Тема 2. Проведение наблюдений

Основные обозначения и термины

X - средняя арифметическая выборки - центр распределения в выборке, вокруг которого группируются все варианты статистической совокупности;

M- средняя величина генеральной совокупности;

mx – ошибка выборочной средней;

d – разность между двумя величинами;

σ- среднее квадратическое отклонение;

Cv – коэффициент вариации;

n – число животных в выборке;

N – число животных в генеральной совокупности;

r - коэффициент фенотипической корреляции;

C, D – дисперсия;

R x/y - коэффициент регрессии - количественное изменение одного признака (функции) в зависимости от изменения второго (аргумента);

l – классовый промежуток;

а – условное отклонение в вариационных рядах;

А – условная средняя вариационного ряда;

W0 - значение начала модального класса;

t – коэффициент достоверности;

lim - границы;

x1, x2…xi - значение данных в выборке;

tst – стандартное значение критерия достоверности Стьюдента;

ν- число степеней свободы;

Р – уровень вероятности;

F – критерий достоверности по Фишеру;

η2 – сила влияния фактора (1) - доля межгрупповой вариации в общем варьировании;

h2 – коэффициент наследуемости (2) - доля генотипической изменчивости в общей (фенотипической) изменчивости;

σ 2 – варианса – квадрат стандартного отклонения;

SD – общий селекционный дифференциал;

SD м – селекционный дифференциал матерей;

SD0 - селекционный дифференциал отцов;

ЭS – эффект селекции;

i – число поколений.

Выборка – часть исследуемой совокупности, по которой делают выводы о распределении признака, справедливые для всей совокупности животных.

Генеральная совокупность – бесконечно большое множество относительно однородных единиц или членов, объединенных по какому-либо признаку.

Критерий достоверности – показывает, удовлетворяют ли выборочные показатели принятой гипотезе.

Нулевая гипотеза – предполагает, что разность между генеральными параметрами сравниваемых групп равна нулю, а различия, наблюдаемые между выборочными характеристиками, носят случайный характер.

Ошибка – разность между результатами измерений и действительно существующими значениями.

Источники ошибок – технические (неисправности и неточности приборов), личные (навыки и мастерство исследователя), случайные.

Вариационный ряд – последовательность показателей признака животных, расположенная в порядке возрастания величин того же признака.

Он обладает рядом закономерностей, которые используются в генетике и селекции животных.

Корреляция – зависимость между переменными величинами.

Наследуемость – доля влияния генотипа на изменчивость признака.

Количественные признаки – признаки, поддающиеся непосредственному измерению. Разделяются на мерные (варьирующие непрерывно) и счетные (дискретные).

Качественные признаки – признаки, не поддающиеся непосредственному измерению и учитываемые по наличию их свойств у отдельных членов изучаемой группы.

Объем выборки – количество отбираемых вариант. Выборка не может содержать менее 2- х вариант.

Дисперсия – сумма квадратов отклонений каждого признака от средней.

Классы вариационного ряда – интервалы, на которые разбивается общая вариация признака в пределах от минимальной до максимальной величины.

Мода – величина, наиболее часто встречающаяся в данной совокупности.

Медиана – средняя, относительно которой вариационный ряд делится на две количественно равные части.

Репрезентативность – мера, в которой выборка представляет генеральную совокупность. По репрезентативной выборке можно судить о генеральной совокупности.

Вероятность – числовая мера возможности осуществления события.

Сила влияния фактора – мера, в которой фактор влияет на формирование признака.

Правило 3-х сигм – при нормальном распределении 99,9% наблюдений входят в цифровой интервал от Х - 3σ до Х + 3σ.

Размах варьирования – разница между максимальным и минимальным значением признака.

Нормированное отклонение – отклонение той или иной варианты от средней, отнесенное к величине среднего квадратического отклонения.

Модальный класс – класс, в котором наблюдается наибольшее количество вариант (особей, наблюдений).

Тема 3. Составление вариационного ряда.

Цель. Знакомство с методом построения вариационного ряда, приобретения навыков их графического изображения и приобретения навыков распознавания характера распределения признаков.

Вариационным рядом называется ряд чисел, показывающий закономерность распределения единиц изучаемой совокупности по ранжированным значениям варьирующего признака. Числа, показывающие, сколько раз отдельные варианты встречаются в данной совокупности, называются частотами или весами вариант и обозначаются p или f. Общая сумма частот всегда равна объему данной совокупности, т.е. $\sum p = n,$ где $\sum$ - суммирование частот вариационного ряда, n – объем выборочной совокупности.

Совокупность числовых значений признака распределяется в безынтервальный или интервальный вариационный ряд в зависимости от того, как варьирует признак – в широком или узком диапазоне. В первом случае частоты распределяются непосредственно по ранжированным значениям варьирующего признака, которые приобретают значение «классов» вариационного ряда, а во втором – частоты распределяются по отдельным интервалам, или промежуткам (от - до), на которые разбивается вариация признака в пределах от минимальной до максимальной варианты совокупности.

Величина равных интервалов определяется делением размаха варьирования признака на число групп или классов (K), намечаемых при построении вариационного ряда:

$$l = \frac{X_{\max} - X_{\min}}{K}, \qquad (1)$$

где l – величина классового интервала, $X_{\max}$ - максимальная, а $X_{\min}$ – минимальная варианты совокупности.

Оптимальное число классов, на которое следует разбить вариацию признака, определяют с помощью таблицы 1.

Таблица 1

Объем выборки (от – до)	Число классов
25-40	5-6
40-60	6-8
60-100	7-10
100-200	8-12
> 200	10-15

Чтобы построить вариационный ряд необходимо:

1. Найти лимиты, т.е. минимальное и максимальное значения варианта.
2. Найти величину классового интервала.
3. Составить классы. К минимальному значению изучаемого признака прибавлять величину классового интервала до тех пор, пока не войдет максимум. Верхние границы классов уменьшают на величину, равную точности, принятой при измерении признака, например, на 1, 0,1,0,01 и т.д., чем и достигается необходимое разграничение классов.
4. Определить величины середин классов (w). Они будут равны полусумме нижних границ данного и следующих классов; можно так же к нижней границе данного класса прибавлять половину классового интервала.
5. Произвести разноску вариант по классам. Для этого составляют таблицу из четырех граф и числа строк по числу классов. В первой графе – границы классов, во второй – середины классов, третья – для учета частот с помощью различных условных знаков, в четвертой – частоте встречаемости вариант в каждом классе (то же, что в графе 3 в цифровом изображении).

При обработке больших выборок удобен следующий шифр частот:

•	• •	• •	• •	• •	• •	• •	• •	• •	• •
		•	• •	• •	• •	• •	• •	• •	• •
1	2	3	4	5	6	7	8	9	10

Если классы вариант (w) выписать в один ряд, а частоты (р) в другой, то получим вариационный ряд, в который входят два ряда цифр, характеризующих классы и частоты. В вариационном ряду существует определенная закономерность. Крайние вариации малочисленны; с приближением к середине ряда частоты вариаций увеличиваются. В середине вариационного ряда или вблизи одна вариация (один класс), на которую приходится наибольшее число частот, ее называют модальной вариацией (модальным классом).

Вариационный ряд, изображенный графически, дает вариационную кривую (диаграмму частот).

При построении графика безынтервального вариационного ряда по оси абсцисс откладываются значения классов, а по оси ординат – частоты. Высота перпендикуляров, восставляемых по оси абсцисс, соответствует частотам классов. Соединяя вершины перпендикуляров прямыми линиями, получают геометрическую фигуру в виде многоугольника, называемую полигоном распределения частот.

При построении графика интервального вариационного ряда по оси абсцисс откладывают границы классовых интервалов. В результате получается столбиковая геометрическая фигура, называемая гистограммой распределения частот. Если из срединных точек вершин прямоугольника гистограммы опустить перпендикуляры на ось абсцисс, гистограмма превращается в полигон распределения. Соединяя точки вершин прямоугольников гистограммы прямыми линиями, получаем вариационную кривую.

При построении графиков вариационных рядов масштабы на осях координаты выбираются произвольно, но с таким расчетом, чтобы высота вариационной кривой относилась к ее основанию примерно как 5:8. Откладывая по оси абсцисс классы вариационного ряда, следует доводить их слева и справа до нулевых классов, которые не содержат ни одной варианты. Несоблюдение этих правил приводит к нежелательным результатам: график получается с остроконечной вершиной или в виде чрезмерно растянутой по ширине уплощенной фигуры. В обоих случаях вариационная кривая оказывается плохо обозримой, недостаточно четко отображающей характерные черты варьирования изучаемого признака.

Задание 1. На основании многолетних клинических наблюдений составлена выборка 100 анализов на содержание кальция (мг%) в сыворотке крови клинически здоровых павианов гамадрилов:

13,6	12,9	12,3	9,9	12,7	11,7	10,8	10,4	10,9	10,2
14,7	10,4	11,6	11,7	12,1	10,9	12,1	9,2	10,1	11,5
13,1	10,9	12,0	11,1	13,5	11,2	13,5	10,1	14,0	10,0
11,6	12,4	11,9	11,4	12,8	11,4	10,9	12,7	13,8	13,2
11,9	10,8	11,0	12,6	10,0	10,3	12,7	11,7	12,1	13,8
12,2	11,9	11,6	10,6	11,1	10,7	12,3	11,5	11,2	11,5

12,7 10,5 11,2 11,9 9,7 13,0 9,6 12,5 11,6 9,0
11,5 12,3 12,8 12,6 12,8 12,5 12,8 11,4 12,5 12,3
14,5 12,3 12,6 11,7 12,2 12,3 11,6 12,0 13,5 12,5
11,6 11,9 12,0 11,4 14,7 11,3 13,2 14,3 13,2 14,2

1. Сгруппировать эти данные в вариационный ряд;
2. Построить гистограмму.

Задание 2. Составить вариационный ряд по следующим данным живой массы коров (кг):

597 673 598 670 657 647 588 646 555 692 635 610
614 650 629 602 584 630 607 652 654 669 503 665
552 685 599 628 655 584 672 550 605 625 645 545
570 644 591 595 664 565 678 540 715 568 688 612
530 660 538 708 535 695 596 675 618 547 638 655
562 571 653 564 648 582 642 559 580 627 567 630
590 576 630 576 630 574 614 586 580 635 610 567
619 633 608 625 522 612 636 604 625 522 612 636
604 625 644 565 617 585 620 658 572 618 634 596
612 603 626 635 611 578 605 595 615 652 615 637
587 601 590 610 592 621 575 606 639 585 512 583

Задание 3. Составить вариационный ряд и построить вариационную кривую по многоплодию 75 янтарь-сапфировых норок (по числу щенков в помете). Составлена следующая выборка:

4 4 2 8 1 6 4 3 4 4 4 6 4 5 2 4 7 4 6 5 6 4 5 4 4
8 4 5 4 4 5 4 3 4 5 4 5 4 4 7 3 4 5 4 5 4 4 3 4 4
4 4 7 5 3 6 4 9 4 6 4 2 6 4 2 4 5 4 4 4 4 3 4 5 7

Задание 4. По данным плодовитости свиноматок построить вариационную кривую, определить ее тип.

Число поросят у свиноматок	4	5	6	7	8	9	10	11	12
Частоты	1	2	8	8	19	27	15	4	1

Контрольные вопросы.

1. Что такое выборки, как они составляются?
2. Дать определение понятиям: вариационный ряд и вариационная кривая.
3. Как составляется вариационный ряд?
4. Какие показатели вариационного ряда характеризуют изменчивость признака?
5. Какие бывают типы распределения и вариационных кривых?

Тема 4. Статистические совокупности

Совокупность, из которой отбирается некоторая часть ее членов для совместного изучения, называется генеральной, а отобранная тем или иным способом часть генеральной совокупности получила название выборочной совокупности или выборки.

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.

Цель. Знакомство с методами вычисления основных биометрических показателей количественных признаков на малых и больших выборках.

Средняя арифметическая – показатель средней величины признака данной группы особей, характеризующий среднюю вариацию этого признака.

Среднее арифметическая – абстрактное число. Если среднее количество зерен в колосьях оказалось равным 5,7, то такое число точно характеризует среднюю величину, хотя в действительности существование 5,7 зерен невозможно.

Средняя арифметическая (если выборка немногочисленна) вычисляется по следующей формуле:

$$\overline{X} = \frac{\overline{x_1} + \overline{x_2} + \overline{x_3} + ... + \overline{x_n}}{n}, \text{ или } \overline{X} = \frac{\sum \overline{x}}{n} \qquad (2)$$

где $\overline{X}$ - средняя арифметическая;

$\overline{x}$ - величина варианты;

n – численность вариант;

$\sum$ (сигма) – знак суммирования (сумма).

Пример. Толщина эпидермиса в коже молодняка овец составляет (мкм): 28; 27,2; 19,6; 25,2; 18,5.

$$\overline{X} = \frac{28 + 27.2 + 19.6 + 25.2 + 18.5}{5} = 23.8$$

Для вычисления средней арифметической для больших выборок (n>30) необходимо оформить вариационный ряд в таблицу, с использованием условной средней по формуле:

$$\overline{X} = A + bK, \qquad (3)$$

где $\overline{X}$ - средняя арифметическая;

A – условная средняя;

b – среднее отклонение от условной средней;

K – величина классового промежутка.

Пример. Требуется вычислить среднее количество поросят на один опорос у 100 свиноматок. Вариационный ряд количества родившихся поросят следующий:

w	p
8	1
9	4
10	19
11	**34**
12	27
13	11
14	4

Сначала из числа вариаций выбирают условную среднюю, обозначаемую буквой A. В качестве таковой обычно берут значение середины того класса, в который входит наибольшее число вариант. В данном случае это будет вариация, выраженная числом 11, около которой ставим букву A.

Затем устанавливается, на какое количество классовых промежутков (в сторону минус или плюс) отклоняется каждый класс от класса, принятого за условную среднюю. Эти отклонения, обозначаемые буквой a, выписываются столбцом, параллельно вариационному ряду; они относятся к разному количеству частот, поэтому для установления среднего отклонения (b) от условной средней необходимо каждую частоту умножить на соответствующее отклонение (p на a) и произведения с тем или иным знаком выписать столбцом параллельно отклонению. Далее производится алгебраическое суммирование произведений частот на отклонение, то есть определяется $\sum$ pa. В данном примере $\sum$ pa =31. Затем определяется величина среднего отклонения от условной средней, приходящаяся на одну варианту. Вышеизложенное можно представить в следующем виде:

	W	p	a	Pa
	8	1	-3	-3
	9	4	-2	-8
	10	19	-1	-19
A	11	34	0	0
	12	27	+1	+27
	13	11	+2	+22
	14	4	+3	+12
n=100				$\sum$ pa = 31

Среднее отклонение от условной средней вычисляет по формуле:

$$b = \frac{\sum pa}{n} \qquad (4)$$

В нашем примере b=0,31.

Теперь вычисляем истинную среднюю арифметическую $\overline{X}$, используя формулу:

$\overline{X}$ =A+bK

$\overline{X}$ =11+0,31x1=11,31.

Взвешенная средняя арифметическая представляет собой результат усреднения средних арифметических нескольких совокупностей. Она вычисляется по формуле:

$$\overline{X} = \frac{\sum \overline{X} n}{\sum n}, \qquad (5)$$

где $\overline{Xn}$ - средняя арифметическая нескольких совокупностей;
n – вес (объем) этих совокупностей.

Пример. Имеются средние арифметические ($\overline{Xn}$) показатели тонины шерсти (мкм) отдельных отар овец и количество овец в этих отарах (n). Требуется вычислить среднюю арифметическую тонины шерсти всех овец.

Вычисление взвешенной средней арифметической тонины шерсти овец производится следующим образом:

№ отар	$\overline{Xn}$	N	$\overline{Xn} \cdot$ n
1	20	210	4200
2	22	150	3300
3	25	240	6000
4	18	100	1800
5	24	300	7200
$\sum$ 1000		$\sum$ 22500	

$\overline{Xn} = \dfrac{22500}{1000} = 22.5$.

Задание 1. На двух птицефабриках число кур несушек составляло 20 000 и 28200 голов, их средняя яйценоскость за год соответственно 294 и 280 штук. Определите среднюю яйценоскость по двум птицефабрикам.

Задание 2. Определите среднюю арифметическую по следующим данным живой массы 60 кобыл адаевской породы:

463	424	573	481	471	425	470	492	480	490
450	490	489	445	520	375	510	400	475	512
449	476	516	460	480	500	530	480	463	490
510	520	430	450	490	460	520	455	480	461
482	451	480	451	480	499	525	475	480	453
519	490	480	491	514	498	490	460	491	475

Контрольные вопросы.

1. Что характеризует средняя арифметическая и как она определяется при большом числе вариантов?
2. В каких пределах колеблется значение средней арифметической генеральной совокупности?
3. Что такое средняя взвешенная? В каких случаях она применяется и как ее вычисляют?
4. Перечислите средние величины и их использование?
5. Какими свойствами обладают средние величины?

Тема 6. Основные статистические показатели.

Цель. Знакомство с методами вычисления основных биометрических показателей количественных признаков на малых и больших выборках.

Основным критерием изменчивости является среднеквадратическое отклонение (σ), которое показывает, насколько в среднем отклоняется по изучаемому признаку каждый член совокупности от средней арифметической этой совокупности. Величина σ всегда именованная (кг, см, % и т.п.) и вычисляется на одну единицу точнее, чем средняя арифметическая.

Определение среднего квадратического отклонения позволяет выявить особенности варьирования признака, если две выборки по значению средних арифметических друг от друга не отличаются. Предположим, что в первой и во второй звероводческих хозяйствах средняя живая масса норок оказалась одинаковой. Анализ среднего квадратического отклонения показывает, что степень генетического разнообразия живой массы норок первого хозяйства в два раза выше, чем во втором. Значит второе хозяйство более однородно по изучаемому признаку.

Крайние величины вариационного ряда – лимиты показывают размах варьирования данного признака. Но крайние варианты не показывают, как распределяются остальные варианты внутри вариационного ряда, то есть лимиты не являются показателями степени варьирования.

Степень варьирования, распределение вариант в вариационном ряду характеризует основной показатель изменчивости (варьирования) – среднее квадратическое отклонение, которое вычисляется по формуле:

$$σ= \pm\sqrt{\frac{\sum D^2}{n-1}}\,, \qquad (6)$$

где σ (сигма)– среднее квадратическое отклонение;

D – центральное отклонение, то есть отклонение варианты от средней арифметической (D=x- $\overline{X}$)

Данная форму применяется если выборка немногочисленна.

Пример. Высший суточный удой у коров двух хозяйств составил (кг): 1 хозяйство – 10, 14, 17, 20, 23, 25, 28, 31, 34, 38, $\overline{X_1}$ = 24 кг; 2 хозяйство – 10, 21, 22, 23, 24, 24, 25, 26, 27, 38 , $\overline{X_2}$ = 24 кг. Вычислить среднее квадратическое отклонение (σ).

1 хозяйство			2 хозяйство		
x	x- $\overline{X}$	(x- $\overline{X}$) ^{2}	x	x- $\overline{X}$	(x- $\overline{X}$) ^{2}
10	-14	196	10	-14	196
14	-10	100	21	-3	9
17	-7	49	22	-2	4
20	-4	16	23	-1	1
23	-1	1	24	0	0
25	+1	1	24	0	0
28	+4	16	25	+1	1
31	+7	49	26	+2	4
34	+10	100	27	+3	9
37	+14	196	38	+14	196
$\sum x = 240$		$\sum (x-\overline{X})^2 = 724$	$\sum x = 240$		$\sum (x-\overline{X})^2 = 420$

$$\overline{X}_1 = \frac{240}{10} = 24\,, \qquad \overline{X}_2 = \frac{240}{10} = 24\,,$$

$$σ_1 = \pm\sqrt{\frac{724}{10-1}} = \pm 8.98 \text{ кг}, \quad σ_2 = \pm\sqrt{\frac{420}{10-1}} = \pm 6.83 \text{ кг}.$$

В первом хозяйстве варьирование высшего суточного удоя сильнее, чем во втором, и σ этого удоя также больше.

При вычислении среднего квадратического отклонения (σ) для многочисленных выборок составляется вариационный ряд и вычисление производится по формуле:

$$σ = \pm K\sqrt{\frac{\sum pa^2}{n} - b^2}\,, \qquad (7)$$

где $\sum$ - знак суммирования;

p – частоты;

a – величина, показывающая на сколько классовых промежутков отстоит данный класс от условной средней;

K – величина классового промежутка;

n – численность варианта.

Вычисление среднего квадратического отклонения производится аналогично вычислению средней арифметической. Для этого требуются те же величины, что и для вычисления $\overline{X}$; дополнительно необходимо произвести определение b^2 и $\sum pa^2$. Поэтому к записанным уже при вычислении $\overline{X}$ столбцам частот (p), отклонений (a), их произведений (Pa) следующим столбцом записываются произведения частот на квадраты отклонения (pa^2). Затем производится их суммирование, то есть определяется $\sum pa^2$.

Среднее квадратическое отклонение показывает степень варьирования (изменчивости): чем больше σ, тем больше изменчивость и, наоборот, чем меньше σ, тем меньше изменчивость изучаемого признака в группе организмов.

Среднее квадратическое отклонение показывает также размах колебания. Обычно этот размах приблизительно равен 3σ, то есть подавляющее количество вариант укладывается в границах ± 3σ от $\overline{X}$.

В вариационному ряду, составленном по значительному количеству достаточно однородных вариант, они располагаются в границах:

$\overline{X}$ ± 1,0 σ…………………………… 68,3% всех вариант
$\overline{X}$ ± 1,5 σ…………………………… 86,6% всех вариант
$\overline{X}$ ± 2,0 σ…………………………… 95,5% всех вариант
$\overline{X}$ ± 2,5 σ…………………………… 98,8% всех вариант
$\overline{X}$ ± 3,0 σ…………………………… 99,7% всех вариант

Среднее квадратическое отклонение дает возможность судить о характере отдельных вариант. Если какая-либо варианта отклоняется от $\overline{X}$ несколько дальше ± 3σ, то большая вероятность того, что эта варианта (особь) принадлежит к другому вариационному ряду, то есть к другой качественной категории.

Задание 1. Вычислить среднее квадратического отклонение живой массы бычков при рождении по следующим данным:

45 47 44 36 56 45 40 33 45 46 32 46 32 46 42 42
49 38 46 48 38 40 40 45 49 45 50 40 49 43 37 46
37 43 44 43 39 45 45 37 47 52 60 34 40 39 54 44
43 42 44 45 50 53 38 44 40 38 43 41 37 44 45 41
43 40 42 37 47 31 51 48 50 46 59 43 45 47 46 50
36 37 44 41 36 36 38 43 38 40 52 40 44 52 46 61
46 38 38 45 46 40 45 50 41 45 40 37 45 46 32 55
45 45 40 37 53 50 45 44 50 50 40 48 48 45 32 36

Задание 2. Определить среднюю арифметическую, среднее квадратическое отклонение по данным следующей выборки суточного прироста, г:

691 587 722 812 573 750 700
660 520 640 650 750 630 650

Задание 3. В хозяйстве было 1500 норок с жемчужной окраской и 2100 с коричневой. Определить величину среднего квадратического отклонения по жемчужной окраске.

Контрольные вопросы.

1. Как вычисляют среднее квадратическое отклонение в малых выборках?
2. Можно ли определить максимальное и минимальное значение изучаемого признака, если известна величина средней арифметической и среднего квадратического отклонения?
3. Какие показатели характеризуют разнообразие признаков?

Тема 7. Корреляционный анализ.

Цель. Знакомство с методами вычисления коэффициента генетической корреляции

Если необходимо определить наследование количественных признаков и их генетическую обусловленность, то применяют метод определения генетического коэффициента корреляции (r_G) между признаками, разработанный Л.Хейзелем.

Генетическая корреляция указывает на изменение вторичных признаков при селекции первичных признаков и может быть вычислена при наличии родственных групп матерей и дочерей, отцов и сыновей, полусибсов, полных сибсов и близнецов.

Суть метода заключается в том, что на родственных группах вычисляют четыре коэффициента корреляции между двумя разными фенотипическими признаками (x и y) в пределах каждой сопоставляемой родственной группы и между группами. В результате получения четырех величин r определяют генетический коэффициент связи между признаками x и y, используя следующую формулу:

$$r_G = \sqrt{\frac{r_{xy^1} \cdot r_{yx^1}}{r_{xx^1} r_{yy^1}}} \qquad (49)$$

где x,y – фенотипическое выражение двух признаков у дочерей; $x^/$, $y^/$ - фенотипическое выражение этих же признаков у матерей; $r_{xy}^/$, $r_{yx}^/$ - коэффициенты фенотипических корреляций; $r_{xx}^/$, $r_{yy}^/$ - коэффициенты фенотипических корреляций между одним и тем же признаком у дочерей и матерей.

Формулу применяют для тех случаев, когда оба r в числителе имеют знак «+» или знак «-», но в знаменателе оба должны быть положительными. Если же под корнем в числителе один из r имеет знак «-», а другой - знак «+», то формула видоизменяется:

$$r_G = \frac{(r_{xy^1} + r_{yx^1}) \div 2}{\sqrt{r_{xx^1} \cdot r_{yy^1}}} \qquad (50)$$

Наличие отрицательной связи между $x_д$ и $x_м$ или между $y_д$ и $y_м$ указывает на сильное взаимодействие генотипа со средой или на сложный тип наследования, и, следовательно, по формуле Хейзеля нельзя выявить связь, так как формула основана на предположении о наличии аддитивного действия генов коррелирующих признаков.

Пример. При вычислении r_G на основании данных фенотипических корреляций получены следующие коэффициенты фенотипических корреляций у кур: между живой массой дочерей в 32-недельном возрасте и годовой яйценоскостью матерей $r_{xy^/}$ = +0,092; между живой массой матерей в 32-недельном возрасте и годовой яйценоскостью дочерей $r_{x^/y}$ = +0,164; между живой массой дочерей и матерей в 32-недельном возрасте r_{xx} = +0,410; между годовой яйценоскостью дочерей и матерей $r_{yy^/}$ = +0,340.

$$r_G = \sqrt{\frac{0.092 \cdot 0.164}{0.41 \cdot 0.34}} = \sqrt{0.1082} = +0.328.$$

Задание 1. Необходимо вычислить генетический коэффициент корреляции между содержанием жира и белка в молоке коров-дочерей и их матерей.

Содержание белка (x)		Содержание жира (y)	
Дочери (x)	Матери ($x^/$)	Дочери (y)	Матери ($y^/$)
3,1	3,0	4,0	3,9
3,3	3,1	4,2	4,0
3,2	3,2	4,1	4,0
3,0	3,1	4,0	3,8
3,4	3,3	4,5	4,2

Задание 2. Вычислить коэффициент генетической корреляции между относительной массой белка и яйценоскостью кур русской белой породы, если коэффициенты фенотипических корреляций равны:

$r_{x^/x}$=+0.42; $r_{x^/y}$=-0.23; $r_{y^/y}$=+0.65; $r_{y^/x}$=-0.28.

где $x^/$, x – относительная масса белка у матерей и дочерей; $y^/$, y – яйценоскость матерей и дочерей; $x^/$, y – относительная масса белка у матерей и яйценоскость дочерей; $y^/$, x – яйценоскость матерей и относительная масса белка у дочерей.

Какие выводы можно сделать о характере наследования указанных признаков? В каком случае наиболее эффективно их сочетание?

Контрольные вопросы.

1. Что такое генетическая корреляция?
2. Какими причинами может быть обусловлена генетическая корреляция?
3. Приведите формулы вычисления коэффициентов генетической корреляции.

- Определить коэффициент регрессии, сделать выводы по полученным данным.

Тема 8. Регрессионный анализ, дисперсионный анализ

Цель. Знакомство с методами вычисления коэффициента регрессии и построения ряда регрессии. Знакомство с методами вычисления дисперсии однофакторного комплекса

Дисперсионный комплекс называется однофакторным, если испытывается действие на признак одного регулируемого фактора. Однофакторные дисперсионные комплексы могут быть равномерными и неравномерными. Независимо от этого техника дисперсионного анализа однофакторных комплексов сводится главным образом к расчету показателей варьирования.

Регрессией называется изменение функции (зависимого признака) в зависимости от изменения аргумента.

Метод регрессивного анализа имеет большое значение в изучении корреляционных связей.

Зная степень зависимости величины изменчивости одного признака от изменчивости другого и зная величину этого другого признака, можно с известной долей вероятности предсказать первый признак, непосредственно величину этого признака не определяя.

Показателями регрессии являются: коэффициент регрессии, ряд регрессии, линия регрессии и др.

Показатели регрессии выражают корреляцию связь двусторонне, учитывая изменение средней величины $\overline{y}_x$ признака У при изменение значений x_i признака Х, и, наоборот, показывают изменение средней величины $\overline{x}_y$ признака Х по измененным значения y_i признака У. исключение составляют временные ряды, показывающие изменение признаков во времени. Регрессия таких рядов является односторонней.

Ряды регрессии дают наглядное представление о форме и тесноте корреляционной связи между признаками, в чем и заключается их ценность. Форма связи между биологическими признаками может быть разнообразной. Задача состоит в том, чтобы любую форму корреляционной связи выразить уравнением определенной функции, что позволяет получать нужную информацию о корреляции между переменными величинами У и Х, предвидеть возможные изменения признака У на основе известных изменений Х, связанного с У корреляционно.

Дисперсия – это варьирование или изменчивость признака, возникающая под влиянием различных факторов. Эти факторы действуют на организм животного или растения независимо друг от друга и с различной силой, а иногда даже и в различных направлениях, В результате такого воздействия варьирующий признак приобретает какую-то определенную величину изменчивости.

Признаки, изменяющиеся под воздействием тех или иных причин, называются результативными, а причины, вызвавшие изменение величины результативного признака или признаков, - факторами. Каждый регулируемый фактор испытывается серийно, т.е. в виде нескольких обособленных друг от друга групп, называемых градациями. Факторы обозначаются заглавными буквами латинского алфавита А, В, С, …, а учитываемые признаки – через Х, У, Z, … Градации обозначаются теми же буквами, которыми обозначаются факторы. Например, градации фактора А обозначаются через $А_1$, $А_2$, $А_3$ и т.д. Числа градаций того или иного фактора определяются условиями опыта. Результативные признаки тоже могут подразделяться на отдельные градации, на которых испытывается действие регулируемых факторов.

Дисперсионный анализ позволяет учитывать не только совместное действие регулируемых факторов, но и действие каждого из них в отдельности, а также действие различных комбинаций этих факторов на результативный признак.

Доля воздействия различных, одновременно действующих факторов неодинакова, часто требуется выявить долю воздействия каждого фактора на изменчивость признака.

Изменчивость, вызываемая всеми одновременно действующими факторами, называется общей дисперсией признака. Общая дисперсия ($С_y$) может быть разложена: на дисперсию, возникающую под влиянием различных учтенных факторов, называемую факториальной ($С_x$), и дисперсию, возникающую под влиянием различных случайных (неучтенных) факторов, называемую остаточной ($С_z$).

При дисперсионном анализе вычисляют величину общей ($С_y$), факториальной ($С_x$) и остаточной ($С_z$) дисперсии. Если изменчивость возникает под влиянием нескольких факторов (возраста, живой массы, продолжительности плодоношения, кормления и т.д.) и требуется определить долю влияния каждого из этих учтенных факторов, то в таком случае факториальная дисперсия ($С_x$) может быть разделена на дисперсии каждого фактора раздельно ($С_A$, $С_B$ и $С_C$) и совместную дисперсию.

При дисперсионном анализе обработке подвергаются выборочные данные, оформленные в статистические комплексы. Выборка может быть малочисленной и многочисленной. Статистические комплексы, в зависимости от того, сколько факторов включено в каждый, могут быть однофакторными, двухфакторными, трехфакторными и с большим числом факторов. Статистические комплексы различаются между собой еще и по соотношению частот в классах факторов, входящих в них. Комплексы, имеющие больше одного фактора, могут быть равномерными, пропорциональными и неравномерными.

Техника расчетов дисперсии для малочисленных и многочисленных выборок не одинакова.

Коэффициент регрессии показывает, на сколько увеличивается одна величина (х) при изменении другой коррелирующей с ней величиной (у) на определенное число.

Коэффициент регрессии – величина именованная, вычисляется по формуле:

$$R_{xy} = r\frac{\sigma_x}{\sigma_y} \text{ и } R_{yx} = r\frac{\sigma_y}{\sigma_x} \qquad (52)$$

Пример. Требуется определить: 1) на сколько килограммов изменится живая масса свиноматок украинской степной породы при изменении обхвата груди на 1 см; 2) на сколько сантиметров увеличится (уменьшится) обхват груди свиноматок при увеличении (уменьшении) их живой массы на 1 кг. Связь между этими признаками равна: r=0,7; σ_y=29,3; σ_x=10,6.

Подставим эти значения в формулу коэффициента регрессии. Регрессия живой массы по обхвату груди:

$R_{xy} = 0,7 \cdot \frac{29,4}{10,6} = 1,94$; итак, с увеличением обхвата груди на 1 см живая масса свиноматок увеличится на 1,94 кг.

Регрессия обхвата груди по живой массе равняется:

$R_{xy} = 0,7 \cdot \frac{10,6}{29,4} = 0,25$; то есть с увеличением живой массы на 1 кг обхват груди увеличится в среднем на 0,25 см.

Ряд регрессии – это двойной ряд цифр: 1-й ряд – значение аргумента, 2-й ряд – соответствующее значение функции коррелирующего признака.

Эмпирический ряд регрессии строят по конкретным данным корреляционной решетки, не прибегая к вычислению коэффициента регрессии. Для этого определяют эмпирические ряды одного признака (функции) по классам другого признака (аргумента).

Пример. Составление эмпирического ряда регрессии живой массы по обхвату груди и обхвата груди по живой массе у свиноматок украинской степной породы по следующим данным (таблица 5.1.1).

Таблица 5.1.1 Вычисление эмпирических рядов регрессии между живой массой и обхватом груди у свиноматок

Класс по живой массе (x)	Класс по обхвату груди (y)								p_x	X_{yx}
	130-135	136-141	142-147	148-153	154-159	160-165	166-171	172-177		
180-199	2								2	133
200-219		5	5						10	142
220-239	3	20	19	13	3	5			63	147
240-259		7	10	10	4	2	3		36	150
260-279		2	2	5	7	4	1		21	155
280-299				4	2	1	1	1	9	158
300-319						2	2	2	6	169
320-339								3	3	175
p_y	5	34	36	32	16	14	7	6	150	
X_{xy}	214	233	235	250	260	260	276	316		

Чтобы составить эмпирический ряд живой массы необходимо середину класса по живой массе (w_1) умножить на соответствующие частоты (p_{xy}), стоящие в классе обхвата груди, затем суммировать эти произведения и разделить на число частот в этом классе обхвата груди. В нашем примере в классе обхвата груди 130-135 имеется частота 2, стоящая в классе живой массы 180-199, и частота 3 – в классе живой массы 220-239; умножаем середину 1-го класса 190 на частоту 2 и середину 2-го класса 230 на частоту 3, суммируем эти произведения и делим на число частот, получим среднюю живую массу этого класса.

$$X_{xy} = \frac{\sum (w_x p_{xy})}{p_y} = \frac{190 \cdot 2 + 230 \cdot 3}{5} = 214.$$

Для второго класса с обхватом груди 136-141 средняя живая масса будет равна:

$$X_{xy} = \frac{210 \cdot 5 + 230 \cdot 20 + 250 \cdot 2}{34} = 233;$$

Для 3-го класса X_{xy}=235;

Для 4-го класса X_{xy}=250;

Для 5-го класса X_{xy}=260;

Для 6-го класса X_{xy}=260;

Для 7-го класса X_{xy}=276;

Для 8-го класса X_{xy}=316.

Полученный ряд из значений X_{xy} дает эмпирический ряд регрессии живой массы по обхвату груди; записывается он строчкой внизу в соответствующих классах (таблица 5.1.1).

Эмпирический ряд обхвата груди по живой массе составляется аналогично предыдущему:

$$X_{yx} = \frac{\sum(w_y p_{yx})}{p_x}.$$

Полученные значения X_{yx} дают эмпирический ряд регрессии обхвата груди по живой массе, записывается он колонкой справа в соответствующих классах (таблица 5.1.1).

Для получения теоретического ряда регрессии пользуются формулой уравнения регрессии:

$y-X_y=R_{yx}(x-X_x)$, или $y= R_{yx}(x-X_x)+X_y$;

$x-X_x=R_{xy}(y-X_y)$, или $x= R_{xy}(y-X_y)+X_x$.

Пример. Составление теоретического ряда регрессии живой масс (х) по обхвату груди (у) и обхвата груди по живой массе у свиноматок по следующим данным:

Живая масса: X_x=247,4 кг; R_{xy}=1,94;

Обхват груди: X_y=149,8 см; R_{yx}=0,25.

Теоретический ряд регрессии живой массы по обхвату груди при обхвате груди (см): у=140, 150, 160, 170.

Для 1-го значения: х=1,94 (140-149,8)+247,4=228,4;

Для 2-го значения: х=1,94 (150-149,8)+247,4=247,6;

Для 3-го значения: х=1,94 (160-149,8)+247,4=267,2;

Для 4-го значения: х=1,94 (170-149,8)+247,4=286,6.

Теоретический ряд регрессии обхвата груди по живой массе при живой массе (кг): х=200, 220, 240, 260, 280.

Для 1-го значения: у=0,25(200-247,4)+149,8=138;

Для 2-го значения: у=0,25(220-247,4)+149,8=143;

Для 3-го значения: у=0,25(240-247,4)+149,8=148;

Для 4-го значения: у=0,25(260-247,4)+149,8=153;

Для 5-го значения: у=0,25(280-247,4)+149,8=158.

Задание 1. Вычислить коэффициенты корреляции и регрессии между густотой (х) и длиной шерсти (у) у десяти овец южноказахский меринос по следующим данным:

Х, шт.- 491 502 526 429 438 410 390 394 360 400

У, шт.- 5,5 10,0 6,6 8,0 7,7 8,0 8,4 9,0 6,0 11,1

Задание 2. При изучении зависимости между содержанием жира (%) и массой зерен (мг) у овса результаты оказались следующими:

Классы по содержанию

жира в зернах (х)……4,5 - 5,0–5,5–6,0–6,5 – 7,0 – 7,5 – 8,0–8,5

Средняя масса зерен..45,0 45,8 44,3 41,9 40,1 39,0 37,5 37,5.

Найдите усредненные, или выравненные, значения членов ряда.

Для вычисления общей дисперсии (C_y) пользуются следующей формулой:

$C_y=\sum v^2 - H$, H – промежуточная величина, она равняется: $H = \frac{\sum(v^2)}{n}$. (53)

Остаточную дисперсию (C_z) вычисляют по формуле:

$C_z= \sum v^2 - \sum h_x$, где $\sum h_x = \frac{\sum(v\ )^2}{n}$. (54)

Факториальную дисперсию (C_x) вычисляют по формуле:

$C_x=\sum h_x - H$. (55)

Пример. Влияние возраста матерей на живую массу телят при рождении. Порядок вычисления и необходимые данные приведены в таблице 6.1.1

Таблица 6.1.1 Обработка однофакторного комплекса при малой выборке

Показатели	Полновозрастные матери	Возраст матерей 31-36 мес.	Возраст матерей 25-30 мес	∑
v (живая масса	35, 36, 40, 38, 43, 42	38, 32, 40,	35, 37, 30,	609

при рождении)		34, 35, 31	31, 32	
v^2	1225, 1296, 1600, 1444, 1849, 1764	1444, 1024, 1600, 1156, 1225, 961	1225, 1369, 900, 961, 1024	22067
N	6	6	5	17
$\sum v$	234	210	165	609
$(\sum v)^2$	234^2=54756	44100	27225	-
$\sum h_x = \frac{\sum(v\ )^2}{n}.$	9126	7350	5445	21921
$\overline{X} = \frac{\sum v}{n}$	39	35	33	35,8

Для вычисления промежуточной величины Н используются сводные показатели таблицы 6.1.1, строчки 4 и 3.

$$H = \frac{\sum(v^2)}{n}. = \frac{609^2}{17} = 21817.$$

Дисперсии C_y, C_x, C_z вычисляются по вышеприведенным формулам, подставляя в них данные из таблицы:

C_y=22067-21817=250;

C_x=21912-21817=104;

C_z=22067-21921=146.

Проверка правильности подсчетов производится суммированием: C_y=C_x+C_z, т.е. 104+146=250. В данном случае подсчеты сделаны правильно.

Степень (доля) влияния разных факторов на варьирующий признак определяется отношением между дисперсиями C_x и C_y, C_z и C_y; обозначают эти отношения через η^2. Так, доля влияния учтенных факторов равняется $\eta^2_x = \frac{Cx}{Cz}$, а для неучтенных факторов $\eta^2_z = \frac{Cz}{Cy}$.

В нашем примере доля учтенных факторов равняется:

$\eta^2{}_x = \frac{104}{250} = 0{,}415,$ или 41,5 %.

Доля неучтенных факторов равняется:

$\eta^2{}_z = \frac{146}{250} = 0{,}585,$ или 58,5 %.

Достоверность факториальной дисперсии, то есть достоверно ли влияние и доля влияния фактора на изменчивость признака определяется коэффициентом достоверности Фишера (F). Для вычисления коэффициента Фишера необходимо определить число степеней свободы (ν) и корректированную дисперсию – девиату (σ^2).

Число степеней свободы для факториальной дисперсии (C_x) равно числу классов (l) по фактору минус единица.

$\nu_x = l_x$-1; в нашем примере =3-1=2.

Для остаточной дисперсии (C_z) число степеней свободы равно численности выборки (n) минус число классов (l).

ν_z = n- l_x; в нашем примере =17-3=14.

Число степеней свободы для общей дисперсии (C_y) равно численности выборки (n) без единицы.

ν_z = n-1; в нашем примере =17-1=16.

Корректированную дисперсию или девиату (σ^2) (факториальную и остаточную) вычисляют делением дисперсии на соответствующее число степеней свободы.

Факториальная девиата равна: $\sigma^2_x = \frac{C_x}{\nu_x}$; в нашем примере $\sigma^2_x = \frac{104}{2} = 57.$

Остаточная девиата равна: $\sigma^2_z = \frac{C_z}{\nu_z}$; в нашем примере $\sigma^2_z = \frac{146}{14} = 10.4$.

Коэффициент достоверности Фишера вычисляется делением факториальной корректированной дисперсии (девиаты) на остаточную корректированную дисперсию.

$F = \frac{\sigma^2_x}{\sigma^2_z}$; в нашем примере F=5,5.

Вычисленное значение F сравнивают с табличным значением F. Табличное значение F для данного примера при трех уровнях вероятности равно:
$F_{0,95}=3,7$; $F_{0,99}=6,5$; $F_{0,999}=11,8$.
В нашем примере вычисленное F равно 5,5, следовательно влияние возраста матерей на живую массу телят при рождении достоверно, при уровне вероятности p=0,95.
Задание 1. На четырех разновозрастных группах мужчин измерялась скорость кровотока в сосудах в 1 с. Результаты оказались следующие:

Возрастные группы мужчин	Варианты опыта (проб)			Средние ($\overline{X_i}$)
	1	2	3	
Первая	7	10	12	9,67
Вторая	9	7	14	10,00
Третья	11	16	20	15,67
четвертая	15	18	17	16,67

Определите, достоверны ли расхождения между средними показателями этих групп.
Задание 2. При изучении влияния светового режима на развитие гусениц дубового шелкопряда оказались следующие результаты наблюдений:

Варианты опыта	Количество гусениц к началу выкормки	Выживаемость гусениц на протяжении 5 дней					Средние групп
		1	2	3	4	5	
Контроль	150	9	8	7	8	17	9,8
Полная темнота	150	10	10	9	8	17	10,8
Свет 4 ч	150	9	8	8	9	16	10,0
Свет 8 ч	150	8	7	8	7	15	9,0
Свет 12 ч	150	9	8	8	7	17	9,4

Какой вывод следует сделать на основании этих данных?
Контрольные вопросы.

Контрольные вопросы.

1. Что означает коэффициент регрессии?
2. В чем различие между коэффициентами R_{xy} и R_{yx}?
3. В чем различие между составлением эмпирического ряда регрессии от теоретического?

4 .В чем заключается цель дисперсионного анализа?
5. Что называется общей, факториальной и остаточной дисперсией?
6. Какие бывают дисперсионные комплексы?

Министерство образования и науки Республики Казахстан

Павлодарский государственный университет им. С. Торайгырова

Агротехнологический факультет

Кафедра Зоотехнологии, генетики и селекции

МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ И УКАЗАНИЯ

К практическим занятиям

по дисциплине Биометрия

для студентов специальности 050607 «Биология»

Павлодар

ождения методических
р и указаний; методических
; методических указаний

Ф СО ПГУ 7.18.3/41

УТВЕРЖДАЮ
Проректор по УР
__________ Пфейфер Н.Э.
«___»____________20__г.

Составитель: _________ к.с/х.н. профессор Бурамбаева Н.Б.

Кафедра Зоотехнологии, генетики и селекции

Методические указания

К практическим занятиям

по дисциплине Биометрия

для студентов специальности 050607 «Биология»

Рекомендовано на заседании кафедры
21.10.2010г., протокол № 3

Заведующий кафедрой _________ Бурамбаева Н.Б. «____» ________20__г.

Одобрена учебно-методическим советом Агротехнологического факультета

09.11. 2010г., протокол № 3
Председатель УМС ___________ Жагипарова М.Е. «____» ________2010г.

ОДОБРЕНО:
Начальник ОПиМОУП _____________ Варакута А.А. «____» ________20__г.

Одобрена учебно-методическим советом университета

10.11.2010 Протокол № 2