

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РФ  
ФГБОУ ВО «БРЯНСКИЙ ГОСУДАРСТВЕННЫЙ  
АГРАРНЫЙ УНИВЕРСИТЕТ»  
*Институт ветеринарной медицины и биотехнологии  
Кафедра кормления животных и частной зоотехнии*

Е.В. ОВСЕЕНКО, Н.П. БАЗУТКО

# **МАТЕМАТИЧЕСКИЕ МЕТОДЫ В БИОЛОГИИ**

**УЧЕБНО - МЕТОДИЧЕСКОЕ ПОСОБИЕ**

**ДЛЯ СТУДЕНТОВ, ОБУЧАЮЩИХСЯ ПО НАПРАВЛЕНИЮ ПОДГОТОВКИ  
36.04.02 «Зоотехния» (уровень высшего  
образования – Магистратура) и аспирантов**

БРЯНСКАЯ ОБЛАСТЬ 2015

УДК519.2:573 (07)

ББК 22.172:28

О.34

Овсеенко Е.В., Базутко Н.П. Математические методы в биологии. Учебно-методическое пособие для студентов, обучающихся по направлению подготовки 36.04.02 «Зоотехния» (уровень высшего образования – Магистратура) и аспирантов. / Е.В. Овсеенко, Н.П. Базутко – Брянская область: Издательство Брянского ГАУ, 2015 – 69 с.

Издание учебного пособия - «Математические методы в биологии», предназначено для студентов, обучающихся по направлению подготовки 36.04.02 «Зоотехния» (уровень высшего образования – Магистратура) и аспирантов.

В учебном пособии изложены вопросы анализа биологических данных. В качестве основы использованы методы теории вероятности и математической статистики; представлено моделирование динамики биологических процессов, основанное на применении математического аппарата дифференциальных уравнений. Изложенный материал в учебном пособии позволяет магистрам реализовать следующие компетенции: ОК-11, ОК-12, ПК-2.

Рекомендовано к изданию решением методической комиссии института ветеринарной медицины и биотехнологии. Протокол № от « » июня 2015 г.

Рецензент: доктор сельскохозяйственных наук, профессор В.Е. Подольников

©Брянский ГАУ, 2015

© Овсеенко Е.В., 2015

© Базутко Н.П., 2015

## ВВЕДЕНИЕ

Развитие современной зоотехнии и ветеринарии невозможно без использования в большей или меньшей степени математики.

В производственной деятельности при планировании и анализе результатов работы на птицефабриках, в племенных заводах, на пунктах искусственного осеменения, в откормочных хозяйствах, зверосовхозах, при обработке производственных и племенных записей, при сводке протоколов хозяйственных опытов и наблюдений специалисту приходится применять простые и сложные математические методы.

В данном пособии изложены наиболее часто применяемые биометрические методы при проведении научных и производственных работ в зоотехнии и ветеринарии: средние величины, разнообразие признака в группе, достоверность выборочных показателей, корреляция, дисперсионный анализ, регрессия, современные способы расчета биометрических показателей.

*Занятие 1.*

## ОСНОВНЫЕ СВЕДЕНИЯ О БИОМЕТРИИ

Цель занятия: ознакомиться со структурой и содержанием биометрии, ее основными задачами, наиболее важными понятиями, свойствами совокупности, а также с основными биометрическими параметрами.

Современный уровень развития любой науки связан с использованием в ней (в большей или меньшей мере) математики. Генетика является не описательной, а точной наукой, изучающей наследственность и изменчивость организмов. Эти фундаментальные свойства организмов проявляются в количественных закономерностях наследования признаков и свойств, как у отдельных индивидов, так и у их совокупностей.

**БИОМЕТРИЯ** (от греч. *bios*- жизнь, *metron* – мера) – наука о применении математических методов в биологических исследованиях при изучении групповых свойств биологических объектов.

Содержанием биометрии является обработка данных наблюдений и экспериментов в биологических исследованиях.

Методы биометрии основаны на теории вероятности и законе больших чисел. Они позволяют дать математически точные характеристики свойств и признаков совокупностей, выявить степень генетического разнообразия признака и влияния на него различных факторов, прогнозировать эффект селекции.

Предмет биометрии (т.е. то, что она изучает) составляют те или иные варьирующие по величине (изменяющиеся в определенных пределах) признаки или свойства объектов совокупности.

Совокупностью в биометрии называют группу биологических объектов. Совокупностями являются породы, стада животных, линии, семейства, дочери определенного производителя, группа овец, на которых проводится опыт, количество эритроцитов в каком-то объеме крови животного и т.д.

Совокупность состоит из единиц или членов. Для стада овец единицей будет каждая овца. Число единиц, входящее в совокупность обозначается буквой *n*. Единица совокупности может характеризоваться определенными признаками. Например, коровы характеризуются удоями за некоторый отрезок времени, жирномолочностью и белкомолочностью, живой массой, промерами, мастью, рогатостью или комолостью, числом эритроцитов и т. д. Сумма отдельных измерений или наблюдений также является совокупностью.

Величину изучаемого признака для какой-то единицы совокупности называют вариантой и обозначают  $X_1; X_2; X_3 \dots$ . Например, при изучении удоя (ц) за третью лактацию у четырех коров получены следующие данные: 42, 40, 38, 45. Эти величины и будут вариантами, т.е.  $X_1 = 42; X_2 = 40; X_3 = 38; X_4 = 45$ .

Различия между отдельными вариантами называются **изменчивостью** или **вариацией**. В этих случаях говорят, что удой или признак «варьирует». Варьирование признака у особей однородной группы (т.е. особей одного вида, возраста, одной породы, находящихся в аналогичных условиях кормления и содержания) обусловлено, с одной стороны, различиями в их наследственности, а с другой – влиянием внешних и внутренних факторов.

**Генеральная совокупность** – группа объектов, которую нужно охарактеризовать исследователю (например, все коровы черно-пестрой или другой породы). Объем генеральной совокупности может быть как очень большим (50 тыс., 500 тыс. и более), так и очень малым.

Генеральная совокупность может быть представлена объектами, которых еще нет фактически, но которые в принципе могут быть получены. Так, при оценке племенных качеств производителя генеральную совокупность образуют *все* его возможные потомки. В этом случае любое количество фактически полученных потомков производителя, используемых для характеристики его племенной ценности, будет лишь часть генеральной совокупности.

Это относится и к большинству опытов по изучению степени влияния различных факторов на животных. Например, при изучении влияния на продуктивность или рост животных определенного рациона, генеральную совокупность образуют все животные, которые в принципе могли бы получать данный рацион. Группа животных, на которых проводят опыт, будет представлять лишь часть генеральной совокупности.

Из этих примеров видно, что генеральную совокупность, как правило, приходится характеризовать по части – по так называемой *выборке*.

**Выборочная совокупность (выборка)** – это часть генеральной совокупности. Выборка является средством изучения и характеристики генеральной совокупности.

Выборка должна быть представительной – **репрезентативной**.

Выборка должна быть **типичной**, т.е. правильно отражать генеральную совокупность.

Каждый член выборки из генеральной совокупности должен быть отобран случайно. Для этого рекомендуется пронумеровать все объекты и отобрать некоторые из них по таблице случайных чисел (жеребьевка).

Существуют и другие способы отбора, обеспечивающие случайность, например механический - в выборку включается или каждый десятый, или каждый двадцатый объект и т. д.

Различают малую ( $n < 20 \dots 30$ ) и большую выборочную совокупности ( $n > 20 \dots 30$ ) Объем совокупности имеет значение при характеристике разнообразия признака и определении параметров генеральной

совокупности по параметрам выборки. Кроме того, различия между малой и большой совокупностью сказываются в технике (приемах) расчета основных параметров совокупности, хотя наличие современных счетных машин фактически ликвидировало эту разницу.

Сельскохозяйственные животные обладают большим разнообразием морфологических, физиологических, хозяйственно-полезных признаков.

Все хозяйственно-полезные признаки животных подразделяют на качественные и количественные.

К качественным признакам животных относятся: пол (мужской и женский), масть черная и белая, тип шерсти (тонкая, грубая, полутонкая, полугрубая), рогатость или комолость и т.д.

Количественные признаки изучают путем измерений, подсчетов и выражают цифрами, например, вес животных, удой, длина шерсти у овец, число сосков у свиней и др.

Качественные признаки четко отличаются друг от друга, и при расщеплении не вызывает трудностей группировка животных по признакам в те или иные фенотипические классы. Количественным же признакам невозможно дать точной качественной характеристики. По таким признакам между индивидами наблюдаются постепенные малозаметные переходы, а при расщеплении не образуются четко выделяемые фенотипические классы. Изменчивость подобных признаков является непрерывной.

Например, в молочном скотоводстве удой коров за лактацию колеблется от 1000 до 10000 кг и выше, и в большой популяции в каждый фенотипический класс, включающий интервал 10 кг молока входит несколько животных. Таким образом, коров можно условно распределить более чем на 100 фенотипических классов, но ни один из них не будет отражать генотипического расщепления.

Непрерывная вариация количественного признака в популяции объясняется, прежде всего, действием многих генов, которые носят название полигены. Каждый из полигенов оказывает незначительное влияние на изменчивость количественного признака, про который можно сказать, что он характеризуется непрерывной изменчивостью и полигенным наследованием.

Современная селекция опирается в основном на оценку и отбор животных по количественным признакам, имеющим экономическое значение. Поэтому изучение закономерностей наследования таких признаков имеет исключительно важное значение при оценке племенных качеств животных.

Для характеристики популяции животных по количественным признакам невозможно использовать частоты генов и распределение генотипов.

На начальном этапе анализа количественных признаков рассчитывают общепринятые статистические (фенотипические) параметры :  $\bar{X}$ ,  $\sigma$ .

С помощью указанных статистических параметров с определенной точностью можно характеризовать родителей и потомков, полученных от прямых и возвратных скрещиваний.

Использование таких параметров при подборе животных для скрещивания позволило установить общие закономерности наследования количественных признаков.

При скрещивании животных заметно различающихся по уровню развития количественного признака, в первом поколении не проявляется доминирование признака одного из родителей, а во втором поколении не наблюдается закономерного расщепления на определенное число фенотипически разных классов.

В общем виде закономерности наследования количественных признаков можно свести к следующим положениям:

1. Средняя арифметическая признака у животного I поколения, как правило бывает промежуточной между средними арифметическими признака у родителей.

2. Между средними величинами признака у животных I и II поколений различия недостоверны. Однако фенотипическая изменчивость животных II поколения значительно выше, чем у животных  $F_1$ , что свидетельствует о расщеплении генотипов.

3. Кривые изменчивости признаков у животных, полученных от возвратных скрещиваний, сдвинуты ближе к кривым распределения тех родителей, с которыми было произведено скрещивание.

4. В основе наследственной передачи количественных признаков лежит тот же механизм расщепления и рекомбинации генов, что и при передаче качественных признаков.

К важнейшим генетическим методам, используемым в селекции скота, относятся:

- методы определения средних величин,
- методы определения разнообразия признаков,
- методы измерения связей между признаками,
- методы измерения соответствия выборочных данных параметрам генеральной совокупности,
- методы определения наследуемости и повторяемости признаков.

Определение указанных генетических показателей позволяет установить фактический уровень селекции и прогнозировать продуктивность животных.

## *Занятие 2.*

# 1. РАСЧЕТ ОСНОВНЫХ СТАТИСТИЧЕСКИХ ПАРАМЕТРОВ

## В БОЛЬШОЙ ВЫБОРОЧНОЙ СОВОКУПНОСТИ $\langle \bar{X}, \sigma, C_v, r, R \rangle$

Цель занятия: ознакомиться с правилами построения вариационных рядов, с графическим изображением вариационного ряда, типами кривых распределения.

Основные статистические параметры в больших выборочных совокупностях вычисляются по вариационному ряду.

### 1.1. ПОСТРОЕНИЕ ВАРИАЦИОННОГО РЯДА И ВАРИАЦИОННЫХ КРИВЫХ

Если имеется многочисленная группа особей, изучаемых по какому-нибудь признаку, то различные значения этого признака встречаются неодинаковое количество раз: одни чаще, другие реже. Это явление называется распределением признака.

При характеристике количественных признаков и большом числе вариантов производят группировку данных и их разность по классам, т.е. строят вариационный ряд.

Вариационный ряд показывает реально существующее распределение особей в группе (совокупности) по величине признака.

**Вариационный ряд** – это двойной ряд чисел, один из которых указывает значение признака в классах, а другой – число животных (объектов) в классах.

**Класс** - это часть совокупности, в которой объединены все сходные по величине изучаемого признака объекты.

Вариационный ряд дает наглядное представление о том, как часто встречаются отдельные варианты в данной совокупности, в каком количестве и как они распределяются, т.е. показывает закономерность варьирования изучаемого признака.

#### ОБЩИЕ ПРАВИЛА ПОСТРОЕНИЯ ВАРИАЦИОННОГО РЯДА.

*Пример 1.* Предположим, что мы собрали данные о суточных удоях 40 коров черно-пестрой породы, записанных в бонитировочную ведомость и получили следующие значения (в кг):

12,0	15,8	18,7	13,3	16,0	13,8	6,0	15,2
8,4	20,4	11,2	8,0	17,4	11,6	8,2	14,0
10,7	16,7	12,3	13,7	13,1	10,2	18,3	9,5
12,3	13,9	14,7	15,1	12,5	14,8	22,9	10,8
7,6	14,1	16,1	9,2	11,5	9,4	11,0	12,0

Для построения вариационного ряда необходимо:

1. Просмотрев весь материал найти максимальное и минимальное значение признака (суточного удоя коров) и тем самым определить значение  $\text{lim}$  :

$$X \text{ min} = 6,0 \text{ и } X \text{ max} = 22,9 ;$$

$$\text{lim} = 6,0 \div 22,9 \text{ кг.}$$

2. Определить размах разнообразия признака:  
 $\rho = X \text{ max} - X \text{ min} = 22,9 - 6,0 = 16,9 \text{ кг.}$
3. Определить число классов.

Число классов зависит от количества особей в выборке.

Объем выборки (n)	Рекомендованное число классов
30-60	6-8
61-100	7-10
101-200	9-12
201-500	12-17

Чем больше число классов, на которые разбит вариационный ряд, тем выше будет точность вычисленных с его помощью параметров.

В данном примере совокупность разбиваем на 6 классов.

4. Определить интервал между соседними классами (классовый промежуток), который обозначается буквой  $K$ .

$$K = \frac{X \text{ max} - X \text{ min}}{\text{Число классов}}$$

**Классовый промежуток** - это величина, на которую отличается значение одного класса от значения другого (соседнего).

$$K = \frac{22,9 - 6,0}{6} = \frac{16,9}{6} = 2,8 = 3$$

Полученный указатель классового промежутка для удобства следует округлить, например, 2,8 до 3; 382 до 400; 2,5 до 3; 17 до 20 и т.д.

5. Найти нижние границы классов.

За нижнюю границу 1-го класса обычно принимают минимальное значение признака. Однако для удобства допускается округление минимальной величины ( $X \text{ min}$ ) в меньшую сторону. Для этого из минимального значения вычисляют величину равную  $1/2$  классового промежутка ( $K$ ). Если, например, минимальное значение признака (удой за лактацию) в выборке составило 2138 кг, а классовый промежуток  $K = 500$  кг, то удобнее нижнюю границу 1-го класса определить в 2000 кг отбросив 138 кг (что несколько меньше  $1/2 K$ ).

Нижняя граница 2-го класса = нижняя граница 1-го класса + классовый промежуток.

Нижняя граница 3-го класса = нижняя граница 2-го класса + классовый промежуток.

Так же находим нижние границы и остальных классов.

В нашем примере минимальное значение признака 6,0, которое и будет нижней границей 1-го класса. Прибавив к ней величину классового промежутка (3 кг) находим нижнюю границу второго класса и т.д.

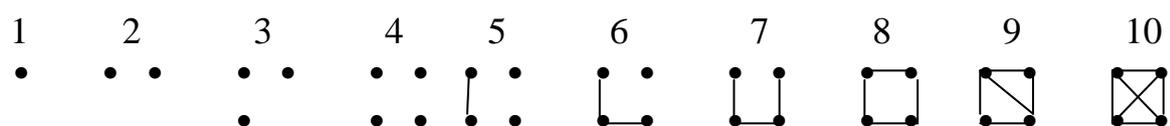
Чтобы при разnosке варианта не попала на границу между классами, проставляют верхнюю границу каждого класса. Для этого уменьшают нижнюю границу следующего класса на 0,1; 0,01; или на 1, в зависимости от точности измерения.

При записи границ классов обязательно обращают внимание на то, чтобы последний класс вариационного ряда включал в себя варианты с максимальным значением признака, а первый - с минимальным.

6. Найти средние значения классов (W).

$W = \text{нижняя граница класса} + \text{половина классового промежутка}$

7. Варианты изучаемой совокупности (X) разносят по классам. Разnosку вариантов по классам производят по порядку их записи (без выбора): отмечают точкой и черточкой в соответствующем классе, пользуясь обозначениями так называемой десятичной системы (или конвертика).



После разnosки вариантов по классам подсчитывают их количество в каждом классе.

Число вариантов в классе называют частотами и обозначают символом  $f$ .

Число частот при разnosке должно быть равно общему числу вариантов в выборке ( $n$ ).

В примере  $f = n = 40$ .

Составленный вариационный ряд будет иметь следующий вид:

Границы классов	Средние значения классов	Разnosка	Частота $f$
6,0 - 8,9	7,5		5
9,0 - 11,9	10,5		10
12,0 - 14,9	13,5		14
15,0 - 17,9	16,5		7

18,0 - 20,9	19,5	• • •	3
21,0 - 23,9	22,5	•	1
			$\sum f = 40 = n$

В таком виде вариационный ряд готов к математической обработке с целью нахождения нужных параметров совокупности.

Рассмотрение вариационного ряда без всяких вычислений позволяет определить величину основных статистических показателей и разнообразия с таким приближением, которое вполне достаточно для первого ознакомления с признаком.

Если взять вариационный ряд, приведенный в примере 1, то без вычислений можно сказать следующее:

а) средняя арифметическая признака находится между 12 и 15, недалеко от 13,5;

б) мода признака (наиболее часто встречающееся значение) равна примерно 12;

в) лимиты (lim) и размах признака примерно 6,0 - 24,0 (18);

В большинстве распределений, с которыми приходится встречаться зоотехнику, проявляется определенная закономерность.

1. Крайние значения (наименьшие и наибольшие) появляются редко;

2. Чем ближе значение признака к средней арифметической, тем оно чаще встречается. На середину вариационного ряда приходится наибольшее число вариантов.

3. Частоты симметрично убывают от середины к краям вариационного ряда.

Подобное распределение значений признака получило название: нормальное распределение.

Такую закономерность можно использовать для решения многих практических задач.

Вариационный ряд можно представить в виде ступенчатой кривой, называемой гистограммой или полигона частот (вариационной кривой)

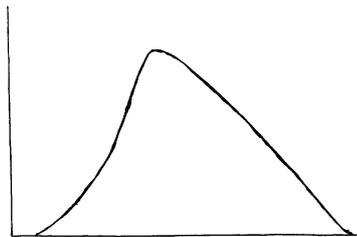
Для построения гистограммы (столбчатой диаграммы) по оси абсцисс откладывают границы классов, а по оси ординат – частоты.

Для построения полигона частот (вариационной кривой) по оси абсцисс откладывают средние значения классов, а по оси ординат – частоты.

Правильно изображенные гистограммы и вариационные кривые должны быть справа и слева доведены до пересечения с осью абсцисс.

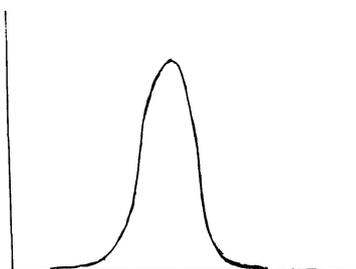
В реальном стаде распределение признака несколько отличается от нормального. Оно может быть:

### АССИМЕТРИЧНЫМ



Вершины вариационных кривых смещены влево или вправо. Причины асимметричного распределения различны: резкая неоднородность условий развития отобранных для изучения особей, преобладание в отобранной группе особей с лучшими или худшими наследственными задатками; разная степень реакции особей на лучшие и худшие условия.

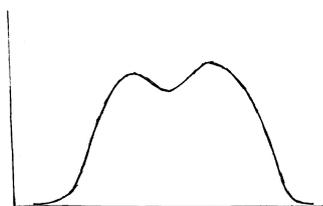
### ЭКСЦЕССИВНЫМ



Кривые вариационных рядов островершинные с резким преобладанием вариант в средних классах, так называемый – *эксцесс*.

Это может быть связано с реакцией организма на условия среды, когда оптимальное развитие признака достигается при довольно широких границах колебания условий среды, и только наиболее крайние отклонения этих условий влияют на большее или меньшее развитие признака. Аналогичное явление может произойти и под влиянием наследственности.

### ТРАНСГРЕССИВНЫМ



Кривые таких типов характерны для смешанного материала, например для животных разных пород, или выращенных в крайне неодинаковых условиях, или резко различающихся по наследственным особенностям.

Эти отличия от нормального распределения, как правило, не большие, и поэтому при анализе этим можно пренебречь.

Благодаря нормальности распределения количественных признаков с помощью биометрических методов можно выполнять следующие анализы:

1. Дать характеристику стада не только по средним величинам признака, но и по их изменчивости.

2. Определить повторяемость признака.
3. Определить наследуемость признака.
4. Дать характеристику взаимосвязи между отдельными признаками.
5. Рассчитать эффективность селекции различных вариантов программ и выбрать оптимальный.

Основными статистическими параметрами количественных признаков являются: средняя арифметическая ( $\bar{X}$ ), среднее квадратическое (стандартное) отклонение или сигма ( $\sigma$ ), ошибка средней арифметической ( $m_x$ ) и коэффициент изменчивости ( $C_v$ ).

**Задание №1.** Построить вариационный ряд и вариационную кривую по живой массе коров холмогорской породы (кг):

475	432	500	488	384	445	520	421	469	460	535	486	556	441	473	
534	505	432	434	417	406	450	491	420	445	426	490	442	439	429	421
470	391	482	390	426	488	411	474	440	539	488	433	546	493	472	441
463	468	483	442	524	447	430	485	440	488	439	445	504	550	495	536
426	388	407	425	390	418	465	391	365	427	448	452	458	387	374	360
545	454	467	456	441	434	519	488	443	493	457	425	429	445	442	393
405	422	400	456	441.											

### Занятие 3.

## 1.2 ВЫЧИСЛЕНИЕ СРЕДНЕЙ

### АРИФМЕТИЧЕСКОЙ ВЕЛИЧИНЫ $\bar{X}$

Цель занятия: ознакомиться с техникой нахождения основных параметров вариационного ряда (средней арифметической -  $\bar{X}$  и среднего квадратического отклонения -  $\sigma$ ), с оценкой структуры разнообразия в совокупности.

Для характеристики большинства хозяйственно полезных признаков используют среднюю арифметическую величину.  $\bar{X}$

Средняя арифметическая ( $\bar{X}$ ) показывает какое значение признака наиболее характерно для данной совокупности. Измеряется средняя арифметическая в тех же единицах, что и признак.

Средняя величина признака приобрела большое значение в генетических исследованиях и в практике племенной работы.

Для больших выборок применяют непрямые методы вычисления средней арифметической. Вычисление производится по формуле:

$$\bar{X} = A + v \quad \text{или} \quad \bar{X} = A + K \frac{\sum fa}{n}$$

где А – произвольно выбираемая условная средняя ;

v – поправка к условной средней.

K- классовый промежуток

f – частота

n – число вариант в выборке

a – отклонение отдельных классов от условной средней, выраженное в числе классовых промежутков.

*Пример:* Вычислить среднесуточный удой 40 коров черно-пестрой породы по вариационному ряду (см. пример выше).

В примере (1) с суточным удоем коров эти расчеты будут выглядеть так:

Среднее значение класса, кг	Частоты (f)	Отклонения (a)	f a
7,5	5	-2	-10
10,5	10	-1	-10
13,5	14	0	0
16,5	7	+1	7
19,5	3	+2	6
22,5	1	+3	3

$$\sum f a = -4$$

Для вычисления средней арифметической величины необходимо:

1. Выбрать условную среднюю (A). За условную среднюю A может быть принято среднее значение любого класса, но удобнее использовать среднее значение класса с наибольшим количеством вариант n, поскольку искомая X скорее всего будет близка к нему в данном примере A= 13,5

2. Устанавливаем на какое количество классовых промежутков отклоняется каждый класс от класса принятого за условную среднюю. Эти отклонения обозначают буквой (a).

Для класса среднее значение которого принято за A, всегда a=0.

Если среднее значение классов меньше условной средней, то отклонениям присваивается знак «минус» (-1, -2, -3 и т.д.), а если среднее значение классов больше условной средней, то отклонениям присваивается знак «плюс» (+1, +2, +3, и т.д.).

3. Умножить отклонение каждого класса (a) на количество вариант (f) в классе (учитывая знаки). Подсчитать сумму произведений f a ( $\sum f a$ ). В нашем случае она составляет  $\sum f a = -20+16 = -4$ .

4. Пользуясь формулой, рассчитать среднюю арифметическую

$$\bar{X} = A + K \frac{\sum fa}{n}$$

В данном примере A= 13,5, K = 3,  $\sum fa = -4$  при n =40

$$\bar{X} = 13,5 + 3 \frac{-4}{40} = 13,5 - 0,3 = 13,2 \text{ кг}$$

Вывод: Данная группа коров характеризуется среднесуточным удоем 13,2 кг молока.

**Задание №2.** Вычислить среднюю арифметическую по следующим данным. Среднее содержание жира в молоке коров айширской породы (%):

3,46	4,13	4,03	4,15	4,16	4,01	3,81	4,54	4,05	3,92	4,29	3,86	4,05	
4,06	4,08	3,59	3,71	4,25	3,82	4,16	3,99	4,01	4,02	4,04	4,03	4,01	3,83
3,96	4,05	4,12	4,01	4,31	4,05	4,27	4,11	4,18	4,02	4,03	4,13	4,11	3,61
4,28	3,95	3,85	3,83	4,11	4,01	3,91	4,28	4,21	4,20	3,92	4,05	4,03	4,40
4,21	3,95	3,90	4,12	4,13	4,03	4,02	3,75	4,02	4,26	4,10	4,15	3,95	4,12
4,03	4,01	3,42	3,81	4,05	4,13	4,05	4,06	4,02	3,76	4,21	4,28	4,11	3,96
4,02	3,59	3,93	4,07	3,73	4,01.								

### 1.3. ПОКАЗАТЕЛИ, ХАРАКТЕРИЗУЮЩИЕ РАЗНООБРАЗИЕ ПРИЗНАКОВ ( $\lim$ , $\sigma$ , $C_v$ ).

Важное свойство совокупности - разнообразие входящих в нее объектов, что обусловлено изменчивостью.

Оценка степени разнообразия животных в изучаемых группах – важнейший элемент селекционной работы. Большое значение придается ей в связи с переводом животноводства на промышленную технологию для обеспечения выпуска стандартной продукции.

Зоотехнику- селекционеру в начале создания новых пород, породных групп, линий важно знать степень разнообразия исходного материала: чем разнообразнее племенные группы, тем больше имеется возможности для отбора и подбора. Однако при завершении работы по созданию пород, линий и т. д. наряду с повышением средней арифметической хозяйственно-полезных признаков требуется уменьшение разнообразия, создание однородных групп по экстерьерным признакам, по качеству шерсти и т.д.

Разнообразие объектов в совокупности характеризуется целым рядом показателей:

А). *Лимиты* ( $\lim$ ) или пределы  $\lim = X_{\min} \dots X_{\max}$ . Границы разнообразия определяются крайними вариантами.

Они отмечают наивысший показатель продуктивности, имеющийся в исследуемой группе (удой, настриг шерсти и т. д.). В то же время лимиты отмечают и наличие наименее продуктивных животных нерентабельных для хозяйства.

Б) *Размах разнообразия*  $\rho = X_{\max} - X_{\min}$

Чем меньше  $\rho$ , тем меньше разнообразие и наоборот.

*Пример:* В двух хозяйствах имеется по пять быков, живая масса которых следующая (кг) :

1 – хозяйство 640 645 650 655 660  $\bar{X}_1 = 650$  кг

2 – хозяйство 610 630 650 670 690  $\bar{X}_2 = 650$  кг

Средняя живая масса быков в обоих хозяйствах одинаковая, однако в первом хозяйстве разнообразие по этому признаку гораздо меньше, чем во втором.

В приведенном примере лимиты таковы:

$\text{Lim}_1 = 640$  -----  $660$  кг;

$\text{Lim}_2 = 610$  -----  $690$  кг.

Предположим, в двух хозяйствах имеется по десять баранов, живая масса которых следующая (кг):

1 – хозяйство 100 110 120 130 140 150 160 170 180 190

$\bar{X}_1 = 145$  кг,  $\text{Lim}_1 = 100 - 190$  кг

2 – хозяйство 100 145 145 145 145 145 145 145 145 190

$\bar{X}_2 = 145$  кг,  $\text{Lim}_2 = 100 - 190$  кг

Средние арифметические и лимиты для обеих групп одинаковы, но степень разнообразия этих групп неодинакова. В первом хозяйстве все бараны по весу различны, а во втором – из десяти баранов восемь имеют одинаковый вес. Разнообразие первой группы несомненно больше, но отметить это с помощью лимитов невозможно. В таком случае необходимо привлечение основного показателя разнообразия – среднего квадратического отклонения ( $\sigma$ ).

В) *Среднее квадратическое отклонение*  $\sigma$  (сигма) – используется и как самостоятельный показатель и для определения целого ряда других параметров совокупности (коэффициента вариации, ошибок репрезентативности, коэффициентов корреляции и регрессии и др.). Сигма – показатель именованный и выражается в тех же единицах, что и изучаемый признак.

Вычисление среднего квадратического отклонения в многочисленных выборках проводят по формуле:

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

Вычислить среднее квадратическое отклонение по вариационному ряду (смотри пример 1).

Среднее значение класса	Частоты (f)	Отклонения (a)	f a	f a <sup>2</sup>
7.5	5	-2	-10	20
10.5	10	-1	-10	10
13.5	14	0	0	0

16.5	7	+1	+7	7
19.5	3	+2	+6	12
22.5	1	+3	+3	9
K=3	n = f = 40		$\sum f a = -4$	$\sum f a^2 = 58$

Подставляем в формулу полученные значения:

$$\sigma = \pm K \sqrt{\frac{58}{40} - \left(\frac{-4}{40}\right)^2} = \pm 3 \sqrt{1,45 - (-0,1)^2} = \pm 3 \sqrt{1,45 - 0,01} = \pm 3 \sqrt{1,44} = \pm 3 \cdot 1,2 = \pm 3,6 \text{ кг.}$$

**Вывод :** В данной совокупности суточные удои коров отклоняются от среднесуточного удоя по группе на величину  $\pm 3,6$  кг.

Сигма имеет два знака (+ и -). Это свидетельствует об отклонении вариант от средней арифметической как в большую сторону так и в меньшую. Отмечено, что в пределах  $\bar{X} \pm 1 \sigma$ , находится 68% вариант совокупности, в пределах  $\bar{X} \pm 2\sigma$  – 95,5 % вариант, а в пределах  $\bar{X} \pm 3\sigma$  – 99,7 %, т.е. практически все варианты.

Г) Коэффициент вариации ( $C_v$ ) используют :

1) для сравнения степени разнообразия признаков выраженных в разных единицах измерения;

2) при сравнении степени разнообразия одинаковых признаков если средние арифметические сильно различаются.

Коэффициент вариации вычисляется по формуле:

$$C_v = \frac{\sigma}{\bar{X}} \cdot 100$$

Чем больше значение  $C_v$ , тем выше разнообразие селекционируемого признака в совокупности.

Выделяют степень разнообразия признака:

Сильную -  $C_v \geq 15\%$ .

Слабую  $C_v \leq 5\%$

Среднюю  $C_v < 15\%$

Пример: Определить степень разнообразия изучаемых признаков в данной группе.

Признаки	$\bar{X}$	$\sigma$	$C_v = \frac{100 \cdot \sigma}{\bar{X}}$
Живая масса (кг)	500	50	10 %
Удой за лактацию, кг.	4000	760	19 %
Содержание жира в молоке, %	4,0	0,3	7,5 %

**Вывод:** В исследуемой группе наиболее разнообразным был удой, а наименее – процент жира.

**Пример:** Сравнить степень варьирования одинаковых признаков у разных групп, если разность между средним арифметическим ( $\bar{X}_1, \bar{X}_2$ ) этих признаков очень большая. Средний удой стада № 1 равен 3000 кг. (

$\sigma = \pm 600$  кг). Средний удой стада № 2 равен 5000 кг. ( $\sigma = \pm 700$  кг). Коэффициент вариации для первого стада

$$C_v = \frac{600 \cdot 100}{3000} = 20 \%$$

а для второго

$$C_v = \frac{700 \cdot 100}{5000} = 14 \%$$

Хотя сигма во втором стаде была больше, чем в первом, но больший коэффициент вариации указывает на большую разнородность удоя коров в первом стаде.

**Задание №3** Вычислить среднее квадратическое отклонение и коэффициент вариации по данным о живой массе бычков костромской породы по следующим данным (кг):

	156	145	140	133	145	146	132	146	142	142	149	138	145	149	145
150	140	149	143	137	146	137	143	144	147	152	160	134	140	139	154
144	143	142	144	145	150	156	146	143	150	131	138	150	145	148	139
153	151	143	141	150	147	138	145	138	148	138	145	138	143	141	130
144	145	141	143	140	142	137	147	159	143	145	147	146	150	136	137
144	141	136	138	140	144	152	146	161	146	138	138	145	146	140	150
145	146	132	155	145	145	140	137	153	150	145	144	148	132	136	138
139	147	138	137	140	136	141	148.								

Занятие 4-5.

## 2. ОШИБКИ РЕПРЕЗЕНТАТИВНОСТИ И ИХ ПРИМЕНЕНИЕ В БИОМЕТРИИ.

Цель занятия: ознакомление с понятиями статистических ошибок (ошибок репрезентативности), с их источниками, методами расчета и использования в решении коренных задач биометрии: оценки генеральных параметров и определении достоверности разности.

При характеристике генеральной совокупности в большинстве случаев пользуются выборкой.

Любая выборка, являясь только частью генеральной совокупности, не может абсолютно точно отразить ее свойство.

Поэтому выборочные параметры будут отличаться от параметров генеральной совокупности.

Ограниченность объема выборки и случайность отбора объектов в выборку являются объективными источниками статистических ошибок (ошибок репрезентативности). При использовании выборочного метода, статистических ошибок нельзя избежать, их необходимо учитывать, чтобы делать научнообоснованные выводы.

**Ошибки репрезентативности** показывают степень соответствия выборочных параметров параметрам генеральной совокупности. Чем меньше значения ошибки, тем точнее вычисленный параметр, тем ближе его значение к значению соответствующего параметра генеральной совокупности.

Согласно закону больших чисел практически маловероятно существенное отклонение выборочного параметра  $\bar{X}$ ,  $\sigma$  и др. от соответствующего параметра генеральной совокупности  $\bar{X}$ ,  $\sigma$  и др., если число наблюдений достаточно велико. При изучении всех членов генеральной совокупности статистических ошибок не может быть, т. к. генеральный параметр рассчитывается непосредственно. Например, если необходимо охарактеризовать по удою группу коров, закрепленных за какой-то дояркой, то, принимая во внимание ограниченный объем данной генеральной совокупности, можно привести сплошное обследование, т. е. учесть удои у всех коров, входящих в данную группу. В этом случае рассчитанные по всем животным средняя арифметическая  $\bar{X}$  и среднее квадратическое отклонение  $\sigma$  будут генеральными параметрами. Ошибки вычисляются для всех выборочных параметров и обычно обозначаются буквой  $m$  с подстрочным указанием знака того параметра, для которого они определяются:  $m_x$ ;  $m_\sigma$   $m_{Cv}$  и т.д.

Выборочные параметры записываются вместе со своей ошибкой:

$$\bar{X} \pm m_x ; \sigma \pm m_\sigma ; C_v \pm m_{Cv} \text{ и т.д.}$$

Ошибки репрезентативности выборочной средней арифметической для больших выборок вычисляют по формуле:

Для малой выборки

$$m_x \equiv \frac{\sigma}{\sqrt{n-1}}$$

Чем больше объем выборки, тем меньше ошибка, тем меньше отклонение выборочной средней арифметической от генеральной средней арифметической.

Ошибка – величина именованная (см., кг., % и т.д.) выражается в тех же единицах, что и признак.

Ошибка среднего квадратического отклонения:

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

Коэффициента вариации

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

Коэффициента корреляции

$$m_r = \sqrt{\frac{1-r^2}{n-2}}, \quad (\text{если } n < 100)$$

$$m_r = \frac{1-r^2}{\sqrt{n-1}} \quad (\text{если } n \geq 100)$$

Коэффициента регрессии

$$m_{Ry/x} = \frac{\sigma_y}{\sigma_x} \cdot m_r ; m_{Rx/y} = \frac{\sigma_x}{\sigma_y} \cdot m_r$$

Ошибки репрезентативности используются для:

1. Установления доверительных границ параметров генеральной совокупности.
2. Определения достоверности выборочных параметров.
3. Определения достоверности разности между средними арифметическими двух выборочных совокупностей.

## **2.1. ОПРЕДЕЛЕНИЕ ДОВЕРИТЕЛЬНЫХ ГРАНИЦ ПАРАМЕТРОВ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ**

Доверительными границами называют крайние значения признака, в пределах которых с заданной вероятностью может находиться параметр генеральной совокупности, а интервал между ними – доверительным.

Доверительные границы для средней арифметической генеральной совокупности рассчитываются:

$$\hat{\bar{X}}_{генер.} = \bar{X}_{выбор} \pm t \cdot m_x, где$$

t (критерий надежности выводов) – показатель вероятности безошибочного прогноза (P). Надежность – это вероятность того, что генеральный параметр в действительности окажется внутри доверительных границ.

Устанавливают вероятность безошибочного прогноза (P).

В практике зоотехнических исследований сложились три основных уровня (порога) надежности выводов (P) и соответствующие им три значения критерия надежности (t).

Первый порог надежности P = 0,95, которому соответствует значение критерия t = 1,96.

Вероятность 0,95 означает, что если из генеральной совокупности сделать 100 выборок в каждой из которых рассчитать 100 интервалов

$\bar{X}_{выб.} \pm 1,96 \cdot m_x$ , то по крайней мере в 95 случаях генеральный параметр  $\bar{X}$  попадет в указанный интервал, а в 5 случаях может оказаться за его пределами.

Критерий надежности t = 1,96 используется при планировании обычных зоотехнических и научно- производственных опытов.

Второй порог надежности – повышенный P = 0,99, ему соответствует значение t = 2,57. Этот порог надежности применяется в проверочных опытах, в экономических и биологических исследованиях.

Третий порог надежности  $P = 0,999$  и соответствующее ему  $t = 3,29$  используется при проверке гипотез и спорных вопросов, а также при изучении действия различных лекарств в ветеринарии и медицине.

Для малых выборок критерий надежности определяют по таблице Стьюдента, для соответствующего числа степеней свободы (приложение):

$$\gamma = n - 1$$

Например, при  $n = 20$ ,  $\gamma = n - 1 = 20 - 1 = 19$ ,  $P = 0,95$   $t = 2,09$

В практике нижняя доверительная граница используется при планировании производства, а верхняя – при планировании материальных затрат (планировании потребности в материалах, кормах, оборудовании), поскольку вероятность того, что генеральный параметр может быть больше верхней и меньше нижней доверительной границы, очень низка (соответственно 0,05; 0,01 или 0,001).

Если нижняя доверительная граница окажется равной нулю или примет отрицательное значение, то выборочный параметр считается недостоверным.

*Пример:* При изучении мясных качеств животных казахской белоголовой породы, получили следующие первичные данные о массе туши двухлетних волов:  $n = 400$ ,  $\bar{X} = 200$  кг,  $\sigma = \pm 20$  кг.

Требуется дать прогноз средней массы туши для всех двухлетних волов этой породы, подготовленных для использования на мясо в аналогичных условиях.

1. Определяем вероятность безошибочного прогноза  $P = 0,95$ .
2. Находим критерий надежности по таблице критериев Стьюдента. При  $P = 0,95$  и при  $\gamma = n - 1 = 400 - 1 = 399$  равен  $t_1 = 2,0$ .
3. Ошибка репрезентативности выборочной средней:

$$m_x = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{400}} = 1,0 \text{ кг}$$

4. Максимальная погрешность оценки генерального параметра:

$$\Delta = t \cdot m = 2,0 \cdot 1,0 = 2,0 \text{ кг.}$$

5. Доверительные границы :

$$\hat{X}_{генер.} = 200 \pm 2,0 \text{ кг}$$

$$\hat{X}_{генер.} \left\{ \begin{array}{l} \text{Не более } \bar{X}_{выб.} + t \cdot m_x = 200 + 2 = 202 \text{ кг} \\ \text{Не менее } \bar{X}_{выб.} - t \cdot m_x = 200 - 2 = 198 \text{ кг} \end{array} \right.$$

Эти границы (198-202) в данном случае имеют вполне определенное производственное значение. По нижней границе, или по гарантированному минимуму, следует планировать общее поступление мяса от переработки всех таких животных, а по верхней границе, или по возможному максимуму, надо готовить технику переработки, емкость холодильников и транспортные средства для перевозок мяса от этого скота.

## **2.2. ОПРЕДЕЛЕНИЕ ДОСТОВЕРНОСТИ ВЫБОРОЧНЫХ ПАРАМЕТРОВ.**

Прежде чем находить доверительные границы генеральных параметров, необходимо установить, а достоверны ли выборочные параметры, т.е. будут ли выборочные параметры соответствовать параметрам генеральной совокупности.

Существует общее правило определения достоверности выборочных параметров: отношение выборочного параметра к его ошибке сравнивают с критерием надежности ( $t$  табл.) из таблицы Стьюдента для соответствующего числа степеней свободы  $\gamma = n - 1$ .

$$t_x = \frac{\bar{X}}{m_{\bar{X}}}; \quad t_\sigma = \frac{\sigma}{m_\sigma}; \quad t_{Cv} = \frac{Cv}{m_{Cv}} \quad \text{и т.д.},$$

где  $t_x$  – критерий достоверности выборочной средней арифметической.

Если критерий достоверности выборочного параметра больше или равен табличному критерию ( $t$  табл.) – то выборочный параметр достоверен, т.е. он будет соответствовать параметру генеральной совокупности, тогда можно указать доверительные границы генерального параметра.

Если критерий достоверности меньше табличного критерия – выборочный параметр недостоверен, что не позволяет судить о значении параметра генеральной совокупности.

## **2.3 .ОПРЕДЕЛЕНИЕ ДОСТОВЕРНОСТИ РАЗНОСТИ МЕЖДУ СРЕДНИМИ АРИФМЕТИЧЕСКИМИ ДВУХ ВЫБОРОЧНЫХ СОВОКУПНОСТЕЙ.**

В исследованиях часто необходимо сравнение средних арифметических двух групп животных (например, средний вес животных опытной и контрольной групп, среднюю продуктивность дочерей двух производителей и т.д.).

Выборочные средние арифметические двух сравниваемых групп всегда в некоторой мере отличаются друг от друга. По полученной в исследовании разности между двумя средними двух выборок нельзя непосредственно сделать вывод о разности между генеральными средними арифметическими сравниваемых совокупностей.

Необходимо определить *достоверность разности* выборочных средних арифметических.

*Достоверность* означает возможность обобщить данные опыта, перенести результаты исследования выборок на соответствующие генеральные совокупности.

Предположим, что в опыте выборочная средняя первой группы ( $\bar{X}_1$ ) – оказалась больше выборочной средней второй группы ( $\bar{X}_2$ ), т.е.  $\bar{X}_1 > \bar{X}_2$ .

Если разность выборочных средних достоверна – это означает, что такую же по знаку разность, какая получена в эксперименте, следует ожидать (с принятым уровнем надежности выводов) и между соответствующими генеральными средними арифметическими, т.е. большей по величине выборочной средней соответствует и большая генеральная средняя, а меньшей по величине выборочной средней соответствует меньшая генеральная средняя. Если выборочная средняя 1-й группа ( $\bar{X}_1$ ) больше выборочной средней 2-й ( $\bar{X}_2$ ) группы и разность достоверна, то и генеральная средняя  $\hat{X}_1$  будет больше генеральной средней  $\hat{X}_2$ .

В том случае, если разность недостоверна, остается неизвестным, находятся ли генеральные средние в таком же соотношении между собой, как и выборочные средние, или их значения противоположны, или они равны. Другими словами, если в эксперименте  $\bar{X}_1 > \bar{X}_2$ , но разность недостоверна, то не ясно:

$$\hat{X}_1 > \hat{X}_2 \quad \text{или} \quad \hat{X}_1 = \hat{X}_2 \quad \text{или} \quad \hat{X}_1 < \hat{X}_2$$

Для выборок оптимального объема недостоверность разности означает принятие "*нулевой гипотезы*" о равенстве генеральных средних арифметических.

Для определения достоверности разности выборочных средних рассчитывают критерий достоверности разности  $td$  по формуле:

$$td = \frac{d}{md} = \frac{X_1 - X_2}{\sqrt{m_1^2 + m_2^2}}$$

где  $d$  - разность средних арифметических выборок (из большей средней вычитают меньшую);

$md$  - ошибка разности;

$m_1^2, m_2^2$  - квадраты ошибок соответствующих средних арифметических.

Полученные фактические значения  $td$  надо сравнить с табличным критерием  $t$ , взятым из таблицы стандартных значений критерия Стьюдента (см. Приложение 1) для числа степеней свободы  $-\gamma = n_1 + n_2 - 2$ , где  $n_1$  - количество вариант первой группы,  $n_2$  - количество вариант второй группы.

Если  $td \geq t_{\text{табл.}}$ , то разность достоверна с соответствующим уровнем надежности выводов ( $P_1 = 0,95$ ;  $P_2 = 0,99$ ;  $P_3 = 0,999$ ).

Если  $td < t_{\text{табл.}}$ , то разность недостоверна, тогда «нулевая гипотеза» не отвергается. Это значит:

а) по выборочной разности нельзя сделать никакой оценки генеральной разности;

б) осталось невыясненным, какая из двух генеральных средних больше;

в) осталось недоказанным как наличие, так и отсутствие различия между генеральными средними.

За минимальный порог достоверности в подавляющем большинстве исследований принимается первый порог, соответствующий вероятности безошибочных прогнозов  $P_1 = 0,95$ .

*Пример 1.* Сравнивали живую массу взрослых индеек двух пород после одинакового откорма. По выборкам получены следующие сводные показатели.

$$\begin{aligned}n_1 &= 20, & \bar{X}_1 \pm m_1 &= 4,0 \pm 0,3 \text{ кг.}, \\n_2 &= 25, & \bar{X}_2 \pm m_2 &= 4,6 \pm 0,4 \text{ кг.}, \\d &= \bar{X}_2 - \bar{X}_1 &= 0,6.\end{aligned}$$

Требуется определить достоверность полученной выборочной разности:

$$td = \frac{\bar{X}_2 - \bar{X}_1}{\sqrt{m_1^2 + m_2^2}} = \frac{0,6}{0,5} = 1,2$$

1) Полученные результаты ( $d \cdot 0,6$ ) свидетельствуют о том, что в некоторых группах индейки второй породы могут оказаться в среднем тяжелее индеек первой породы;

2) Так как полученная разность оказалась недостоверной, то ничего нельзя заключить о всех представителях обеих пород. Осталось невыясненным, какая порода (вся) может иметь больший средний вес. Нельзя считать доказанным, что разницы в среднем весе между породами нет и что эти породы в среднем одинаковы по весу.

*Пример 2.* В пригородном свиноводстве испытывали рационы с добавкой витаминов при беконном откорме подсвинков. Были образованы опытная (О) и контрольная (К) группы на 100 откормочников в каждой. В течении месяца опытная группа получала сверх обычного рациона комплекс витаминов. В конце месяца для каждого подсвинка определена прибавка в весе и затем высчитан средний суточный прирост по каждой группе.

Первичные данные:

$$\begin{aligned}n_k &= 100, & \bar{X}_k &= 700 \text{ г}, & \sigma_k &= 50 \text{ г}, & m_k &= 5 \text{ г}; \\n_o &= 100, & \bar{X}_o &= 800 \text{ г}, & \sigma_o &= 60 \text{ г}, & m_o &= 6 \text{ г}; \\d &= \bar{X}_o - \bar{X}_k &= 800 - 700 &= 100\end{aligned}$$

$$td = \frac{800 - 700}{\sqrt{5^2 + 6^2}} = \frac{100}{7,8} = 12,8$$

$$\gamma = 198 \quad t_{st} = (2,0-2,6-3,3)$$

Действие витаминов оказалось в высшей степени достоверным. Все поголовье совхоза переведено на витаминизированные рационы.

Аналогично рассчитывают достоверность разности между всеми другими выборочными параметрами. Например, достоверность разности между двумя выборочными коэффициентами корреляции определяют так:

$$td_r = \frac{d_r}{md_r} = \frac{r_1 - r_2}{\sqrt{m_{r_1}^2 + m_{r_2}^2}} \geq t_{табл} (v = n_1 + n_2 - 4)$$

Такая необходимость может возникнуть при сравнении разных пород, линий и т.д., но особенно часто – при оценке производителей по качеству потомства: сравнивая степень связи между секционированными признаками у матерей и у их потомков от данного производителя, можно определить степень наследственного влияния производителя на связь между признаками.

#### Достоверность разности по критерию Фишера

Определить достоверность разности можно по критерию Фишера, по формуле:

$$F = \frac{d^2}{\sigma_z^2} \cdot \frac{n_1 n_2}{n_1 + n_2} \geq F_{st} \left\{ \begin{array}{l} v_1 = 1 \\ v_2 = n_1 + n_2 - 2 \end{array} \right\},$$

где F- критерий Фишера,

$d^2$  – квадрат разности средних  $(\bar{X}_1 - \bar{X}_2)^2$ ,

$\sigma_z^2$  – случайная вариация, которая может быть определена на основе обеих дисперсий, или на основе обоих квадратов ошибок, сравниваемых средних по формулам :

$$\sigma_z^2 = \frac{C_1 + C_2}{n_1 + n_2 - 2} = \frac{(n_1 - 1) \cdot \sigma_1^2 + (n_2 - 1) \sigma_2^2}{n_1 + n_2 - 2} = \frac{n_1 (n_1 - 1) \cdot m_1^2 + n_2 (n_2 - 1) \cdot m_2^2}{n_1 + n_2 - 2}$$

$n_1, n_2$  – объемы обеих сравниваемых выборок;

$F_{st}$  – стандартные значения критерия Фишера при двух указанных степенях свободы (находят по таблице).

*Пример* :  $n_1 = 20 \quad \bar{X}_1 \pm m_1 = 4,0 \pm 0,3$

$n_2 = 25 \quad \bar{X}_2 \pm m_2 = 4,6 \pm 0,4$

$$\sigma_z^2 = \frac{20 \cdot 19 \cdot 0,09 + 25 \cdot 24 \cdot 0,16}{20 + 25 - 2} = 3,03$$

$$F = \frac{0,6^2}{3,03} \cdot \frac{20 \cdot 25}{20 + 25} = 1,3 \text{ не достоверно} \quad \begin{array}{l} v_1 = 1 \\ v_2 = 43 \end{array},$$

$F_{st} = (4,1-7,3-12,7)$ .

**Задание №4.** Достоверны ли различия в живой массе у овец гиссарской породы трех типов конституции при доверительных вероятностях 0,95 и 0,99 по следующим данным:

Тип конституции	Грубая	Крепкая	Нежная
Живая масса, кг	82,4±1,10	78,9±0,84	69,9±0,88

**Задание №5.** Каковы достоверность разницы и доверительные границы живой массы при P=0,95 у животных симментальского скота Мордовской АССР разного типа конституции?

Показатель	Широкотелый тип	Узкотелый тип
Живая масса, кг	601±9,2	551±7,3
Наивысший удой за лактацию, кг	4431±90,0	4150±103,0
Содержание жира в молоке, %	3,90±0,03	3,91±0,04

#### **2.4. ОПРЕДЕЛЕНИЕ ОБЪЕМА ВЫБОРОК**

Прежде чем проводить научный эксперимент необходимо правильно составить выборку, а также определить минимальный объем выборки, чтобы получить достаточно надежные, достоверные и объективные выводы.

Чем больше изучено особей в выборке, тем больше выборочное исследование приближается к сплошному, тем меньше ошибка репрезентативности.

Часто необоснованно одни исследователи для опыта берут очень мало животных, другие наоборот слишком много. Поэтому, прежде, чем начинать опыт необходимо рассчитать объем выборки. Минимальный объем выборки определяют по формуле:

$$n = \frac{t^2 \cdot \sigma^2}{\Delta^2} \quad \Delta = k \cdot \sigma \quad n = \frac{t^2 \cdot \sigma^2}{k^2 \cdot \sigma^2} = \frac{t^2}{k^2} = \left( \frac{t}{k} \right)^2, \text{ где}$$

t – табличный критерий Стьюдента

k – показатель точности

Δ – допустимая неточность

Показатель точности условно принимают для:

- ориентировочного ознакомления с популяцией = 0,3-0,5
- исследования средней точности = 0,1-0,3
- с повышенной точностью = 0,1

*Пример:* Требуется определить количество поголовья, необходимое для получения достоверной средней арифметической по содержанию жира в молоке коров. Допустимую неточность устанавливаем априорно 0,1% жира, стандартное отклонение (на основе предыдущих исследований) в генеральной совокупности равно 0,25% жира. Критерий достоверности при уровне значимости 0,05 равен 1,96. Тогда

$$k = \frac{0,1}{0,25} = 0,4 \qquad n = \left( \frac{1,96}{0,4} \right)^2 = 24 \text{ головы}$$

Занятие 6-7.

### **3. СТАТИСТИЧЕСКИЕ ВЗАИМОСВЯЗИ И СПОСОБЫ ВЫЧИСЛЕНИЯ ИХ ВЕЛИЧИН**

*Цель занятия:* ознакомиться с техникой нахождения основных параметров фенотипического коэффициента корреляции и рангового коэффициента корреляции Спирмена.

#### **3.1. КОРРЕЛЯЦИОННАЯ СВЯЗЬ И ВЫЧИСЛЕНИЕ КОЭФФИЦИЕНТА ФЕНОТИПИЧЕСКОЙ КОРРЕЛЯЦИИ**

В племенном деле при проведении отбора животных для спаривания очень большое значение имеет знание закономерностей соотносительной изменчивости и обязательный их учет в повседневной практической работе.

В биологии термин «корреляция» употребляется для обозначения взаимной связи между отдельными частями, признаками и свойствами организма, проявляющейся в том, что изменение одной части (или признака) ведет к соответствующему изменению другой, связанной с ней части (или признака).

Так, высота в холке в определенной мере связана с живой массой. Но такая связь не абсолютна, поскольку всегда найдутся особи, у которых, например, живая масса увеличена, а высота в холке нет. Довольно часто от обильномолочных матерей рождаются маломолочные дочери и т.д. Но, так как изучаются не отдельные особи, а целые их группы (совокупности), то эти индивидуальные отличия сглаживаются и отчетливо становится видно, что в массе закономерная связь существует.

Обнаружить наличие или отсутствие корреляционной связи между какими-либо признаками можно на основании обследования большой

группы животных, в результате обработки полученных цифр определенным биометрическим методом.

Для оценки связи между биологическими признаками и свойствами используется **коэффициент корреляции** ( $r$ ). Он является величиной неименованной и показывает **направление и силу** корреляционной связи.

**Силу связи** показывает абсолютная величина « $r$ », а ее **направление** – знак перед ним. Наличие знака «+» означает, что между признаками существует **положительная корреляция**, когда при увеличении одного признака у особей совокупности другой признак также возрастает или при уменьшении одного признака другой также снижается. Следовательно, изменение признаков происходит в одном направлении. Например, с увеличением живой массы коров матерей (до определенного оптимума) увеличивается и живая масса получаемых от них телят и т. п. Знак «-» указывает на наличие между признаками **отрицательной (обратной), разнонаправленной связи**, при которой увеличение одного признака сопровождается уменьшением другого (например, с увеличением удоя жирномолочность уменьшается). Значения коэффициента корреляции варьируют в пределах от +1 до -1.

Различают слабую, среднюю и высокую корреляционную зависимость (связь):

- если  $r > 0,7$ , то связь сильная (тесная, высокая);
- если  $r < 0,3$  – слабая (низкая);
- если  $r = 0,3+0,7$  – средняя.

Если  $r = 0$ , то связь между признаками криволинейная, т.е. с увеличением одного признака, другой увеличивается до определенного момента, а затем уменьшается или наоборот. Например, с увеличением возраста удои увеличивается до 6-7 отела, а затем у большинства коров снижается.

При криволинейной корреляции связь сначала положительная, затем отрицательная.

### **Техника расчета коэффициента корреляции в больших выборках.**

Для определения коэффициента корреляции в большой выборке сначала строят корреляционную решетку. Затем корреляционную решетку обрабатывают и находят коэффициент корреляции по формуле:

$$r = \frac{\sum \sum f a_x a_y - n \cdot \beta_x \cdot \beta_y}{n \cdot S_x \cdot S_y}$$

$$\beta = \frac{\sum f a}{n}$$

Порядок работы рассмотрим на примере:

$$S = \sqrt{\frac{\sum f a^2}{n} - \beta^2}$$

*Пример 1* Вычислить коэффициент корреляции (r) между живой массой и обхватом груди у коров ярославской породы, по приведенным данным.

Живая масса и обхват груди у коров ярославской породы

Живая масса, кг	Обхват груди, см	Живая масса, кг	Обхват груди, см	Живая масса, кг	Обхват груди, см
489	184	440	185	445	180
467	186	524	194	420	179
462	185	447	179	491	191
441	182	430	179	450	181
473	186	485	187	406	172
491	190	440	173	417	174
545	196	488	187	434	177
433	183	439	180	432	177
488	191	445	180	505	193
539	196	504	191	534	195
440	182	550	203	473	188
475	186	495	190	441	179
411	178	536	192	556	197
488	189	426	186	486	185
426	177	388	169	535	196
390	172	407	176	460	183
482	191	425	184	469	180
391	174	390	172	421	180
470	185	418	179	520	195
421	180	465	189	445	183
429	176	391	174	384	173
439	180	365	167	488	182
442	183	383	172	500	190
490	189	427	183	432	178
426	177	448	186	475	185

1. Произвольно обозначим живую массу через X, а обхват груди –У. Построить вариационный ряд по каждому признаку.

Живая масса -X	Обхват груди – У
min = 365 кг, max = 556 кг	min = 167 см, max = 203 см
$K = \frac{556-365}{8} = 23,9 \approx 24 \text{ кг}$	$K = \frac{203-167}{8} = 4,5 \approx 5 \text{ см}$

6. Строим корреляционную решетку. Произвести разноску вариант одновременно по двум признакам. Например, первая корова имеет жи-

вую массу 489 кг и обхват груди 184 см. По живой массе она должна быть отнесена в класс 485-508, а по обхвату груди – в класс 182-186. Животное с этими показателями помещают в ячейку, находящуюся на пересечении указанных классов по живой массе и обхвату груди.

7. Подсчитать число частот в каждом классе, как по горизонтали ( $f_y$ ), так и по вертикали ( $f_x$ ).

8. Выделить в каждом вариационном ряду условный средний класс.

9. Вычислить отклонение от модального класса ( $a_x, a_y$ ), произведения частот на отклонение ( $f_x a_x, f_y a_y$ ) и частот на квадраты отклонений ( $f_x a_x^2, f_y a_y^2$ ).

10. Условные классы разделили корреляционную решетку на четыре квадранта ( I, II, III, IV).

Вычислить  $\sum f a_x a_y$  для каждого квадранта. Для этого частоту каждого квадранта умножают на отклонение  $a_x, a_y$ . Частоты в условных средних не учитываем.

I квадрант	II квадрант	III квадрант	IV квадрант
$\sum f a_x a_y$	$\sum f a_x a_y = 0$	$\sum f a_x a_y$	$\sum f a_x a_y$
$2 \cdot (-2) \cdot (-3) = 12$		$1 \cdot 1 \cdot (-2) = -2$	$3 \cdot 2 \cdot 1 = 6$
$2 \cdot (-2) \cdot (-2) = 8$		$7 \cdot 1 \cdot (-1) = -7$	$10 \cdot 3 \cdot 1 = 30$
$6 \cdot (-1) \cdot (-2) = 12$		$1 \cdot 2 \cdot (-1) = -2$	$1 \cdot 3 \cdot 2 = 6$
$1 \cdot (-1) \cdot (-1) = 1$			$2 \cdot 4 \cdot 2 = 16$
			$5 \cdot 5 \cdot 2 = 50$
			$1 \cdot 5 \cdot 3 = 15$
			$1 \cdot 5 \cdot 4 = 20$
$\sum f a_x a_y = 33$		$\sum f a_x a_y = -11$	$\sum f a_x a_y = 143$

Найти сумму сумм квадрантов:

$$\sum \sum f a_x a_y = 33 + 0 + (-11) + 143 = 165$$

11. Подставляем значения в формулы:

$$\beta_x = \frac{\sum f_x a_x}{n} = \frac{105}{75} = 1,4$$

$$\beta_y = \frac{\sum f_y a_y}{n} = \frac{-11}{75} = -0,147$$

$$S_x = \sqrt{\frac{\sum f_x a_x^2}{n} - \beta_x^2} = \sqrt{\frac{411}{75} - 1,4^2} = \sqrt{5,48 - 1,96} = \sqrt{3,52} = 1,88$$

$$S_y = \sqrt{\frac{\sum f_y a_y^2}{n} - \beta_y^2} = \sqrt{\frac{151}{75} - (-0,147)^2} = \sqrt{2,01 - 0,02} = \sqrt{1,99} = 1,41$$

$$r = \frac{\sum \sum f a_x a_y - n \cdot \beta_x \beta_y}{n \cdot S_x \cdot S_y} = \frac{165 - 75 \cdot 1,4 \cdot (-0,147)}{75 \cdot 1,88 \cdot 1,41} = \frac{165 - (-15,44)}{30 \cdot 198,81} = +0,91$$

Корреляционная решетка												
Y - охват X - живая масса	167-171	172- 176	177- 181	182 - 186	187- 191	192- 196	197- 201	202- 206	$f_x$	$a_x$	$f_x a_x$	$f_x a_x^2$
365-388	. . 2	. 2							4	-2	-8	16
389-412		L <sub>6</sub>	.						7	-1	-7	7
413-436		. 2	⊗ <sub>10</sub>	∴ <sub>4</sub>					16	0	0	0
437-460		.	U <sub>7</sub>	U <sub>7</sub>					15	1	15	15
461-484			. 1	L <sub>6</sub>	∴ <sub>3</sub>				10	2	20	40
485-508				∴ <sub>3</sub>	⊗ <sub>10</sub>	. 1			14	3	42	126
509-532						. 2			2	4	8	32
533-556						L <sub>5</sub>	. 1	. 1	7	5	35	175
$f_y$	2	11	19	20	13	8	1	1				
$a_y$	-3	-2	-1	0	1	2	3	4	*			
$f_y a_y$	-6	-22	-19	0	13	16	3	4	$\Sigma f_y a_y = -11$		$\Sigma f_x a_x = 105$	
$f_y a_y^2$	18	44	19	0	13	32	9	16	$\Sigma f_y a_y^2 = 151$		$\Sigma f_x a_x^2 = 411$	

**ВЫВОД:** Между живой массой и обхватом груди сильная, положительная коррелятивная связь. С увеличением живой массы у коров увеличивается обхват груди.

### ***Достоверность выборочного коэффициента корреляции***

Критерий достоверности выборочного коэффициента корреляции определяется по формуле:

$$t_r = \frac{r}{m_r} \geq t_{st} (v=n-2),$$

При  $t_r \geq t_{st}$  коэффициент корреляции достоверен.

В этом случае с определенной вероятностью можно считать, что между коррелируемыми признаками имеется связь и в генеральной совокупности такая же по знаку, какая получилась в выборке (прямая или обратная).

При  $t_r < t_{st}$  выборочный коэффициент корреляции недостоверен, что не дает возможности сделать какое-либо заключение о связи признаков в генеральной совокупности. Для выяснения этого вопроса требуется провести повторные исследования на более многочисленном материале.

*Пример:* При проверке гипотезы о связи крупноплодности с жирномолочностью был рассчитан коэффициент корреляции между процентом жира в молоке у 50 коров и весом при рождении телят от этих же коров.

Получено:

коэффициент корреляции  $r = \pm 0,15$ ;

его ошибка  $m_r = \sqrt{\frac{1-0,15^2}{50-2}} = 0,14$ ,

критерий достоверности:

$$t_r = \frac{0,15}{0,14} = 1,0; v = 48; t_{st} (1,7-2,0-2,7-3,5).$$

Выборочный коэффициент оказался явно недостоверным. На основе проведенного исследования нельзя ожидать связи между крупноплодностью и жирномолочностью.

### **Задание №6**

*Пример:* Вычислить коэффициенты корреляции между признаками: суточный удой и скорость молокоотдачи, суточный удой и индекс вымени у коров черно-пестрой и костромской пород.

№ п/п	Черно-пестрая порода			Костромская порода		
	Суточный удой , кг	Скорость молокоот- дачи, кг- мин	Индекс вымени, %	Суточный удой , кг	Скорость молокоот- дачи, кг- мин	Индекс вымени, %
1	2	3	4	5	6	7
1	19	2,03	44	17	1,80	42
2	18	2,15	42	15	1,67	43
3	14	2,10	43	12	1,35	37
4	17	1,45	39	10	1,20	38
5	22	1,80	40	15	1,88	42
6	17	1,95	42	14	1,60	40
7	16	2,02	44	18	2,3	46
8	15	1,33	37	20	2,17	42
9	19	2,09	45	14	2,15	45
10	20	1,77	42	13	1,18	36
11	14	1,80	44	17	1,45	40
12	12	1,55	42	15	1,82	42
13	17	2,15	46	16	1,70	41
14	18	1,60	41	17	1,44	40
15	21	2,10	44	19	2,15	45
16	18	1,75	40	21	2,10	42
17	14	1,60	38	11	1,45	44
18	15	2,12	43	12	1,18	43
19	17	1,85	42	14	1,75	40
20	18	2,05	44	16	1,50	39
21	17	2,03	45	12	1,33	37
22	15	1,25	37	14	1,28	42
23	16	1,39	39	15	1,25	44
24	20	2,05	42	18	2,11	39
25	18	1,99	41	14	2,03	42
26	19	2,13	45	17	1,52	43
27	21	2,21	45	15	1,66	40
28	17	1,95	41	12	1,40	38
29	19	1,47	43	14	1,28	42
30	16	1,95	42	19	2,18	45
31	14	2,03	45	21	2,35	43
32	20	2,31	42	14	1,56	40
33	15	1,40	39	15	1,65	37
34	18	1,58	40	13	1,22	35
35	17	2,2	44	18	1,50	42
36	14	1,95	43	16	1,95	41
37	16	1,44	40	14	1,60	39
38	18	1,90	39	17	1,83	42
39	16	1,45	42	15	1,25	40
40	17	2,22	44	15	1,48	37
41	22	1,75	40	17	1,85	39
42	20	2,20	44	18	2,23	44
43	21	2,35	47	11	1,20	35
44	17	2,09	44	13	1,55	39
45	15	1,84	39	12	1,30	37
46	19	2,16	44	14	1,27	40
47	14	2,03	42	22	2,11	38
48	12	1,33	40	14	1,45	41
49	15	1,40	37	14	1,70	40
50	19	2,17	44	13	1,55	42

### 3.2 ВЫЧИСЛЕНИЕ РАНГОВОГО КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ СПИРМЕНА ( $r_s$ )

Есть признаки, которые нельзя или не требуется измерять, но по которым особи могут быть расположены по возрастающей или убывающей степени выраженности признака.

Ряд объектов, расположенных по степени выраженности значения признака, называется *ранжированным рядом*, процесс составления краткого ряда называется *ранжированием*, порядковый номер особи в ранжированном ряду называют *рангом* этой особи. Входящие в выборку варианты обязательно ранжируют сначала по одному, а затем по другому признаку.

Если в группе имеются особи, неразличимые между собой по изучаемому порядковому признаку, то каждой паре или тройке, или четверке и т.д. таких признаков присуждается средний ранг, равный средней арифметической из тех рангов, какие имели бы эти особи, если бы были различимы. Например, имеется ряд ранжированных значений какого-то признака: 2, 4, 4, 4, 5, 6, 6, 8, 9, 10 (табл. ). Если бы цифры 4 и 6 не повторялись, то ранги этих значений были бы соответственно: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. Но в этом ряду цифра 4, имеющая одинаковый ранг, повторяется трижды, а цифра 6 – дважды. Их ранг, независимо от месторасположения в ранжированном ряду, одинаковый. Он равен средней арифметической величине из соответствующих им порядковых чисел:

а) для цифры 4, занимающей в ранжированном ряду 2-е, 3-е, и 4-е место, ранг равен  $(2+3+4) : 3 = 3$ ;

б) для цифры 6 ранг равен  $(6+7) : 2 = 6,5$

#### ОПРЕДЕЛЕНИЕ РАНГОВ

Номер объекта	1	2	3	4	5	6	7	8	9	10
Значение признака	2	4	4	4	5	6	6	8	9	10
Ранг значений	1	3	3	3	5	6,5	6,5	8	9	10

Формула рангового коэффициента корреляции:

$$r_s = 1 - \frac{6 \sum (x - y)^2}{n \cdot (n^2 - 1)}$$

где  $x$  и  $y$  – разность между рангами сопряженных рядов, а  $n$  – число пар значений признаков.

Абсолютное значение рангового коэффициента корреляции изменяется от 0 до 1 и выражается десятичной дробью со знаком «+» (положительная корреляция) или «-» (отрицательная корреляция).

*Пример:* При изучении зависимости густоты спермы от степени упитанности отобрано 7 быков- производителей. Они ранжированы: по упитанности – от самой низкой до ожиревшего (ранги 1-7), по густоте спермы – от очень плохого до самого лучшего (ранги 1-7). По данным ранжированных рядов вычислить коэффициент ранговой корреляции.

Производители	А	Б	В	Г	Д	Е	Ж	
Упитанность (x)	1	2	3	4	5	6	7	
Густота спермы (y)	3	5	7	6	4	1	2	
x – y	-2	-3	-4	-2	1	5	5	
(x – y) <sup>2</sup>	4	9	16	4	1	25	25	$\sum(x-y)^2=84$

$$r_s = 1 - \frac{6\sum(x-y)^2}{n \cdot (n^2 - 1)} = 1 - \frac{6 \cdot 84}{7 \cdot (7^2 - 1)} = 1 - \frac{504}{336} = 1 - 1,5 = -0,5$$

**Вывод:** Между упитанностью производителей и густотой продуцированной ими спермы существует отрицательная зависимость. Следовательно, быки высокой упитанности дают сперму худшего качества, чем быки средней упитанности.

*Пример:* Определить направление и силу связи между балльной оценкой экстерьера и длиной туловища свиней крупной белой породы по следующим исходным данным.

#### Расчет коэффициента ранговой корреляции

Но- мер объ- екта	Абсолютное значение при- знака		Ранг по при- знаку		d = x-y	d <sup>2</sup>	Вспомогательная таблица по расчету рангов второго при- знака y		
	x	y	x	y			№ объ- екта	Абсолют- ные зна- чения	Ранг
1.	81	156	1	7	-6	36	1	148	1
2.	84	168	2	15,5	13,5	182,25	2	152	2
3.	86	160	3	11,5	-8,5	72,25	3	155	4
4.	89	155	5	4	+1	1	4	155	4
5.	89	156	5	7	-2	4	5	155	4
6.	89	156	5	7	-2	4	6	156	7
7.	90	168	7,5	15,5	-8	64	7	156	7
8.	90	172	7,5	17	-9,5	90,25	8	156	7
9.	91	157	9,5	9,5	0	0	9	157	9,5
10.	91	160	9,5	11,5	-2	4	10	157	9,5
11.	92	148	11,5	1	+10,5	110,25	11	160	11,5

12.	92	157	11,5	9,5	+2	4	12	160	11,5
13.	93	161	13,5	13	+0,5	0,25	13	161	13
14.	93	155	13,5	4	+9,5	90,25	14	162	14
15.	94	152	15	2	+13	169	15	168	15,5
16.	95	155	16	4	+12	144	16	168	15,5
17.	96	162	17	14	+3	9	17	172	17

Определяем

$$r_s = 1 - \frac{6 \sum d^2}{n \cdot (n^2 - 1)} = 1 - \frac{6 \cdot 984,5}{17 \cdot (17^2 - 1)} = 1 - \frac{5907}{4896} = 1 - 1,206 = -0,21$$

Вывод : Между балльной оценкой экстерьера и длиной туловища свиней крупной белой породы установлена отрицательная зависимость. Следовательно, с увеличением длины туловища свиней снижается их балльная оценка экстерьера.

**Задание №7** . При изучении связи качества меха норок с агрессивностью поведения отобраны 9 самцов, которые занимают ранги от самого спокойного, т.е. слабой агрессивности (1), до сильного (9). По качеству меха животные расположены также по рангу от худшего (1) до лучшего (9). Можно ли вести селекцию по поведению норок с целью повышения экономической эффективности? Какова достоверность полученной величины коэффициента ранговой корреляции?

Норки .....	А	Б	В	Г	Д	Е	Ж	З	И
Ранги: по агрессивности	1	2	3	4	5	6	7	8	9
по качеству меха	3	1	2	5	4	8	9	6	7

**Задание №8.** Приводятся ранжированные ряды густоты шерсти овец от меньшей (1) до большей (7) и упитанности от низшей (1) до высшей (7). Что можно сказать о характере взаимосвязи между густотой шерсти и упитанностью овец?

Номера овец.....	20	21	22	23	24	25	26
Ранги: По густоте шерсти	1	2	3	4	5	6	7
По упитанности	2	3	4	1	7	6	5

Занятие 8.

### **3.3 КОЭФФИЦИЕНТ РЕГРЕССИИ (R) И ЕГО ИСПОЛЬЗОВАНИЕ В ЗООВЕТЕРИНАРНОЙ РАБОТЕ**

Цель занятия: получить представление о регрессии, способах ее определения, построения номограмм.

Коэффициент корреляции дает представление о степени и направлении связи, но не показывает насколько в среднем изменяется один из признаков, если другой, связанный с ним изменяется на единицу.

Ответ на этот вопрос дает *коэффициент регрессии* (R), который вычисляют по формуле:

$$R_{x/y} = r \cdot \frac{\sigma_x}{\sigma_y}$$

Коэффициент регрессии показывает насколько в среднем изменится признак X, если признак Y изменится на единицу.

$$R_{y/x} = r \cdot \frac{\sigma_y}{\sigma_x}$$

Коэффициент регрессии показывает насколько в среднем изменится признак Y, если признак X изменится на единицу.

Коэффициент регрессии имеет тот же знак (+ или -), что и коэффициент корреляции, но в отличие от него всегда является величиной именованной и выражается в тех же единицах в каких измеряется признак.

*Пример:* Рассчитать коэффициенты регрессии живой массы по обхвату груди и обхват груди по живой массе свиноматок крупной белой породы, если коэффициент корреляции между живой массой – X и обхватом груди – Y составит 0,7, а среднее квадратическое отклонение  $\sigma_x = 29,4$  кг,  $\sigma_y = 10,6$  см.

$$R_{x/y} = 0,7 \cdot \frac{29,4}{10,6} = 1,94 \text{ кг.}$$

$$R_{y/x} = 0,7 \cdot \frac{10,6}{29,4} = 0,25 \text{ см.}$$

В практической работе обычно вычисляют только одно из двух значений коэффициента регрессии, которое имеет реальный смысл. Например, в рассматриваемом нами примере определенный хозяйственный интерес представляет знание того, как по промеру обхвата груди можно судить о живой массе свиноматок.

Коэффициент  $R_{x/y}$  здесь означает, что если из совокупности отобрать свиноматок с обхватом груди на 1 см больше (или меньше) среднего, то их живая масса будет в среднем на 1,94 кг больше (или соответственно меньше), чем средняя живая масса свиноматок совокупности».

Для описания регрессии можно использовать уравнение прямолинейной регрессии, которое выражается формулой:

$$Y = a + b \cdot X$$

$$\left\{ \begin{array}{l} a = \bar{X}_y - R_{y/x} \cdot \bar{X}_x \end{array} \right.$$

$$b = R_{y/x}$$

По этой формуле, зная значение «X», можно определить значение "Y" без непосредственного его измерения.

$$\begin{cases} X = a + b \cdot y \\ a = \bar{X}_x - R_{x/y} \cdot \bar{X}_y \\ b = R_{x/y} \end{cases}$$

По этой формуле, зная значение «y», можно определить значение "X".

Пример: Для разработки способа определения живой массы лошадей по обхвату груди (без взвешивания) было взвешено 1618 лошадей и у каждой из них измерен обхват груди. Получены следующие показатели:

X-обхват груди;  $n = 1618$ ;  $\bar{X}_x = 174$  см;  $\sigma_x = 7,9$  см.

Y – живая масса;  $n = 1618$ ;  $\bar{X}_y = 424$  кг,  $\sigma_y = 56,8$  кг.

$r = +0,89$

Коэффициент регрессии живой массы по обхвату груди равен:

$$R_{y/x} = \frac{56,8}{7,9} \cdot (+0,89) = +6,4.$$

Таким образом, можно ожидать, что при увеличении (или уменьшении) обхвата груди на 1 см, живая масса лошадей возрастает (или уменьшается) в среднем на 6,4 кг.

Уравнение регрессии для этого примера определяется следующим образом:

$$a = \bar{X}_y - R_{y/x} \cdot \bar{X}_x = 424 - (+6,4) \cdot 174 = -690$$

$$b = R_{y/x} = +6,4$$

$$Y = a + b \cdot x = -690 + 6,4x;$$

$$Y = 6,4x - 690$$

Следовательно, чтобы определить живую массу лошади без взвешивания, по этому способу надо обхват груди умножить на постоянный коэффициент 6,4 и из полученного произведения вычесть число -690.

На основе уравнений прямолинейной регрессии можно заранее рассчитать значение Y для каждого значения X.

Таблица для определения живой массы лошадей по обхвату груди может иметь простую форму:

X - обхват груди (см)	140	150	160	170	180	190	200	210	220	230	240
Y - вес лошади (кг)	206	270	334	398	462	526	590	654	718	782	846

Например, при обхвате груди  $X = 140$  см, живая масса лошади равна 206 кг.

$$Y = 6,4 \cdot 140 - 690$$

$$Y = 206 \text{ кг.}$$

Если цифры такой таблицы нанести на график, на котором по оси абсцисс отложить через равные интервалы значения  $X$  - обхвата, а по оси ординат - значения  $Y$  - живой массы, то получится номограмма для определения веса лошадей без взвешивания и без вычисления.

#### Занятие 9.

### 4. РАСЧЕТ ОСНОВНЫХ СТАТИСТИЧЕСКИХ ПАРАМЕТРОВ В МАЛОЙ ВЫБОРКЕ

Цель занятия: ознакомиться с техникой нахождения основных статистических параметров в малых выборках.

Если в обработке используются данные меньше чем по 30 животных, обработку материалов проводят по малым выборкам

*Вычисление средней арифметической  $\bar{X}$*

Вычисляют  $\bar{X}$  путем суммирования значений всех вариантов ( $X_1 + X_2 + X_3 + \dots + X_n$ ) с последующим делением суммы на количество вариантов

$$\bar{X} = \frac{\sum X}{n},$$

где  $\bar{X}$  - средняя арифметическая,

$\sum X$  - сумма значений вариантов,

$n$  - объем выборки (количество вариантов).

Например, необходимо определить среднесуточный удой в группе из шести коров, если удой каждой в отдельности составил : 12,3; 10,7; 9,5; 14,5; 13,5; 15,0 (кг):

$$\bar{X} = \frac{12,3 + 10,7 + 9,5 + 14,5 + 13,5 + 15,0}{6} = \frac{75,5}{6} = 12,58 \text{ кг}$$

*Вычисление средней взвешенной  $\bar{X}_{взв.}$ .*

Средняя взвешенная представляет собой результат усреднения средних арифметических нескольких совокупностей и вычисляется по формуле:

$$\bar{X}_{взв.} = \frac{\sum \bar{X} \cdot n}{\sum n} = \frac{\bar{X}_1 \cdot n_1 + \bar{X}_2 \cdot n_2 + \dots + \bar{X}_n \cdot n_n}{n_1 + n_2 + \dots + n_n}$$

Где  $\bar{X}_1, \bar{X}_2, \bar{X}_n$  - средние арифметические отдельных совокупностей,

$n_1, n_2, n_n$  - объем совокупностей.

***Вычисление среднего квадратического отклонения ( $\sigma$ )***

В малочисленных выборках сигму можно рассчитать, используя формулу:

$$\sigma = \pm \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

*Пример:* Обработать данные по глубине груди (см) 10 свиноматок.  
 Рассчитать  $\bar{X} \pm m_x$ ;  $\sigma$

Глубина груди, см. X	45	46	46	48	50	51	55	57	57	60
Отклонения $x - \bar{X}$	-6,5	-5,5	-5,5	-3,5	-1,5	-0,5	3,5	5,5	5,5	8,5
Квадраты отклонений $(x - \bar{X})^2$	42,25	30,25	30,25	12,25	2,25	0,25	12,25	30,25	30,25	72,2

$$\bar{X} = \frac{515}{10} = 51,5 \quad \sum X = 515 \quad \sum (x - \bar{X})^2 = 262,5$$

$$m_x = \frac{\sigma}{\sqrt{n-1}} = \frac{5,4}{3} = 1,8 \text{ см} \quad \sigma = \pm \sqrt{\frac{262,5}{9}} = \pm 5,4 \quad \bar{X} \pm m_x = 51,5 \pm 1,8 \text{ см.}$$

**Вычисление коэффициента корреляции  
и коэффициента регрессии**

Коэффициент корреляции определяется по формуле:

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{(\sum x^2 - \frac{(\sum x)^2}{n}) \cdot (\sum y^2 - \frac{(\sum y)^2}{n})}}$$

Коэффициент регрессии:

$$R_{y/x} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} \quad R_{x/y} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum y^2 - \frac{(\sum y)^2}{n}}$$

*Пример 2.* Оценить связь удоя за лактацию у матерей (X) и дочерей (Y). (см. таблицу)

Лактационный удой 10 пар мать- дочь				
X	Y	X <sup>2</sup>	Y <sup>2</sup>	X•Y
31,0	33,2	961,00	1102,24	1029,30
32,5	33,4	1056,25	1115,56	1079,00

30,5	33,0	930,25	1089,00	1006,50
37,5	34,1	1406,25	1162,81	1278,75
36,0	34,0	1296,00	1156,00	1224,00
35,0	33,9	1225,00	1149,21	1186,50
34,8	33,9	1211,04	1149,21	1179,72
37,2	34,3	1383,84	1176,49	1275,96
32,4	33,4	1049,76	1115,56	1082,16
30,5	32,9	930,25	1082,41	1003,45
$\Sigma 337,4$	336,2	114496,40	113032,49	113542,4

По данным таблицы можно оценить коэффициент корреляции между молочной продуктивностью матерей и дочерей:

$$r = \frac{113542,4 - \frac{336,2 \cdot 337,4}{10}}{\sqrt{(114496,40 - \frac{337,4^2}{10}) \cdot (113032,49 - \frac{336,2^2}{10})}} = 0,98$$

Можно сделать заключение, что молочная продуктивность дочерей в большой степени зависит от уровня удоя матерей.

Материал таблицы позволяет рассчитать и коэффициент регрессии:

$$R_{y/x} = \frac{113542,4 - \frac{336,2 \cdot 337,4}{10}}{114496,4 - \frac{337,4^2}{10}} = \frac{118602}{657640} = 0,18$$

Полученные результаты свидетельствуют, что при повышении удоя матерей на 100 кг средняя молочная продуктивность дочерей увеличивается на 18 кг. Другими словами, коэффициент регрессии дочерей на матерей указывает на степень эффективности массовой селекции по отбору матерей.

**Задание №9.** Пример: вычислить среднюю арифметическую, среднеквадратическое отклонение, коэффициент вариации, ошибки репрезентативности, достоверность выборочных параметров, указать лимиты используя данные о живой массе телят при рождении (кг):

28 30 32 36 34 38 36 38 36 32

**Задание №10.**

Пример: вычислить среднюю арифметическую, среднеквадратическое отклонение, их ошибки и показатель достоверности разности по живой массе пород крупная белая и ландрас при рождении по следующим данным:

Крупная белая	1,2	1	1,1	1,4	0,9	1,2	1,2	1	1,1	0,8	1,1	1,2	0,9	1,3
Ландрас	1,1	1,1	0,8	0,9	1	1	1	1,1	1,2	0,9	1	1,1	1,1	0,9

**Задание №11.** Пример: вычислить коэффициент корреляции и регрессии между живой массой (x) и настригом шерсти (y) у 10 овец по следующим данным:

X, кг	70	71	72	74	75	70	80	85	75	80
Y, %	4,0	4,5	5,0	5,5	6,0	5,0	8,5	8,0	7,0	7,5

Построить номограмму для определения настрига шерсти овец по их живой массе.

Занятие 10.

## 5 СТАТИСТИЧЕСКИЙ АНАЛИЗ НАСЛЕДУЕМОСТИ И ПОВТОРЯЕМОСТИ

Цель занятия: ознакомиться с техникой нахождения коэффициентов наследуемости и повторяемости.

Развитие и степень выраженности отдельных полезных для хозяйства и биологических признаков у животных зависит от наследственной природы и конкретных условий среды обитания.

В практике животноводства известно, что изменчивость одних признаков в большей мере зависит от условий внешней среды (удой, упитанность, живая масса и др.), а других в меньшей мере (экстерьер, конституция, жирномолочность, масть и др.). Следовательно, одни признаки в меньшей степени подвержены наследственной детерминации, а другие - в большей степени.

В селекционной работе важно знать, насколько стойко, даже при некоторых изменениях внешней среды, будут передаваться желательные особенности родителей их потомству. Если величина признака стойко передается по наследству, то такой признак легко селекционируется. Отбор можно вести по фенотипу, т.е. осуществлять массовую селекцию.

Для определения статистической доли влияния наследственности на величину развития признака вычисляют *коэффициент наследуемости*.

### 5.1. КОЭФФИЦИЕНТ НАСЛЕДУЕМОСТИ И СПОСОБЫ ЕГО ВЫЧИСЛЕНИЯ ( $h^2$ )

Коэффициент наследуемости ( $h^2$ ) характеризует степень генетического разнообразия особей в популяции, какая доля общего разнообразия признака обусловлена наследственностью.

Коэффициент наследуемости колеблется от 0 до 1. Он является величиной неименованной.

Чем выше наследуемость, тем в большей степени разнообразие признака в популяции обусловлено наследственными различиями, и тем устойчивее этот признак к воздействиям внешней среды. Чем ниже  $h^2$ , тем сильнее влияние паратипических факторов и тем труднее правильно оценить по фенотипу наследственные особенности животных при отборе.

Коэффициент наследуемости вычисляют разными способами в зависимости от используемой первоосновы.

Формула 
$$h^2 = 2 \cdot r_{m/d}$$

применяется, если признак учитывается у животных одного пола (удой, яйценоскость, качество спермы и др.); коэффициент корреляции используется между признаками матерей и дочерей.

Формула

$$h^2 = 4 \cdot r_{m/c}$$

используется, когда признаки матерей неизвестны, а имеются только сведения об отцах; коэффициент корреляции вычисляют между полусибсами (полубратья и полусестры).

При неодинаковой степени отбора среди родителей и детей показатели разнообразия признака резко различаются. Тогда целесообразно пользоваться формулой

$$h^2 = 2 \cdot R_{d/m}$$

Преимущество формул с использованием коэффициента прямолинейной регрессии в том, что  $h^2$  не реагирует на изменение всех видов генетической информации, получаемой от родителей.

Часто при использовании формул с удвоенным и учетверенным коэффициентами корреляции получают биологически абсурдные значения коэффициента наследуемости - отрицательные или превышающие единицу.

Причины получения абсурдных величин  $h^2$  могут быть следующими:

- недостаточный объем выборки;
- взаимодействие генотипа со средой, что приводит к отрицательному коэффициенту корреляции между родственниками;
- межallelельные взаимодействия генов, доминирование, сверхдоминирование, эпистаз, оказывающие более сильное влияние на признак, чем аддитивное действие генов;
- популяция родителей не является случайной выборкой.

Рассмотрим примеры вычисления коэффициентов наследуемости.

*Пример 1.* В хозяйстве на маточном поголовье крупного рогатого скота вычислена теснота зависимости между удоями коров-матерей и

коров-дочерей, которая составила  $r = 0,13$ . Следовательно, коэффициент наследуемости:

$$h^2 = 2 \cdot r_{м/д} = 2 \cdot 0,13 = 0,26 \text{ (наследуемость удоя низкая).}$$

По показателю жирномолочности коэффициент корреляции  $r = 0,13$ . Следовательно, коэффициент наследуемости  $h^2 = 0,54$  (наследуемость средняя).

Коэффициент корреляции между живой массой родителей и потомков равнялся  $0,15$ . Этот признак учитывается у обоих родителей, тогда коэффициент наследуемости

$$h^2 = r_{р/п} = 0,15$$

Имеются другие методы вычисления коэффициента наследуемости, устраняющие получение его абсурдных показателей.

Широко используется формула, основанная на явлении регрессии, где учитываются средние показатели продуктивности лучших и худших животных – дочерей и матерей.

$$h^2 = \frac{\overline{X}_{\text{ДОЧЕРИ ЛУЧШИЕ}} - \overline{X}_{\text{ДОЧЕРИ ХУДШИЕ}}}{\overline{X}_{\text{МАТЕРИ ЛУЧШИЕ}} - \overline{X}_{\text{МАТЕРИ ХУДШИЕ}}} \cdot 2$$

где  $\overline{X}$  - средняя продуктивность группы лучших или худших дочерей и матерей.

Например, удои группы лучших молодых коров, полученных от высокоудойных матерей и хозяйстве, составил в среднем 3790 кг, удои их матерей - 4405 кг. Дочери худших коров имели средний удои 3600 кг, а их матери 2944 кг. Подставив данные в формулу, получим:

$$h^2 = \frac{3790 - 3600}{4405 - 2944} \cdot 2 = 0,26$$

Коэффициент наследуемости можно вычислить также по формуле:

$$h^2 = \frac{d_{\text{ПОТОМКОВ}}}{d_{\text{РОДИТЕЛЕЙ}}}$$

где  $d_{\text{ПОТОМКОВ}}$  - среднее превосходство потомков по данному признаку над всеми средними показателями стада;

$d_{\text{РОДИТЕЛЕЙ}}$  - среднее превосходство обоих родителей над всеми средними показателями стада.

*Пример 2.* Средняя живая масса свиней крупной белой породы в возрасте одного года по свиноферме хозяйства составила 148 кг. Живая масса лучших молодых животных равнялась 160 кг, а средняя живая масса их родителей в этом возрасте – 188 кг.

$$\text{Отсюда } h^2 = \frac{160 - 148}{188 - 148} = \frac{12}{40} = 0,3$$

В селекционной работе коэффициент наследуемости применяется прежде всего для выбора методов селекции и прогноза эффективности селекции. При низкой ( $h^2 < 0,4$ ) наследуемости следует проводить ин-

дивидуальный отбор, используя метод оценки производителей по качеству потомства. Если наследуемость признака высокая, можно проводить массовый отбор. При этом влияние внешних факторов на изменение величины признаков менее существенно.

Однако генотипическая структура популяции и особенности средовых факторов оказывают влияние на величину  $h^2$ . Поэтому по одним и тем же признакам  $h^2$  в разных стадах может сильно изменяться. Это следует учитывать селекционеру. Показатели величины  $h^2$  имеют реальное значение только для данной конкретной популяции. Чем более генетически однородна популяция, тем ниже коэффициент наследуемости.

Известно и то, что одни организмы более устойчиво сохраняют величину наследуемости при изменении паратипических факторов, а другие менее устойчиво. Эту сторону взаимосвязи наследуемости и среды характеризуют повторяемость.

## **5.2 Коэффициент повторяемости и способы его вычисления**

Повторяемость – это способность организма удерживать свои показатели на определенном уровне при постоянстве факторов среды, а при изменении условий – сохранять свое ранговое место в сравнении с другими животными. Повторяемость и наследуемость связаны между собой. Чем выше повторяемость, тем выше наследуемость. Мерилем повторяемости является коэффициент повторяемости  $r_n$ . Он позволяет судить о той части различий между животными, которая остается устойчивой во все время их племенного использования. Коэффициенты  $h^2$  и  $r_n$  широко применяются при расчетах эффективности селекции.

Коэффициент повторяемости ( $r_n$ ) определяют путем вычисления коэффициента корреляции между последовательными изменениями признака, например, между удоем за первую лактацию и последующие лактации. Высокие значения коэффициента повторяемости (свыше 0,7) позволяют рассчитывать на положительный эффект селекции.

Высокопродуктивные и животные, отобранные по продуктивности за первую лактацию, сохраняют свой ранг по удою и в последующие лактации, т.е. лучшие по первой лактации останутся лучшими и в последующих лактациях.

Напротив, при низкой повторяемости признака, животных приходится оценивать по средней продуктивности за ряд лактаций, что естественно, снижает скорость генетического прогресса.

## Занятие 11.

### 6. ИСПОЛЬЗОВАНИЕ ОСНОВНЫХ БИОМЕТРИЧЕСКИХ ПОКАЗАТЕЛЕЙ В ПЛАНИРОВАНИИ РЕЗУЛЬТАТОВ ОТБОРА И ПОДБОРА

Цель занятия: ознакомиться с элементами селекционно-племенной работы и использованием основных статистических параметров в прогнозировании результатов отбора и подбора.

Для того чтобы повысить продуктивность и племенную ценность животных важно правильно вести племенную работу.

Главная задача племенной работы состоит в постоянном из года в год, из поколения в поколение увеличении средних показателей животных стада по основным хозяйственно полезным признакам, в частности по продуктивности.

К важным элементам племенной работы относятся отбор и племенной подбор отобранных животных.

Отбор – когда оставляют на племя особей желательного типа и устраняют из дальнейшего разведения тех особей, которые в племенном отношении являются нежелательными.

Подбор – закрепление производителей за матками с целью получения потомства с заранее намеченными качествами.

Для планирования результатов отбора и подбора используют такие показатели, как  $\bar{X}$ ,  $\sigma$  и  $h^2$ .

Зная  $\sigma$  и  $h^2$  можно определить:

1. результаты отбора, т.е. предсказать какой будет величина учитываемого признака у потомства, если известна доля животных родительского поколения, оставляемых для размножения;
2. каких животных надо отбирать, чтобы получить от них потомство с желательным развитием учитываемого при отборе признака.

В самом простом случае отбор состоит в разделении животных исходного стада на две группы:

**племенное ядро** - группа лучших животных, потомками которых ремонтируют стадо.

**племенной брак** - группа животных потомки которых непригодны для ремонта стада, а могут иметь только пользовательное назначение.

Начиная отбор прежде всего устанавливают определенную минимальную границу величины признака. Эта минимальная для отбора величина признака называется селекционной точкой, границей отбора, селекционной границей.

Параметры отбора можно определить, исходя из закономерностей нормального распределения (рис. 1).

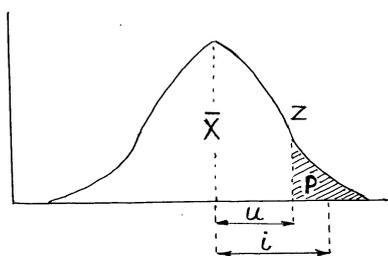


Рис. 1. Графическое изображение интенсивности отбора.

На рисунке 1,  $p$  – доля отобранных животных (заштрихованная часть площади нормальной кривой, ограниченной ординатой  $Z$ , отсекающей уровень признака отобранных животных);  $Z$  – минимальная варианта селекционируемого животного;  $u$  – отклонение от  $\bar{X}$  (средняя родительской генерации) абсциссы в точке ординаты отсека отобранных животных, выраженное в долях  $\sigma$ ;  $i$  – интенсивность отбора (селекции) или селекционный дифференциал, выраженный в единицах среднего квадратического отклонения. Селекционный дифференциал равен  $SD = i \cdot \sigma$ .

Указанные показатели интенсивности отбора можно определить по доле отобранных животных.

#### Определение показателей интенсивности отбора по доле отобранных животных

P	U	Z	i	P	U	Z	i
0,1	1,28	0,18	1,76	0,6	-0,25	0,39	0,64
0,2	0,84	0,28	1,40	0,7	-0,52	0,35	0,50
0,3	0,52	0,35	1,16	0,8	-0,84	0,28	0,35
0,4	0,25	0,39	0,97	0,9	-1,28	0,18	0,19
0,5	0	0,40	0,80	1,0	-	-	-

При отборе животных со значениями признака меньше установленной границы их относят в группу племенного брака. Остальных животных исходной популяции отобранных в племенное ядро допускают к репродукции, обозначают буквой  $P$ .

Тогда доля бракуемых животных составляет  $1-P$ .

Следовательно, чем меньше  $P$  (доля отобранных животных), тем интенсивнее отбор, и наоборот. При  $P = 1$ , отбора нет, т.е. всех животных популяции оставляют для дальнейшего воспроизводства.

$$SG = \bar{X} + u \cdot \sigma, \text{ где}$$

$SG$  – селекционная граница (минимальная продуктивность коров, отобранных для селекции)

$\bar{X}$  – среднее значение признака в стаде

$\sigma$  – показатель разнообразия признака

$$u = \frac{SG - \bar{X}}{\sigma}$$

показывает насколько сигма отклоняется от  $\bar{X}$  селекционная граница.

$$\bar{X}_{\text{ОТОБРАННОЙ ГРУППЫ (ПЛЕМЯДРА)}} = \bar{X}_{\text{СТАДА (ПОПУЛЯЦИИ)}} + i \cdot \sigma$$

Селекционный дифференциал и селекционный эффект являются основными показателями прогнозирования продуктивных качеств животных при разведении и селекции.

Эффективность селекции зависит от многих факторов: количества селекционных признаков, величины коэффициента наследуемости, взаимосвязи между признаками, характера отбора и др.

Селекционный дифференциал – это разница между средней величиной признака у группы животных, отобранных для селекции (племенное ядро), и средней величиной признака всего стада (популяции).

$$SD = \bar{X}_{\text{ПЛЕМЯДРА}} - \bar{X}_{\text{ПОПУЛЯЦИИ}}$$

Чем выше  $S_d$ , тем выше потенциал селекции.

В животноводстве селекционный дифференциал выше у самцов и значительно ниже у самок. Чем выше селекционный дифференциал, тем больше эффект селекции в последующих поколениях.

*Пример* : Определить минимальную границу отбора, средний удой отобранных для воспроизводства коров и величину селекционного дифференциала, если в группу отбора войдет 80 % коров ( $P=0,8$ ), а средний удой по стаду 3000 кг,  $\sigma \pm 500$  кг,  $h^2 = 0,4$ .

По таблице 1 можно определить селекционную границу ( $SG$ ) и среднюю продуктивность отобранных коров: Селекционная граница (минимальная продуктивность отобранных коров) рассчитывается по формуле

$$SG = \bar{X} + u \cdot \sigma = 3000 - 0,84 \cdot 500 = 2580 \text{ кг молока.}$$

Коров с удоем 2580 кг молока и выше оставляют на племя, а с более низкой продуктивностью – выбраковывают.

Средняя молочная продуктивность отобранных коров определяется по формуле:

$$\bar{X}_{\text{ОТОБР.}} = \bar{X}_{\text{СТАДА}} + i \cdot \sigma = 3000 + 0,35 \cdot 500 = 3175 \text{ кг молока.}$$

Отсюда селекционный дифференциал

$$SD = \bar{X}_{\text{ОТОБР.}} - \bar{X}_{\text{СТАДА}} = 3175 - 3000 = 175 \text{ кг молока}$$

Таким образом, отобранные коровы превосходят по продуктивности коров исходной популяции на 175 кг молока (5,8 %).

Низкая доля отбора (ремонта стада) увеличивает его интенсивность и, следовательно селекционный дифференциал, а высокая доля

отбора, наоборот, снижает эти параметры селекции. Чем выше уровень SG, тем выше SD.

Однако нельзя ожидать, что потомки, полученные от отобранной группы животных, превзойдут по продуктивности исходную популяцию в размере SD. Как было отмечено выше, фенотип животных определяется взаимодействием генотипа и среды. Следовательно, необходимо выявить долю генотипического разнообразия селекционируемого признака, которое выражается  $h^2$ . Тогда можно предполагать, что дочери отобранных матерей в аналогичных условиях среды будут превосходить исходную популяцию на  $SD \cdot h^2$ .

Сдвиг в среднем значении признака за одно поколение (при нормальных условиях содержания, кормления, использования) показывает эффект отбора (SE).

Т.О. SE за одну генерацию по одному признаку =  $SD \cdot h^2 = 175 \cdot 0,4 = 70$  кг. молока.

Это значит, что дочери отселекционируемых по молочной продуктивности матерей будут превосходить исходную популяцию на 70 кг. молока, а генетически обусловленный уровень удоя новой (дочерней) популяции составит:

$$\bar{X}_{\text{дочерней популяции}} = \bar{X}_{\text{стада}} + SE = 3000 + 70 = 3070 \text{ кг. молока.}$$

Таким образом, сдвиг средней молочной продуктивности исходной популяции как результат селекции произошел с 3000 до 3070 кг молока на одну корову.

На ускорение процесса совершенствования стад по селекционируемым признакам в значительной степени может оказывать влияние интервал между поколениями. При меньшем интервале в смене поколений можно быстрее получить результаты отбора и, если нужно, изменить направление.

Интервал между поколениями (t) представляет промежуток времени между рождением родителей (отца и матери) и их потомков.

Для исчисления интервала между поколениями Иоганссон и Лаш рекомендуют следующие показатели.

*Интервал смены поколений.*

Животные	По Иоганссону, интервал между поколениями, лет			По Лашу, интервал, лет
	отцовский	материнский	В среднем	В среднем
Молочный скот	4,6	6,0	5,3	4...4,5
Лошади	9,5	8,9	9,3	10...13
Овцы	3,6	4,3	3,9	4...4,5
Свиньи	2,4	3,0	2,7	2,5
Куры	-	-	-	1,5

Однако, приведенные интервалы могут изменяться в зависимости от скороспелости пород, специализации хозяйств, интенсивности селекции и т.д. Поэтому при прогнозировании эффекта селекции для каждой конкретной популяции (стада) необходимо рассчитывать соответствующий интервал между сменой поколений.

Эффект отбора по матерям в расчете за год вычисляют по формуле:

$$SE = \frac{SD \cdot h^2}{t_m}$$

где  $SD$  – селекционный дифференциал со стороны матерей,

$h^2$  – коэффициент наследуемости признака в стаде,

$t_m$  – интервал между поколениями по матерям.

Известно, что на проявление качеств животных нового поколения оказывает влияние наследственность не только матери, но и отца.

Следовательно, необходимо учитывать и результаты подбора. Эффект селекции от подбора производителей рассчитывают:

$$\text{За поколение } SE_o = SD_o \quad \text{за год } SE_o = \frac{SD_o}{t_o}$$

$$SD_o = \bar{X}_{\text{ДОЧЕРЕЙ}} - \bar{X}_{\text{МАТЕРЕЙ}}$$

$$SD_o = \bar{X}_{\text{ДОЧЕРЕЙ ОТЦОВ}} - \bar{X}_{\text{СВЕРСТНИЦ}}, \quad \text{где}$$

$SD_o$  – селекционный дифференциал отцов

$t_o$  – интервал смены поколений по отцам.

Суммарный эффект селекции за поколение с учетом селекционного дифференциала по материнской и отцовской линиям:

$$SE_{\text{за год}} = \left( \frac{SD_m \cdot h^2}{t_m} + \frac{SD_o}{t_o} \right) : 2$$

$$SE_{\text{за поколение}} = (SD_m \cdot h^2 + SD_o) : 2$$

Например, селекционный дифференциал по матерям равен 550 кг, по отцам 350 кг, наследуемость  $h^2$  удою 0,3; интервал между отцовскими поколениями 4,6, материнским 6.

Поставим эти значения в формулу, получаем эффект селекции по удою за год:

$$SE_{\text{за год}} = \left( \frac{550 \cdot 0,3}{6} + \frac{350}{4,6} \right) : 2 = 51,7 \text{ кг. молока}$$

## 7. ИСПОЛЬЗОВАНИЕ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ В БИОЛОГИИ

Цель занятия: ознакомиться с использованием свойств нормального распределения для решения ряда биологических задач.

Закономерности нормального распределения позволяют небольшим числом параметров охарактеризовать основные свойства совокупности и решить ряд задач. Рассмотрим их несколько подробнее.

1. Зная  $\bar{X}$  и  $\sigma$ , можно приближенно (с заданной вероятностью – уровнем надежности) указать минимальное и максимальное значение признака в изучаемой совокупности, т.е. указать вероятные пределы разнообразия.

Так, с вероятностью  $P = 0,95$  минимальное значение признака будет не меньше, чем  $\bar{X} - 1,96 \cdot \sigma$  (или  $\bar{X} - 2\sigma$ ), т.е.  $X_{\min} = \bar{X} - 2\sigma$ ,

$$\text{а максимальное} - X_{\max} = \bar{X} + 2\sigma.$$

При повышении надежности выводов до  $P = 0,999$  соответствующие пределы составят:

$$X_{\min} = \bar{X} - 3,3\sigma ; \quad X_{\max} = \bar{X} + 3,3\sigma.$$

2. Сравнивая фактические максимум и минимум в совокупности с теоретически рассчитанным указанным выше способом, можно сделать предварительный вывод о направлении и интенсивности отбора животных в совокупности по данному признаку. Действительно, если животных с недостаточным развитием признаков браковать, оставляя лучших, это приведет к асимметрии распределения: фактический минимум в совокупности будет больше, чем теоретически рассчитанный по формуле  $X_{\min} = \bar{X} - 2\sigma$ ; причем эта разность больше, чем интенсивнее была браковка животных. В то же время, максимальные значения могут значительно превышать теоретическое значение максимума, рассчитанное даже для  $P = 0,999$ , т.е.  $X_{\max} = \bar{X} + 3,3\sigma$

3. Зная  $\sigma$ , можно определить, с каким материалом мы имеем дело – с однородным или неоднородным. Чем  $\sigma$  меньше, тем крайние отклонения ( $\bar{X} + 3\sigma$ )

меньше и, следовательно, материал однороднее и наоборот.

4. Зная  $\bar{X}$  и  $\sigma$ , можно найти «выскакивающие» (выходящие за пределы  $\pm 3\sigma$ ) варианты и проверить их на принадлежность к изучаемой совокупности.

5. Зная  $\bar{X}$  и  $\sigma$ , можно для любой варианты в совокупности рассчитать нормированное отклонение:

$$t = \frac{(X - \bar{X})}{\sigma}$$

Нормированное отклонение является неименованной величиной, что позволяет использовать его для сравнения отдельных объектов, принадлежащих к разным генеральным совокупностям, по степени развития у них того или иного признака, а также определять, какой признак у данного объекта развит лучше.

Например, при отборе коров в племенную группу (или при покупке племенных животных) необходимо решить, какое из двух животных разного возраста лучше по молочной продуктивности. Известно, что одна корова за первую лактацию дала 3500 кг молока, а другая после третьего отела имела удой 4580 кг. Известно также, что средний удой первотелок в данном хозяйстве в данном году составил 2500 кг и  $\sigma = 500$  кг, а удой полновозрастных животных (третий отел и старше) – 3500 кг и  $\sigma = 600$  кг. Если для этого случая рассчитать абсолютные отклонения, то вторая корова будет выглядеть предпочтительнее первой:

$$d_1 = X_1 - \overline{X}_1 = 3500 - 2500 = 1000 \text{ кг.}$$

$$d_2 = X_2 - \overline{X}_2 = 4580 - 3500 = 1080 \text{ кг.}$$

Однако этот вывод был бы ошибочным.

Расчет нормированного отклонения  $t$  покажет их действительное место в соответствующих вариационных рядах по степени развития данного признака:

$$t_1 = \frac{1000}{500} = 2,0; \quad t_2 = \frac{1080}{600} = 1,8$$

Отсюда видно, что первотелка более обильномолочна и ей следует оказать предпочтение.

Соответствующим образом решается вопрос о том, какой из двух или нескольких признаков развит лучше у данного животного. Если, например, по удою данная корова характеризуется значением  $t = 2,0$ ; по жирномолочности –  $t = 2,5$ ; А по живой массе –  $t = 1,7$ , то совершенно ясно, что у нее сильнее развит признак жирномолочности.

6. Нормированное отклонение используется при решении селекционных задач для сравнения и оценки интенсивности отбора и селекционной границы в разных стадах и по разным признакам:

$$i = \frac{X_{отоб} - X_{стада}}{\sigma},$$

где  $i$  – нормированное отклонение, характеризующее интенсивность отбора;

$\overline{X}_{отоб}$  - среднее значение признака в группе отобранных на племя животных;

$\overline{X}_{стада}$  - среднее значение признака в группе до отбора;

$$U = \frac{(SI - \overline{X}_{стада})}{\sigma}$$

$U$  – нормированное отклонение, характеризующее границу отбора.

Например, средний удой в стаде  $\bar{X}_{1 \text{ стада}}$  составил 3500 кг при среднем квадратическом отклонении  $\sigma_1 = 600$  кг. Средняя жирномолочность коров в стаде  $\bar{X}_{2 \text{ стада}} = 3,7\%$  при  $\sigma_2 = 0,2\%$ . Группа животных, отобранных из стада для племенного использования (племенное ядро), имела средний удой  $X_1 \text{ отоб} = 3800$  кг и среднюю жирномолочность  $\bar{X}_{2 \text{ отоб}} = 3,77\%$ . Рассчитав  $i$  можно установить по какому из признаков провели более интенсивный отбор:

$$i_1 = \frac{(3800 - 3500)}{600} = 0,5 \quad i_2 = \frac{(3,77 - 3,7)}{0,2} = 0,35$$

В стаде большее внимание уделялось отбору по удою, чем по жирномолочности.

### Занятие 13.

## 8. ОСНОВЫ ДИСПЕРСИОННОГО АНАЛИЗА

*Цель занятия* – ознакомиться с основными положениями дисперсионного анализа и его практическим применением в работе.

### 8.1. Техника дисперсионного анализа

В зоотехнической практике и научных исследованиях довольно часто возникают ситуации, когда необходимо подвергнуть сравнительному анализу одновременно не две, а несколько групп животных: например, при испытании пород, линий, при оценке производителей по качеству потомства, определении стандартности линии или отобранной для каких-либо целей группы животных и в других случаях, связанных с изучением влияния различных факторов (биологических, кормовых, гигиенических) на рост, развитие, продуктивность, здоровье животных и другие интересующие зоотехника (или биолога) признаки.

В большинстве случаев важно не только установить факт воздействия на изучаемый объект того или иного фактора (т.е. достоверность влияния), но и выяснить степень этого воздействия, т.е. оценить его относительную силу. Сравнить одновременно несколько средних, определить достоверность и силу влияния различных факторов позволяет метод дисперсионного анализа – один из сложнейших и не до конца разработанных разделов биометрии.

Фактором может быть любое воздействие на изучаемые объекты, в частности на животных. Все факторы можно разделить на четыре группы:

I. Внешние по отношению к объекту воздействия: температура, влажность, место обитания, условия жизни и питания, стимуляторы жизненных функций и т.п.

II. Состояние животных: беременность (разные периоды), лактация, возраст, пол, работа, здоровье и т.п.

III. Отдельные признаки животных, влияющие на развитие других признаков: например, удои коров в зависимости от их живой массы. Здесь дисперсионный анализ соприкасается с корреляционным, и одновременное их проведение составляет так называемый *п о л н ы й корреляционный анализ*.

IV. Происхождение животных от разных пород, линий, семейств, от отдельных производителей или отдельных маток, от групп производителей или от групп маток разного качества. Изучение влияния факторов четвертой группы играет важную роль в селекции животных, а также в генетических исследованиях.

Обычно приходится изучать разные степени действия фактора – разные его дозы, называемые *г р а д а ц и я м и ф а к т о р а*. Так, например, при изучении наследственных влияний градами фактора будут отдельные производители (или отдельные матки).

При изучении действия только одного фактора (в нескольких градах) необходимую информацию о его силе и достоверности получают путем анализа *о д н о ф а к т о р н о г о д и с п е р с и о н н о г о к о м п л е к с а*. При изучении действия одновременно двух или более факторов анализируют двухфакторный (или многофакторный) дисперсионный комплекс.

Следует отметить, что трехфакторный и более дисперсионный анализ используют в практике довольно редко вследствие ненадежности получаемых оценок силы влияния. Кроме того, существуют некоторые обязательные условия организации многофакторных комплексов. Одно из них – полная независимость изучаемых факторов друг от друга – обычно трудно выполнимо.

Например, нельзя изучать влияние таких факторов, как рост и живая масса, на продуктивность животных методом двухфакторного дисперсионного анализа, поскольку рост и масса животных взаимосвязаны. Лучше в этом случае провести два однофакторных дисперсионных анализа.

Группа животных, которая подвергается воздействию определенной градации фактора, образует *г р а д а ц и ю д и с п е р с и о н н о г о к о м п л е к с а*. Таким образом, число градаций комплекса соответствует числу градаций фактора.

Наиболее важный момент в исследовании – отбор животных в группы – градации комплекса. Предположим, что исследуется влияние использовавшихся в стаде быков-производителей на удои их дочерей. Может случиться так, что дочери одного быка оценены по удою только за первую лактацию, дочери другого быка – по удою за вторую и третью лактации, а дочери третьего – за третью, четвертую и пятую лакта-

ции. Если не обратить внимание на этот факт, что первотелки в среднем дают за лактацию молока меньше, чем коровы более старших возрастов (удой повышаются до 5-8 отелов), то в приведенном примере даже при абсолютной одинаковости быков – производителей по их наследственным качествам, обуславливающим молочную продуктивность, можно прийти к неправильным выводам, отдав предпочтение третьему быку-производителю, (дочери которого будут иметь более высокий удой, как более старшие по возрасту). Очевидно, что для выяснения влияния производителей на удои дочерей необходимо учитывать возраст телок. В данном примере сильное влияние на результат могут оказать и коровы-матери (дочери одного быка могут быть получены только от лучших матерей, а другого – от плохих и средних), а также условия кормления и содержания животных (если дочери сравниваемых производителей выращивались и продуцировали в разных условиях). Чтобы избежать наложения действия других факторов на действие изучаемых градаций фактора (производителей), следует дочерей разных быков от матерей примерно одинакового качества вырастить в одинаковых условиях.

Градации комплекса составляют таким образом, чтобы обеспечить случайность действия всех остальных факторов (кроме изучаемого), их неорганизованность, равновероятную направленность (рандомность) действия, что создаст фон, на котором можно выявить закономерность действия организованного в градации фактора.

*Градации комплекса* – это выборки, сделанные из заведомо разных генеральных совокупностей, поэтому наиболее приемлемым принципом формирования градаций комплекса является принцип случайного отбора в них отдельных объектов из совокупностей.

В тех случаях, когда не удается избежать систематического влияния на объект неизучаемого фактора (не удастся рендомизировать его влияние), прибегают к двухфакторному дисперсионному анализу. И хотя второй фактор сам по себе может не интересовать исследователя, его влияние учитывают, чтобы выявить степень влияния изучаемого фактора. Влияние фактора оценивают по изменениям у животных отдельного признака, который называют *результативным признаком*.

Основные понятия и принципы использования дисперсионного анализа для обработки зоотехнических данных можно изучить на примере анализа однофакторного дисперсионного комплекса, имеющего наиболее широкое применение в практике селекционно-генетических исследований.

*Пример:* При испытании новой минеральной подкормки для подсосунков организован однофакторный дисперсионный комплекс. Изуча-

ется несколько степеней интенсивности действия фактора: пять увеличивающихся доз подкормки – нулевая (контроль), одинарная, двойная, тройная и учетверенная.

Для каждой дозы подбирают в случайном порядке по несколько подсвинков. По дозам подкормки подобраны аналоги: для первой – три, для второй – четыре и далее пять, четыре, четыре.

Степени интенсивности фактора называются градациями фактора.

Частные группы, подобранные для каждой градации фактора, называются градациями комплекса.

У каждой особи, вошедшей в комплекс, измеряют результативный признак, например, у каждого подсвинка определяют прирост веса в килограммах за опытный период. И, наконец, для каждой градации вычисляют частные средние значения признака. В результате получается дисперсионный комплекс, приведенный в таблице.

Однофакторный дисперсионный комплекс

Градация фактора	0	1	2	3	4	r=5
Даты X	2, 3, 1	4,3,6,3	5,6,4,6,9	9,7,6,6	3, 6, 5, 6	N=20 ΣX=100
Объем градаций n	3	4	5	4	4	
Суммы дат ΣX	6	16	30	28	20	
Частные средние $\bar{X}_{i=}$ $\frac{\sum X}{n}$	2	4	6	7	5	$\bar{X}_{\Sigma} = \frac{100}{20} = 5$

Этот комплекс имеет следующую структуру: объем комплекса N=20, число градаций r=5, повторность по градациям 3, 4, 5, 4, 4.

Это комплекс однофакторный, неравномерный. В равномерных комплексах повторности по градациям равны:  $n_1=n_2=\dots n_r$ , поэтому N=rn.

Решение комплекса начинается с рассмотрения ряда частных средних.

В комплексе ряд частных средних:  $\bar{X}_{i=}$  2-4-6-7-5 указывают на заметное влияние изучаемого средства. Средний привес увеличивается в достаточном соответствии с усилением фактора: при слабых дозах (0, 1) привесы меньшие, при сильных дозах (2, 3, 4) привесы большие.

После рассмотрения ряда частных средних рассчитывают показатели дисперсионного анализа.

### Расчет подсобных величин

Вначале рассчитываются следующие вспомогательные величины:

1. Сумма дат по всему комплексу:

$$\sum X = 6 + 16 + 30 + 28 + 20 = 100$$

2. Общая подсобная величина:

$$H_{\Sigma} = \frac{(\sum X)^2}{N} = \frac{100^2}{20} = 500$$

3. Сумма частных подсобных величин:

$$\sum H_i = 12 + 64 + 180 + 196 + 100 = 552$$

Каждую частную подсобную величину рассчитывают по каждой градации – это квадрат суммы дат по градации, деленный на объем этой градации. Например:

$$H_1 = \frac{(\sum X_1)^2}{n_1} = \frac{6^2}{3} = 12; \quad H_2 = \frac{16^2}{4} = 64 \text{ и т.д.}$$

4. Сумма квадратов дат по всему комплексу –  $\sum X^2$ . Чтобы избежать ошибок внимания, можно предварительно рассчитать сумму квадратов дат отдельно по каждой градации, а затем эти частные суммы сложить:

$$\sum \sum X^2 = 14 + 70 + 194 + 202 + 106 = 586.$$

#### Расчет дисперсий (сумм квадратов)

В однофакторном дисперсионном комплексе рассчитывают три дисперсии (суммы квадратов): факториальную ( $C_x$ ), случайную ( $C_z$ ) и общую ( $C_y$ ).

$C_x$  – факториальная (межгрупповая) дисперсия, сумма взвешенных квадратов центральных отклонений частных средних по градациям (или по группам) от общей средней по всему комплексу:

$$C_x = \sum n (\bar{X}_i - \bar{X}_{\Sigma})^2. \text{ Рассчитывают по рабочей формуле:}$$

$$C_x = \sum H_i - H_{\Sigma} = 552 - 500 = 52.$$

$C_z$  – случайная (внутригрупповая) дисперсия, сумма квадратов центральных отклонений каждой даты от своей частной средней:

$$C_z = \sum (X - \bar{X}_i)^2. \text{ Рассчитывают по рабочей формуле:}$$

$$C_z = \sum \sum X^2 - \sum H_i = 586 - 552 = 34.$$

$C_y$  – общая дисперсия, сумма квадратов центральных отклонений дат от общей средней по всему комплексу:  $C_y = \sum (X - \bar{X}_{\Sigma})^2$ . Рассчитывают по рабочей формуле:

$$C_y = \sum \sum X^2 - H_{\Sigma} = 586 - 500 = 86.$$

На основе трех дисперсий однофакторного комплекса можно рассчитать оба заключительных показателя дисперсионного анализа.

1. Показатель силы влияния:

$$\eta_x^2 = C_x / C_y = 52 / 86 = 0,605.$$

2. Показатель достоверности влияния:

$$F = \frac{C_x}{C_z} \cdot \frac{N-r}{r-1} = \frac{52}{34} \cdot \frac{20-5}{5-1} = 5,7$$

Однако для некоторых деталей дисперсионного анализа, учитывая, что в большинстве работ показатель достоверности рассчитывают по отношению варiances, необходимо остановиться на использовании и этих величин.

### **Расчет варiances (средних квадратов)**

В однофакторном дисперсионном комплексе рассчитывают две варiances (средние квадраты): факториальную (межгрупповую) и случайную (внутригрупповую). Каждая варiances равна дисперсии, деленной на число степеней свободы:

$$\sigma_i^2 = C_i/v_i.$$

$\sigma_x^2$  – факториальная варiances равна факториальной дисперсии, деленной на число степеней свободы (число классов без одного):

$$\sigma_x^2 = \frac{C_x}{r-1} = \frac{52}{5-1} = 13,00.$$

$\sigma_z^2$  – случайная варiances равна случайной дисперсии, деленной на число степеней свободы (объем комплекса без числа градаций):

$$\sigma_z^2 = \frac{C_z}{N-r} = \frac{34}{20-5} = 2,27.$$

### **ОСНОВНОЙ ПОКАЗАТЕЛЬ СИЛЫ ВЛИЯНИЯ**

Показатель силы влияния равен отношению факториальной дисперсии к общей. Он определяет ту долю общей дисперсии, которая приходится на факториальную дисперсию, или долю влияния изучаемого фактора в общей сумме влияния всех вообще факторов, определяющих величину результативного признака.

Вычисляют показатель силы влияния по формуле:

$$\eta_x^2 = C_x/C_y = 52/86 = 0,605.$$

Полученный показатель силы влияния ( $\eta_x^2 = 0,605$ ) вскрывает достаточно сильное действие проверяемого кормового средства среди 20 исследованных подсвинков. Из всех факторов, определивших разнообразие привесов на действие изученной минеральной добавки пришелся 61%, конечно, при всех тех условиях, в которых проходил эксперимент.

Описанный способ определения силы влияния можно применять во всех случаях изучения действия одного фактора на количественный результативный признак, причем не требуется делать никаких поправок на к дисперсиям, ни к варiancesам.

В основу показателя силы влияния  $\eta_x^2 = C_x/C_y$  положен основной закон аддитивности дисперсий:

$$C_x + C_z = C_y \quad ; \quad \frac{C_x}{C_y} + \frac{C_z}{C_y} = \frac{C_y}{C_y} \quad ; \quad \eta_x^2 + \eta_z^2 = 1; \quad \eta_x^2 = C_x/C_y .$$

Кроме того, этот показатель есть отношение двух основных показателей дисперсионных комплексов – факториальной и общей дисперсий (сумм квадратов). Поэтому этот показатель можно назвать основным показателем силы влияния в отличие от дополнительных, в основу которых положены отношения компонентов математических ожиданий вариантов (средних квадратов).

Путь, который мы использовали для расчета основных показателей дисперсионного комплекса, дал возможность в учебных целях продемонстрировать математическую сущность рассчитываемых показателей и логический ход анализа.

На практике анализ проводят по алгоритму, разработанному Плохинским Н.А.

**АЛГОРИТМ ОДНОФАКТОРНОГО ДИСПЕРСИОННОГО АНАЛИЗА  
(для количественных признаков для малых групп)**

	Градации					Число градаций r=5	Факториальная дисперсия $C_x = \sum H_i - H_\Sigma = 552 - 500 = 52.$
	1	2	3	4	5		
							Случайная дисперсия
Даты X	2 3 1	4 3 6 3	5 6 4 6 9	9 7 6 6	3 6 5 6	$H_\Sigma = \frac{\sum X^2}{N} = \frac{100^2}{20} = 500$	$C_z = \sum \Sigma X^2 - \Sigma H_i = 586 - 552 = 34$
							Общая дисперсия
n	3	4	5	4	4	Объем комплекса $N = \sum n = 20$	$C_y = \sum \Sigma X^2 - H_\Sigma = 586 - 500 = 86$
$\sum X$	6	16	30	28	20	$\sum \Sigma X = 100$	
							Факториальная дисперсия
$H_i = \frac{(\sum X)^2}{n}$	12	64	180	196	100	$\sum H_i = 552$	$\sigma_x^2 = \frac{C_x}{r-1} = \frac{52}{4} = 13,00$
$\sum X^2$	14	70	194	202	106	$\sum \Sigma X^2 = 586$	
Частные средние $\bar{X}_i$	2	4	6	7	5	Общая средняя $\bar{X}_\Sigma = 5$	Случайная дисперсия $\sigma_z^2 = \frac{C_z}{N-r} = \frac{34}{15} = 2,27$
Показатель силы влияния $\eta_x^2 =$						$C_x = 52$	$= \underline{0,605}$

*Критерий достоверности*

$$F = \sigma_x^2 / \sigma_z^2 = 13,00 / 2,27 = \underline{5,7}$$

$$v_1 = r - 1 = 4; v_2 = N - r = 15; F_{st} = \{3,1 - 4,9 - 8,3\}$$

(табл. Стандартные значения критерия Фишера  $\sigma_1^2 / \sigma_2^2$ )

*Общий вывод:* влияние фактора достоверно с вероятностью  $p > 0,99$

Форма итоговой записи				
Разнообразие	Дисперсии (суммы квадратов) С.	Числа сте- пеней сво- боды v	Вариансы (средние квадраты) $\sigma^2$	$\eta_x^2 =$ <u><u>0,605</u></u>
Факториальное (межгрупповое)	52	4	13,0	
Случайное (внутригрупповое)	34	15	2,27	
Общее	86	19	4,53	

*Вывод:* Влияние фактора достоверно, т.е. такого же действия подкормки следует ожидать при ее массовом применении на поросятах того же возраста и пола (на большом поголовье).

### Занятие 13.

#### ИСПОЛЬЗОВАНИЕ МЕТОДА $\chi^2$ (ХИ – КВАДРАТ) ДЛЯ АНАЛИЗА РЕЗУЛЬТАТОВ ГЕНЕТИЧЕСКИХ ЭКСПЕРИМЕНТОВ

*Цель занятия* – освоить метод хи-квадрат в применении к анализу результатов генетических опытов и ознакомиться с формой записей результатов генетического анализа.

В основе генетического расщепления лежат биологические механизмы гаметогенеза и оплодотворения.

При образовании гамет биологический механизм мейоза обеспечивает расхождение гомологичных хромосом и аллелей по разным гаметам, а также независимое друг от друга расхождение и комбинаторику негомологичных хромосом. Лишь часть из огромного числа образующихся гамет (выборка) участвует в образовании зигот. Сочетание гамет разного типа при оплодотворении происходит случайно. Вследствие этого проявление закона расщепления носит статистический характер. Чем больше объем выборки, т.е. большее количество потомков соответствующего поколения будет получено в эксперименте, тем ближе

окажутся результаты опыта – фактически полученное расщепление – к теоретически ожидаемым.

Отклонения расщепления в опыте от теоретически ожидаемого могут быть вызваны разными причинами:

- а) случайностью выборки;
- б) различиями в жизнеспособности гамет и зигот;
- в) неслучайным характером оплодотворения;
- г) нечетким различием фенотипов;
- д) участием нескольких генов в определении признака;
- е) недостаточной чистотой исходного материала.

Для проверки соответствия опытных данных теоретическим используют метод  $\chi^2$  (хи-квадрат), разработанный К. Пирсоном (1900); его часто называют «критерием согласия» или «Критерием расхождения распределений».

С помощью этого метода можно оценить, являются ли отклонения, наблюдаемые в опыте, случайными. Если отклонения оказываются неслучайными, эксперимент следует повторить, а также применить другие методы генетического исследования.

Следует помнить, что метод  $\chi^2$  не может быть применен, если значения величины в опыте (количество объектов в классах) выражены в процентах или в относительных числах (долях), а также если в выборке число особей в каком-либо из теоретических классов меньше пяти (оптимальным считается число особей не менее 50). Критерий  $\chi^2$  вычисляется по формуле:

$$\chi^2 = \sum \frac{(\Phi - T)^2}{T},$$

Где  $\Phi$  – фактическое число особей (объектов) в каждом классе;

$T$  – теоретически ожидаемое число особей (объектов) в классе.

Полученное значение  $\chi^2$  следует сравнить с табличным. При использовании критерия хи-квадрат следует помнить: чем ответственнее исследования, тем при меньшем пороге вероятности различия должны считаться достоверными, и наоборот.

Таким образом, табличные значения (соответствующие  $P_1 = 0,95$ ) используются при высокой ответственности исследований,  $\chi_2^2$  ( $P_2 = 0,99$ ) – при обычной, а  $\chi_3^2$  ( $P_3 = 0,999$ ) – при малой (!) ответственности исследований.

Различия считаются случайными, если фактическое значение не достигает табличного для соответствующего уровня вероятности, т.е.  $\chi^2_{\text{факт}} < \chi^2_{\text{табл.}}$ .

Рассмотрим пример. В опыте по моногибридному скрещиванию изучали исследование нормальных и зачаточных крыльев у плодовой

мушки *Drosophila melanogaster*. Анализировали F<sub>2</sub> от скрещивания родительских форм, различающихся между собой по форме крыльев.

P: нормальные крылья × зачаточные крылья

В F<sub>1</sub> от реципрокных вариантов скрещивания наблюдалось единообразие: все мухи были с нормальными крыльями. При анализе F<sub>2</sub> получены следующие результаты: 285 мух с нормальными крыльями и 91 – с зачаточными.

Найдем теоретически ожидаемые результаты. Предполагалось расщепление 3:1. Рассчитаем количество мух в каждом классе, которое бы точно соответствовало этому соотношению. Сложим общее количество объектов и разделим на 4, узнав, сколько объектов приходится соответственно на 1 и на 3 части. 285 + 91 = 376. Ожидаемые результаты – 282 : 94. Все расчеты оформляются в таблицу.

### Вычисление $\chi^2$ (хи- квадрат)

Класс	Данные опыта Ф	Теоретически ожидаемые результаты Т	Ф-Т	$(\Phi - T)^2$	$(\Phi - T)^2$	$\chi^2$
					Т	
Нормальные крылья	285	282	+ 3	9	0,03	0,12
Зачаточные крылья	91	94	- 3	9	0,09	

$$\chi^2_{\text{табл}} \quad v = g - 1 = 2 - 1 = 1, \quad \chi^2_{\text{табл}} \{3,8 - 6,6 - 10,8\}.$$

g – число классов. 0,12 < 3,8. Разность недостоверна.

Отклонения случайны, и расщепление в F<sub>2</sub> можно считать соответствующим теоретически ожидаемому 3:1. Из этого следует, что наблюдаемое разнообразие особей по изученному признаку обусловлено различиями родительских особей по одной паре аллельных генов при полном доминировании нормальных крыльев над зачаточными. Обозначим аллель нормальных крыльев буквой А, зачаточных - а.

### Схемы скрещиваний

Прямое	Обратное
P : ♀ AA × aa ♂	P: ♀ aa × ♂ AA
F <sub>1</sub> : Aa	F <sub>1</sub> : Aa
F <sub>2</sub> : AA : 2Aa : aa	F <sub>2</sub> : AA : 2Aa : aa
3 : 1	3 : 1

Выводы:

1. Признак наследуется моногенно (одна пара аллелей).
2. Нормальные крылья доминируют над зачаточными.

3. Признак не сцеплен с полом (аутосомное наследование), так как в  $F_1$  в реципрокных вариантах получены одинаковые результаты – наблюдалось единообразие.

**Задание № 12.** Вычислите критерий  $\chi^2$  и определите соответствие между фактическим и ожидаемым расщеплением, если при скрещивании медленнооперяющихся петухов с розовидным гребнем с быстрооперяющимися курами с листовидным гребнем во втором поколении получено 655 особей с розовидным гребнем, из которых оказалось 157 с быстрой оперяемостью и 498 с медленной оперяемостью, 201 особь с листовидным гребнем, среди которых было 60 с быстрой оперяемостью и 141 с медленной оперяемостью.

**Задание № 13.** Вычислите критерий  $\chi^2$  и оцените соответствие между фактическими и ожидаемыми данными в случае, когда от помесей свиней первого поколения (черный гемпшир×красный дюрок) получено потомство следующих фенотипов: 81 поросенок черной масти и 26 красной масти. При нулевой гипотезе (согласно которой масть обусловлена одной парой генов) ожидается расщепление по масти в соотношении 2 черных: 1 красный.

## П Р И Л О Ж Е Н И Е

### Стандартные значения критерия t для малых выборок (по Стьюденту)

Число степеней свободы	Вероятность (p)				
	0,90	0,95	0,98	0,99	0,999
1.	6,31	12,17	31,82	63,66	-
2.	2,92	4,30	6,97	9,93	31,60
3.	2,35	3,18	4,54	5,84	12,94
4.	2,13	2,78	3,75	4,60	8,61
5.	2,02	2,57	3,37	4,03	6,86
6.	1,94	2,45	3,14	3,71	5,96
7.	1,90	2,37	3,00	3,50	5,41
8.	1,86	2,31	2,90	3,36	5,04
9.	1,83	2,26	2,82	3,25	4,78
10.	1,81	2,23	2,76	3,17	4,59
11.	1,80	2,20	2,72	3,11	4,44
12.	1,78	2,18	2,68	3,06	4,32
13.	1,77	2,16	2,65	3,01	4,22
14.	1,76	2,15	2,62	2,98	4,14
15.	1,75	2,13	2,60	2,95	4,07
16.	1,75	2,12	2,58	2,92	4,02
17.	1,74	2,11	2,57	2,90	3,97
18.	1,73	2,10	2,55	2,88	3,92
19.	1,73	2,09	2,54	2,86	3,88
20.	1,73	2,09	2,53	2,85	3,85
21.	1,72	2,08	2,52	2,83	3,82
22.	1,72	2,07	2,51	2,82	3,79
23.	1,71	2,07	2,50	2,81	3,77
24.	1,71	2,06	2,49	2,80	3,75
25.	1,71	2,06	2,49	2,79	3,73
26.	1,71	2,06	2,48	2,78	3,71
27.	1,70	2,05	2,47	2,77	3,69
28.	1,70	2,05	2,47	2,76	3,67
29.	1,70	2,05	2,46	2,75	3,66
30.	1,70	2,04	2,46	2,75	3,05
$\infty$	1,64	1,96	2,33	2,58	3,29

## ЛИТЕРАТУРА

### ОСНОВНАЯ

1. Бакай А.В. , Кочиш И.И., Скрипниченко Г.Г. и др. Практикум по генетике. М.: КолосС, 2010. – 301 с.
2. Бакай А.В. и др. Генетика. М: Колос,2006. –с.448.
3. Бессчетнов В.П. и др. Дисперсионный анализ многоуровневых иерархических комплексов: Методические указания для студентов и аспирантов очного и заочного отделений факультета лесного хозяйства по специальности 25.02.01.65 – Лесное хозяйство/Бессчетнов В.П., Бессчетнова Н.Н., Храмова О.Ю, Орнатский А.Н., Горелов Н.И. – Нижний Новгород: Нижегородская ГСХА. 2012. – 33 с.
4. Н.А.Зиновьева, В.Р.Харзинова, Е.И.Сизарева, Е.А.Гладырь, О.В.Костюхина, Л.Н. Гамко, Овсеенко Е.В. Оценка вклада различных популяций в генетическое разнообразие свиней корня крупной белой породы. Сельскохозяйственная биология, 2012, № 6
5. Меркурьева Е.К. и др. Генетика. М.: Агропромиздат, 1991. – с. 165-228.
6. Мятлев В.Д. и др. Теория вероятностей и математическая статистика, математические модели. Серия: университетский учебник. Высшая математика и ее приложения к биологии.// В.Д. Мятлев, Л. А Панченко, Г.Ю. Ризниченко, А.Т. Терехин. – М.: «Академия», 2009. – 320 с. Для студентов биологических сений.
7. Овсеенко Ю.В , Овсеенко Е.В. Словарь физиологических терминов. Учебно-методическое пособие./ - Брянск: Брянская ГСХА, 2014.
8. Петухов В.Л. и др. Ветеринарная генетика. – М.: Колос, 1996. – с. 125-156.

### ДОПОЛНИТЕЛЬНАЯ

1. Жученко А.А. и др. Генетика. М.: КолосС, 2003. – 480 с.
2. Меркурьева Е.К., Шангин-Березовский Г.Н. Генетика с основами биометрии. – М.: Колос, 1983. с. 170-200.
3. Ларцева С.Х. , Муксинов М.К. Практикум по генетике. – М.: Агропромиздат, 1985.- с. 5-68.

### СПРАВОЧНАЯ

1. Плохинский Н.А. Руководство по биометрии для зоотехников. М.: Колос, 1969. – с. 256.

## СОДЕРЖАНИЕ

Введение.....	3
Основные сведения о биометрии.....	4
1. Расчет основных статистических параметров в большой выборочной совокупности.....	8
1.1 Построение вариационного ряда и вариационных кривых.....	8
1.2 Вычисление средней арифметической величины.....	13
1.3 Показатели, характеризующие разнообразие признаков.....	15
2 Ошибки репрезентативности и их применение в биометрии.....	18
2.1 Определение доверительных границ параметров генеральной совокупности.....	20
2.2 Определение достоверности выборочных параметров.....	22
2.3 Определение достоверности разности между средними арифметическими двух выборочных совокупностей .....	22
2.4 Определение объема выборок.....	26
3. Статистические взаимосвязи и способы вычисления их величин... ..	27
3.1 Корреляционная связь и вычисление коэффициента фенотипической корреляции.....	27
3.2 Вычисление рангового коэффициента корреляции Спирмена.....	34
3.3 Коэффициент регрессии и его использование в зооветеринарной работе.....	36
4. Расчет основных статистических параметров в малой выборке.....	39
5 Статистический анализ наследуемости и повторяемости.....	42
5.1 Коэффициент наследуемости и способы его вычисления.....	42
5.2 Коэффициент повторяемости и способы его вычисления.....	45
6. Использование основных биометрических показателей в планировании результатов отбора и подбора.....	46
7. Использование нормального распределения в биологии.....	51
8. Основы дисперсионного анализа.....	53
8.1 Техника дисперсионного анализа.....	53
8.2 Использование метода $\chi^2$ (хи –квадрат) для анализа результатов генетических экспериментов.....	60
Приложение.....	64
Список литературы.....	65

УДК 636.082.12 (07)  
ББК 45.3  
О 34

### УЧЕБНО-МЕТОДИЧЕСКОЕ ПОСОБИЕ

для студентов, обучающихся по направлению подготовки 36.04.02 «Зоотехния» (уровень высшего образования – Магистратура) и аспирантов.

Овсеенко Е.В., Базутко Н.П. Математические методы в биологии. Учебно-методическое пособие для студентов, обучающихся по направлению подготовки 36.04.02 «Зоотехния» (уровень высшего образования – Магистратура) и аспирантов / Е.В. Овсеенко, Н.П. Базутко – Брянская область: Издательство Брянского ГАУ, 2014 – 69 с.

Учебно-методическое пособие предназначено для студентов для студентов, обучающихся по направлению подготовки 36.04.02 «Зоотехния» (уровень высшего образования – Магистратура) и аспирантов.

Рецензент: доктор сельскохозяйственных наук, профессор В.Е. Подольников

Рекомендовано к изданию решением методической комиссии института ветеринарной медицины и биотехнологии от \_\_\_\_\_ протокол

© ФГБОУ ВПО Брянский ГАУ, 2015

© Овсеенко Е.В., 2015

© Базутко Н.П., 2015

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РФ  
ФГОУ ВО «БРЯНСКИЙ ГАУ»  
*ИНСТИТУТ ВЕТЕРИНАРНОЙ МЕДИЦИНЫ И BIOTEХНОЛОГИИ*  
КАФЕДРА КОРМЛЕНИЯ ЖИВОТНЫХ И ЧАСТНОЙ ЗООТЕХНИИ

Е.В. Овсеенко, Н.П. Базутко

## **МАТЕМАТИЧЕСКИЕ МЕТОДЫ В БИОЛОГИИ**

учебно-методическое пособие  
для студентов, обучающихся по направлению подготовки 36.04.02  
«Зоотехния» (уровень высшего образования – Магистратура)  
и аспирантов

Брянская область 2015

Учебное издание

Овсеенко Елена Викторовна

Базутко Наталья Петровна

## **МАТЕМАТИЧЕСКИЕ МЕТОДЫ В БИОЛОГИИ**

учебно-методическое пособие  
для студентов, обучающихся по направлению подготовки 36.04.02  
«Зоотехния» (уровень высшего образования – Магистратура)  
и аспирантов

---

Подписано к печати г. Формат 60×84 1/16.

Бумага печатная. Усл. п. л. 3,02. Тираж 25 экз. Изд. №

---

Издательство Брянского ГАУ.

243365 Брянская область., Выгоничский район, с. Кокино, Брянский ГАУ