ЗАНЯТИЕ 1.
БИОМЕТРИЯ. ХАРАКТЕРИСТИКА СТАДА ПО СРЕДНИМ ПОКАЗАТЕЛЯМ. СОСТАВЛЕНИЕ СПИСКА КОРОВ

Цель занятия. Основные понятия. Использование методов биометрии в генетике и зоотехнии. Усвоение методов вычисления средних величин в зависимости от поставленных задач и числа животных в выборке. Составление списка коров и взятие данных по их продуктивности: удой, % жира, живая масса матерей и дочерей из форм 2-МОЛ (15 животных).

Методические указания. *Биометрия* – наука о статистическом анализе групповых свойств в биологии. Эти свойства могут изучаться по характеру межгруппового и внутригруппового разнообразия при любой численности (начиная с двух).

Методы биометрии позволяют дать математически точные характеристики свойств и признаков совокупностей, выявить степень генетического разнообразия признака и влияния на него различных факторов, прогнозировать эффект селекции.

Методы биометрии основаны на теории вероятности законе больших чисел.

Вероятность – объективная возможность наступления какого-либо события. Событие может наступить (при благоприятных условиях), а может не наступить (при неблагоприятных условиях).

Генеральная совокупность – большой массив животных, интересующих исследователя (например, животных бестужевской или другой породы крупного рогатого скота). Объем генеральной совокупности может быть как очень большим (50 тыс., 500 тыс. и более), так и очень малым.

Метод сплошного обследования, т. е. изучение всех членов генеральной совокупности, требует большой затраты времени и труда. В

случаях когда изучение генеральной совокупности связано с убоем животных, метод сплошного обследования совершенно неприемлем.

Для изучения генеральной совокупности составляют выборку.

Выборка должна быть типичной, т. е. правильно отражать генеральную совокупность. Например, при изучении молочной продуктивности коров нельзя включать в выборку больных животных и животных с атрофией сосков, т. к. они нетипичны для изучаемой совокупности. Выборка должна быть однородной (одна порода, один пол). Она составляется по принципу случайного отбора. Это значит, что в выборку с равной вероятностью может попасть каждый член генеральной совокупности.

При изучении генеральной совокупности по выборке, т. е. характеристике целого по его части при случайном отборе особей, неизбежны ошибки репрезентативности, указывающие на степень соответствия выборочных показателей параметрам генеральной совокупности.

Материалом для составления выборки служат первичные зоотехнические, ветеринарные, а также экспериментальные данные.

Величина признака у отдельной особи называется вариантой и обозначается буквой x (икс). Величина варианты зависит от многих факторов. Например, суточный удой коровы зависит от генетических факторов, физиологического состояния организма, условий кормления и содержания, климатических факторов и др. При этом возможны многообразные сочетания факторов, положительно и отрицательно влияющих на признак и обусловливающие его разнообразие.

Очень важным вопросом является объем выборки. Определение ее объема зависит от изучаемых вопросов и степени их изученности.

Число особей в выборке обозначается буквой n, в генеральной совокупности – N. Различают многочисленные (большие) и малочисленные (малые) выборки, для которых различны методы

обработки показателей признаков.

Большими называют выборки с численностью 30 особей и более, малыми – численностью менее 30 особей.

Следует отметить условность такого разделения, т. к. при использовании компьютерной техники деление на большие и малые выборки отпадает. При отсутствии счетно-вычислительной техники и наличии многозначных показателей обработка методами больших и малых выборок имеет значение: расчет в больших выборках ведется не прямым способом, а путем группировки.

После составления выборки приступаем к ее изучению.

К числу важнейших показателей, используемых в генетике и зоотехнии, относятся следующие.

I. Средние величины – средняя арифметическая (M), средняя взвешенная ($M_{взв}$), средняя геометрическая (G), средняя гармоническая (H), мода (M_o), медиана (M_e).

II. Показатели разнообразия признака: лимиты (lim), среднее квадратическое отклонение (σ), коэффициент вариации (C_v), нормированное отклонение (t).

III. Показатели связи между признаками: коэффициенты фенотипической и генетической корреляции (r, r_a), регрессии (R_{xy}) и др.

IV. Показатели соответствия выборочных данных параметрам генеральной совокупности, т. е. репрезентативности.

Средние величины – важные биометрические показатели, используемые в науке и практике. Средняя арифметическая является основным показателем, характеризующим совокупность по величине изучаемого признака. Свойства средних величин: срединное расположение между минимальным и максимальным значениями признака, абстрактность и единство суммарного действия.

В зависимости от поставленных целей в биологии используются несколько средних величин: средняя арифметическая, средняя взвешенная,

средняя геометрическая, средняя гармоническая.

Вычисление средней арифметической (M) в малочисленных выборках. Средняя арифметическая величина в малочисленных выборках вычисляется прямым способом, который заключается а суммировании всех вариант последующим делением суммы на число вариант в совокупности (*n*):

$$M = \frac{\sum x_n}{n}$$

Вычисление средней вешенной ($X_{взв}$).

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

$$M_{взв} = \frac{x_1 n_1 + x_2 n_2 + ... + x_s n_s}{n_1 + n_2 + ... + n_s}$$

где $X_{взв}$ – средняя взвешенная; M_1, M_2, ... M_s – средние арифметические первой, второй и т. д. совокупностей; n_1, n_2,..., n_s – вес (объем) этих совокупностей.

Занятие 2

ПОКАЗАТЕЛИ РАЗНООБРАЗИЯ ПРИЗНАКОВ. ДОСТОВЕРНОСТЬ ВЫБОРОЧНЫХ ПОКАЗАТЕЛЕЙ

Цель занятия. Освоение методов вычисления показателей разнообразия признаков и практическое применение их в селекции. Вычисление критерия достоверности, уровня значимости и достоверности разности двух средних величин.

Методические указания. Показателем разнообразия признака в совокупности могут в известной мере служить *лимиты,* которые характеризуют минимальное и максимальное значение изучаемого признака в выборочной совокупности и указывают на амплитуду вариации.

Однако эти показатели недостаточны, так как животные с такими показателями могут быть нехарактерны для данного стада. Кроме того, лимиты не отражают индивидуальных различий внутри выборки, т.е. при одной и той же средней величине группы неоднородны.

Установление степени разнообразия признака в популяциях имеет важное значение в селекции. Наилучшим показателем разнообразия признака является среднее квадратическое отклонение σ, которое учитывает отклонение каждой варианты от средней арифметической.

Вычисление среднего квадратического отклонения в малочисленных выборках (n<30). При небольшом числе вариант среднее квадратическое отклонение вычисляется по формуле:

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

Можно вычислить среднее квадратическое отклонение по данным удоя за 305 дней лактации (таблица).

В первую графу вписывают варианты (удой за 305 дней лактации, ц). Суммировав их и разделив *(n)* число вариант, получают средний удой за

305 дней, ц *(M)*.

Затем вычитают х из каждой варианты и разности (отклонения от средней) и вписывают во вторую графу. Для проверки правильности вычислений суммируют все разности *(x – M)*, их сумма должна быть равна нулю. Далее каждое отклонение возводят в квадрат и вписывают квадраты отклонений $(x—M)^2$ в третью графу. Квадраты отклонений всегда положительны. Суммируя все числа третьей графы, получают сумму квадратов отклонений $\sum(x—M)^2$, которую вписывают в итог третьей графы. Среднее квадратическое отклонение вычисляют по формуле.

Выражение *n—1* называемся числом степеней свободы *(v)*, которое указывает на ограничение, имеющее при вычислении среднего квадратического отклонения одно условие: сигма является показателем разнообразия изучаемого признака для группы, имеющей определенную среднюю арифметическую, поэтому $v = n—1$. Полученная величина σ указывает, что в среднем отклонения вариант данного признака от средней арифметической составляют σ.

Вычисление коэффициента вариации *(Cv)*. Среднее квадратическое отклонение – величина именованная. При изучении удоев она выражается в килограммах, при изучении жирности молока – в процентах, при изучении промеров – в сантиметрах и является показателем признака для группы с определенной средней арифметической величиной. При изучении разнообразия признаков, выраженных в различных единицах измерения (см, кг, % и др.), и при больших различиях средних арифметических величин сравниваемых групп сигма не может быть использована. В таких случаях используют другой показатель – коэффициент вариации (Cv), вычисляемый по формуле:

$$C_V = \frac{\sigma}{M} \cdot 100\%$$

На предыдущих занятиях Мы с Вами рассмотрели биометрические показатели выборочных совокупностей. Можно ли, пользуясь ими, характеризовать генеральную совокупность? Например, можно ли по выборкам, состоящим из 100 или 1000 голов холмогорского скота, судить о продуктивности всего поголовья этой породы?

Характеристика генеральной совокупности на основе выборки, составленной по принципу случайности, будет всегда неточной. Эта неточность всегда возникает, когда целое (генеральная совокупность) характеризуется на основе его части (выборки). Ошибки, возникающие при характеристике генеральной совокупности показателями, полученными при изучении выборки, называются ошибками репрезентативности. Эти ошибки возникают даже при организации исследований на самом высоком уровне. Величину ошибок репрезентативности определяют только для выборочных показателей. Если группа животных выступает как генеральная совокупность, ошибки репрезентативности не вычисляют.

Данные ошибки используются для установления доверительных границ в генеральной совокупности, достоверности выборочных показателей и разности, установления объема выборок при научно-исследовательских работах.

Доверительными границами называют крайние значения, в пределах которых может находиться параметр генеральной совокупности.

Для оценки параметров генеральной совокупности, кроме выборочного показателя, необходимо иметь: величину ошибки (m)— показатель точности оценки генерального параметра.

В биологии используют три основных порога вероятности безошибочных прогнозов (таблица).

Три порога надежности или вероятности безошибочных прогнозов для больших выборок

Порог	Применение	Вероятность безошибочных прогнозов	Число ошибочных случаев	Критерий надежности	Объем"выборок (равен или больше)
	В производственных и научно-производственных исследованиях	0,95	5 из 100	1,96	30
	В большинстве биологических, зоотехниических и ветеринарных исследований	0,99	1 из 100	2,57	100
	В работах с очень высокими требованиями (проверка гипотез, исследования лекарственных и других веществ)	0,999	1 из 1000	3,29	200

Если генеральная совокупность велика, ее приравнивают к бесконечности. В этом случае ошибку выборочной средней арифметической *(M)* вычисляют по формуле:

$$m_M = \pm \frac{\sigma}{\sqrt{n}}$$

где m_M — ошибка средней арифметической *(M);* σ — среднее квадратическое отклонение; n — объем выборки.

Согласно этой формуле, ошибка средней арифметической величины зависит от σ и п (чем меньше разнообразие признака, тем меньше ошибка). При полной однородности совокупности по изучаемому признаку ($\sigma = 0$) средняя ошибка становится равной нулю, т. е. *M* выборки становится равной *M* генеральной совокупности. Величина средней ошибки находится в обратной зависимости от *n.* Чем больше вариант вошло в выборку, тем меньше ошибка выборочной *M.*

В малочисленных выборках она вычисляется по следующей формуле:

$$m_M = \pm \frac{\sigma}{\sqrt{n-1}}$$

В выборке из 10 коров определен средний удой за 305 дней лактации M=41,26 ц, а σ=±6,68 ц. Ошибка средней арифметической в данном случае составит

$$m_M = \pm\frac{\sigma}{\sqrt{\text{n -1}}} = \frac{6{,}68}{\sqrt{10-1}} = 2{,}23$$

Это означает, что средняя ошибка на 10 голов составляет 2,23 кг. Следовательно, удои за 305 дней лактации изучаемой выборки характеризуются $M \pm m_M = 41{,}26 \pm 2{,}23$ кг. Доверительные границы для средней арифметической генеральной совокупности M находятся в пределах: $M \pm t \cdot m_M$. Верхняя граница 41,26+2,23 = 43,49; нижняя – 41,26–2,23=39,03.

Ошибки других выборочных показателей вычисляют по следующим формулам:

среднего квадратичного отклонения:

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

коэффициента вариации

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

коэффициента корреляции

$$m_r = \frac{1-r^2}{\sqrt{n}}$$

коэффициента регрессии

$$m_{\text{R}} = m_r \frac{\sigma_2}{\sigma_1}$$

Величину выборочного показателя записывают с величиной его

ошибки со знаком «±»: M±т, σ±$т_σ$, Cv±m_{cv}, r±m_r и т. д. Достоверность выборочных показателей (±t) определяется отношением выборочного показателя к его средней ошибке по формулам:

$$t_M = \frac{M}{m_M} \qquad t_\sigma = \frac{\sigma}{m_\sigma} \qquad t_{Cv} = \frac{Cv}{m_{Cv}}$$

$$t_r = \frac{r}{m_r}$$

Во многих исследованиях возникает необходимость сравнить средние арифметические двух групп животных (например, среднюю живую массу животных опытной и контрольной групп или среднюю продуктивность дочерей двух производителей и т. д.). Средние двух сравниваемых групп всегда в некоторой мере отличаются друг от друга. Поэтому необходимо установить, достоверна ли разность между средними.

При решении задач такого рода_ определяют разность (d) между двумя средними (M_1 и M_2) путем вычитания: d= M_1 – M_2. Среднюю ошибку разности (t_d) вычисляют по формуле:

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

Достоверность разности (t_d) определяется по формуле:

$$t_d = \frac{d}{m_d} = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}}$$

В таблице три порога надежности приведены три стандартные величины и соответствующие им вероятности. Так, например, при t_d = 2,57 вероятность того, что разность достоверна, составляет 0,99 (т. е. 99%), а

при t_d = 3,2 вероятность достигает 0,999 (99,9%). Если же величина t_d меньше 1,96, то разность между средними сравниваемых групп не может быть признана достоверной.

При сравнении малых выборок вычисление m_d по формуле можно производить только в том случае, если объемы выборок (n) не отличаются резко друг от друга, в других случаях пользуются таблицей Стьюдента.

Занятие 3

Показатели взаимосвязи между признаками.

Коэффициенты корреляции между альтернативными признаками (r_a)

Цель занятия. Понятие корреляции. Освоение методов вычисления показателей связи между признаками и приобретение навыков по использованию этих показателей в селекционной работе и прогнозировании селекции, а также установления связи между альтернативными признаками.

Методические указания. При одновременном изучении совокупности животных по нескольким признакам нередко обнаруживается, что между признаками существует взаимная связь. Например, более крупные животные имеют более высокую продуктивность. Эта взаимосвязь обнаруживается только в совокупности. У отдельных животных при одинаковой живой массе продуктивность может значительно различаться, так как продуктивность зависит не только от размера животных, но и от многих других факторов.

Взаимная связь признаков в их изменении называется корреляцией. По форме корреляция может быть прямолинейной и криволинейной, по направлению – прямой (положительной) и обратной (отрицательной).

При прямолинейной связи равномерным изменениям одного признака соответствуют равномерные изменения второго признака при незначительных отклонениях. Например, при увеличении длины тела на 1 см ширина его тоже увеличивается на определенную величину.

При криволинейной связи с увеличением одного признака другой увеличивается до определенного момента, а затем уменьшается (или наоборот). Например, с увеличением возраста удой увеличивается до 6–7-го отела, а затем у большинства коров снижается. При криволинейной корреляции связь сначала положительная, затем отрицательная – при увеличении первого признака второй, коррелирующий с ним, уменьшается.

Степень связи между признаками измеряется при помощи коэффициентов корреляции (r), тетрахорического показателя, коэффициентов регрессии.

Изучение связи между признаками имеет большое значение при решении генетико-селекционных вопросов. Установление фенотипической и генотипической связи между признаками позволяет вести косвенную селекцию по коррелирующим признакам и используется для прогноза селекции.

Вычисление коэффициента фенотипической корреляции (r) в малочисленных выборках.

$$r = \frac{\sum X \times Y - \frac{\sum X \times \sum Y}{n}}{\sqrt{C_x \times C_y}}, \quad \text{где:}$$

$$C_x = \sum X^2 - \frac{(\sum X)^2}{n}$$

$$C_Y = \sum X^2 - \frac{(\sum Y)^2}{n}$$

Величину C_x вычисляют отдельно: C_x для ряда *x,* C_y для ряда *y,* C_d для ряда разностей между ними *(x–y)*.

Техника вычисления r рассматривается на примере удой за 305 дней лактации дочерей и матерей по данным малой выборки (*n=10*).

Вычисление коэффициента прямолинейной регрессии (*R*).

Коэффициент прямолинейной регрессии указывает, насколько в среднем изменяется один из признаков при изменении другого на единицу измерения. В многочисленных выборках ин вычисляется по формулам:

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

В малочисленных выборках – по формулам:

$$R_{x/y} = \frac{\sum X \times Y - \frac{\sum X \times \sum Y}{n}}{\sum y^2 - \frac{(\sum y)^2}{n}}$$

$$R_{y/x} = \frac{\sum X \times Y - \frac{\sum X \times \sum Y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

Для установления связи между альтернативными признаками строят четырехклеточную корреляционную решетку. Коэффициент корреляции или тетрахорический показатель (**r_a**) вычисляют по формуле:

$$ra = \frac{(ad - bc) - \frac{n}{2}}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

Дочери	Матери		Всего
	Больные	Здоровые	
Больные	a=23	b=119	a+b=142
Здоровые	c=14	d=149	c+d=163
Всего	a+c=27	b+d=268	n=a+b+c+d=600

$$ra = \frac{(23 \times 149 - 14 \times 119) - \frac{600}{2}}{\sqrt{(23+119)(14+149)(23+14)(119+149)}} = 0{,}097$$

Ошибка коэффициента корреляции:

$$m_r = \frac{1 - 0{,}097^2}{\sqrt{600}} = 0{,}04$$

Достоверность коэффициента корреляции

$$t_r = \frac{r}{m_r} = \frac{0{,}097}{0{,}04} = 2{,}4$$

Так как t=2,4 больше табличного значения t_{st}=1,96, но меньше 2,57, следовательно коэффициент корреляции достоверен (p>0,95).

Таким образом, между частотой заболеваемости лейкозом бурой

латвийской породы матерей-коров и их дочерей существует слабая положительная связь, т.е. у больных матерей дочери чаще заболевают лейкозом, чем у здоровых.