30	0.36 0.46	125	0.18 0.23
8	0.70 0.82	35	0.33 0.43	150	0.16 0.21
9	0.66 0.78	40	0.31 0.40	200	0.14 0.18
10	0.63 0.75	45	0.29 0.38	300	0.11 0.15
12	0.57 0.70	50	0.28 0.36	500	0.09 0.12
14	0.53 0.65	60	0.25 0.33	1000	0.06 0.08
16	0.50 0.61	70	0.24 0.31	10000	0.02 0.03

Примечание: верхний ряд чисел соответствует доверительной вероятности – 0.95, нижний ряд –

0.99

Занятие 9. Вычисление коэффициента корреляции в малочисленных выборках (n<30)

Цель занятия. Освоение метода вычисления показателей связи между признаками в малочисленных выборках.

Методические указания. Коэффициент корреляции в малых выборках широко применяют при определении связи между признаками родственных групп животных для решения вопросов связанных с выявлением закономерностей наследования и наследственности.

Наиболее удобны для вычисления коэффициенты фенотипической корреляции в малых выборка следующие формулы:

$$1. r = \frac{\sum x \times y - \frac{\sum x \times \sum y}{n}}{\sqrt{C_x \times C_y}}$$

$$2. r = \frac{C_x + C_y - C_d}{2\sqrt{C_x \times C_y}},$$

где n – число животных, изучаемых по 2-м признакам;

x и y – значение варианта первого и второго признака;

C – сумма квадратов центральных отклонений, вычисляемая по формуле:

$$C_x = \sum x^2 - \frac{\left(\sum x\right)^2}{n}$$

$$C_y = \sum y^2 - \frac{\left(\sum y\right)^2}{n}$$

$$C_d = \sum d^2 - \frac{\left(\sum d\right)^2}{n}$$

Задание: Вычислим коэффициент корреляции между содержанием белка в молоке коров (x, %) и их дочерей (y, %) по данным малой выборки (n=10) таблица 12.

Таблица 12 – Вычисление коэффициента корреляции между содержанием белка в молоке коров и их дочерей.

Занятие . Коэффициент наследуемости *(h^2)*.

Цель занятия. Освоение методов вычисления коэффициента наследуемости различными способами.

Методические указания. Изменчивость любого количественного признака обусловлена совместным влиянием наследственности и среды. Наследуемость – это доля генетического разнообразия в общей изменчивости признака. Количественные признаки формируются в процессе индивидуального развития на основе наследственности и факторов внешней среды. Степень влияния этих факторов на развитие разных признаков не одинаковы.

Поэтому при селекции животных важно знать относительную долю наследственной изменчивости в общей фенотипической изменчивости признака в конкретной популяции. Для этого вычисляется особый генетический параметр, называемый коэффициентом наследуемости *(h^2).*

Коэффициент наследуемости применяется в селекции животных для характеристики генетической структуры популяции по различным признакам, указывает на эффективность отбора, необходим для прогнозирования результатов отбора и обоснованного планирования роста продуктивности, используется при оценке производителей по качеству потомства, а также для оценки гетерозиса и прогнозирования результатов отбора у гетерозиготных гибридов.

Коэффициент наследуемости выражает долю генетической изменчивости в общей фенотипической изменчивости по данному признаку в популяции.

Показатель наследуемости относится только к проявлению признаков в популяции, а не у особи и поэтому его не следует смешивать с понятием наследственность. Наследуемость разделяется на два вида в широком и узком смысле слова. Наследуемость в широком смысле является отношением всей генетической варианты (σ_G^2), обусловленной аддитивным и эпистатическим действием, влиянием доминирования и сверхдоминирования, к общей фенотипической вариансе (σ_P^2), отсюда следует что коэффициент наследуемости равен:

$$h^2 = \frac{\sigma_G^2}{\sigma_P^2}$$

На практике в основном используется коэффициент наследуемости в узком смысле слова, который определяется как отношение аддитивной генетической вариансы (σ_A^2) к общей фенотипической вариансе (σ_P^2). Тогда формула расчета коэффициента наследуемости будет выглядеть так:

$$h^2 = \frac{\sigma_A^2}{\sigma_P^2}$$

Коэффициенты наследуемости вычисляются на основе степени сходства между родственниками.

Методы популяционной генетики позволяют для каждого признака определить долю влияния каждого фактора вычислением коэффициента наследуемости. Значение степени наследуемости различных признаков в конкретных условиях имеет важное значение для определения эффективности селекции. Чем выше коэффициент наследуемости, тем больше развитие данного признака зависит от наследственности, тем эффективнее отбор.

В зоотехнии наиболее удобны методы, основанные на учете коррелятивных и регрессивных связей между родственниками (между матерями и дочерями, полусибсами и полными сибсами), а также установлении долей различных источников изменчивости с помощью дисперсионного анализа.

Для вычисления коэффициента наследуемости необходимо определить коэффициент корреляции или регрессии и воспользоваться следующими формулами:

1. С помощью коэффициента корреляции (r) между показателями дочерей и теми же показателями матерей по формуле:

$$h^2 = 2 \times r_{м/д}$$

где h^2 – коэффициент наследуемости;

$r_{м/д}$ – коэффициент корреляции между признаками матерей и их дочерей (или сыновей и отцов).

По мнению некоторых ученых использование данного метода приводит к завышению h^2, а иногда и к явно ошибочным результатам, когда h^2 больше единицы. Поэтому они предлагают за h^2 брать коэффициент корреляции между родителями и потомками без его удвоения $h^2 = r_{м/д}$. Формула используется в тех случаях, когда различия признаков у родителей и потомков незначительны.

2. При разной интенсивности отбора среди родителей и потомков показатели разнообразия признака резко различаются. В таких случаях целесообразно пользоваться формулой:

$$h^2 = 2 \times R_{м/д}$$

где $2 \times R$ – 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 формуле:

$$h^2 = \frac{Длучшие - Дхудшие}{Млучшие - Мхудшие} \times 2$$

где - коэффициент наследуемости; $М_{лучшие}$ – средние показатели группы лучших животных стада; $М_{худшие}$ – средние показатели группы худших животных стада; $Д_{лучшие}$ – средний показатель потомства, полученного от лучших животных стада; $Д_{худшие}$ – средний показатель потомства от худших животных стада.

Например, в группе лучших животных стада средний удой был 8504 кг, а в группе худших – 7350 кг. Продуктивность дочерей, полученных от лучших коров была 7530 кг, а от группы худших – 7390 кг.

Подставим данные в формулу получим:

$$h^2 = \frac{7530 - 7390}{8504 - 7350} \times 2 = \frac{140}{1154} \times 2 = 0{,}24$$

Этот показатель (0,24) свидетельствует о низкой наследуемости и большой зависимости величины удоя от условий кормления и содержания животных.

3. Коэффициент наследуемости можно вычислить по следующей формуле:

где - коэффициент наследуемости; *Дп* - превосходство потомков (в среднем) над средним показателем стада; *Др* - среднее превосходство обоих родителей над средним показателем стада.

Применение этой формулы можно проиллюстрировать на следующем примере. Средняя масса овец в хозяйстве была в пределах 43,0 кг. В элитную группу отобраны овцы средней живой массой 54,0 кг. Если бы такая живая масса родителей была полностью унаследована потомством, то и оно в среднем имело бы живую массу 54,0 кг, но фактически масса потомства оказалась 46,5 кг, что обусловлено регрессией.