МИНИСТЕРСТВО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Математические методы в зоологии

Учебно-методическое пособие
по специальности 011600 – «Биология»

Воронеж 2003

Утверждено научно-методическим советом биолого-почвенного факультета 27 октября 2003 г., протокол № 17.

Составитель Пантелеева Н.Ю.

Учебно-методическое пособие подготовлено на кафедре экологии и систематики беспозвоночных животных биолого-почвенного факультета Воронежского государственного университета.

Рекомендуется для студентов 3 курса биологического отделения дневной формы обучения.

Содержание

Введение

В зоологических работах большое место занимают вопросы статистической обработки материала, которые включают как анализ опытных данных, так и оценку результатов исследований.

В России и за рубежом издан целый ряд учебников, пособий по математической статистике, но большинство из них доступно лишь для узких специалистов. В определенной мере негативной чертой является то, что они не указывают корректность применения тех или иных математических методов для обработки результатов наблюдений и не дают биологической трактовки многих статистических параметров в том объеме, который требуется для объективизации биологических данных.

Настоящее пособие рассчитано на студентов и аспирантов, занимающихся систематикой, фауной, экологией и другими направлениями в зоологии. В пособии приводятся все основные методы обработки зоологических данных и на примерах рассматривается применение элементов вариационной статистики, корреляционного, регрессионного, дисперсионного, многофакторного и кластерного анализов.

Тема 1. Вариационный ряд

Занятие 1. **Параметры вариационного ряда**

Вариационным рядом называется ряд числовых параметров – вариант (V), характеризующих признаки из одной выборки.

Ранжировка вариант представляет собой выстраивание от максимальной до минимальной величины (нисходящий) или наоборот (восходящий) ряды.

Основные выборочные показатели:

Средняя арифметическая – $\mathbf{x} = \dfrac{\Sigma\, V}{n}$;

Среднее квадратичное отклонение (**σ**) – $\sigma = \sqrt{\dfrac{\Sigma\,(V-x)^2}{n}}$;

Простое отклонение (**а**) – $a = V - x$;

Варианса (**σ** ^{2}) - $\sigma^2 = \dfrac{\Sigma\,(V-x)^2}{n}$;

Коэффициент вариации (**CV**) – $CV = \dfrac{\sigma}{x}\ 100\%$.

Занятие 2**. Типы распределения признаков**

1. Нормальное распределение – распределение признаков, когда средние параметры выборки имеют наибольшую частоту встречаемости.

Показатели распределения:

Асимметрия – $As = \dfrac{\Sigma\,(V-x)^3}{n\,\sigma^3}$;

Эксцесс – $Ex = \dfrac{\Sigma\,(V-x)^4}{n\,\sigma^4} - 3$.

2. Биноминальное распределение – распределение признаков, когда два средних параметра выборки имеют наибольшую частоту встречаемости.
3. Случайное распределение (распределение Пуассона) – распределение признаков, когда минимальное или максимальное значение вариационного ряда имеет наибольшую частоту встречаемости.

Статистические ошибки – допустимые для определенного уровня точности исследований отклонения переменных вариант.

$$m_x = \frac{\sigma}{\sqrt{n}}; \quad m_\sigma = \frac{\sigma}{\sqrt{2n}}; \quad m_{CV} = \frac{CV}{\sqrt{2n}}.$$

Нормированное отклонение – объективная величина, дающая возможность сравнивать вариационные ряды с разными единицами исчисления.

$$T_M = \frac{V - x}{\sigma};$$

Выскакивающие значение в вариационном ряду – **T** -

$$T = \frac{V - x}{\sigma} \rangle\rangle Tst;$$ если T › Tst – артефакт отбрасывается.

Задание 1.

Имеется два вариационных ряда по признакам, свойственным одному объекту исследований:

1. V: 4.8, 4.9, 5.1, 5.6, 5.7, 5.9, 6.0.
   p: 3, 3, 4, 18, 10, 6, 2
2. V: 12.7, 10.8, 9.4, 10.2, 11.4, 10.6, 10.2.
   p: 2, 10, 3, 4, 6, 18, 3

Оба ряда именованные, с разными единицами измерения. Сравнимы ли эти ряды по типу распределения и степени вариабельности признаков?

Задание 2.

Какой из признаков может быть применен в диагностике вида: А – размеры тела (мм); В – вес (г):

А: V: 7 8 11 12 14 17 19
p: 18 24 12 9 7 3 2

В: V: 7.5 7.6 7.7 7.8 7.9 8.0 8.2
p: 21 16 12 10 7 4 2

Определите тип распределения признаков в обоих случаях.

Задание 3.

Сравниваются два признака: А – длина крыла у Syrphus corelli; В – длина усика у того же вида. Какой признак может быть применен в систематике? Каков тип распределения признаков?

А: V: 2.4 2.5 2.6 2.7 2.8 2.9 3.0
p: 17 10 8 6 4 2 2

В: V: 0.7 0.8 0.9 1.0 1.1 1.2
p: 4 6 12 17 6 3

Тема 2. Достоверность исследований

Занятие 3. **Методы определения достоверности исследований**

Достоверность исследований в вариационной статистике определяется чаще всего следующим образом:

1. Критерий Стьюдента – $\mathbf{t} = \dfrac{\boldsymbol{\sigma}}{\mathbf{m}_{\sigma}}$;

2. Критерий Фишера – $\mathbf{F} = \dfrac{\sigma^2_1}{\sigma^2_2}$; если F › F st при P = 0.01-0.05, то различие между рядами достоверно; если F ‹ F st, то это один вариационный ряд.

3. Хи-квадрат (χ^2) – критерий соответствия или согласия Пирсона

$$\chi^2 = \Sigma \frac{(V-A)^2}{A}$$; где – V – фактическое, A –теоретическое число.

Нулевая гипотеза – гипотеза, подразумевающая, что различий между двумя рядами нет или имеется соответствие установленным биологическим закономерностям. В зависимости от уровня достоверности полученный χ^2 сравнивается с табличным значением, и если χ^2 › χ^2_{st}, то нулевая гипотеза отбрасывается.

Число степеней свободы определяется возможностями вариационного ряда и количеством операций, производящихся с ним. Таким образом, $df_1 = n - 1$; $df_2 = N - r$.

Задание 1.

Для трех видов жуков определены показатели веса

1. V: 4.8, 4.9, 5.2, 5.7, 6.1, 6.4, 6.7
   p: 2 3 5 7 4 1 1
2. V: 5.1, 5.7, 6.0, 6.2, 6.4, 6.8, 7.3
   p: 3 4 7 8 12 16 21.
3. V: 12.7, 12.8, 13.0, 13.4, 13.5, 13.7, 13.9
   p: 3 7 8 10 12 15 17.

Какой из видов отличается наибольшей вариабельностью признаков? Подтвердите достоверность выводов.

Задание 2.

Во время учета жужелиц на полях в окрестностях г. Воронежа в 1998-1999 гг. было собрано 112 самок и 70 самцов, а в Воронежском госзаповеднике –174 самки и 130 самцов.

Нарушена ли популяционная характеристика в одном из обследованных пунктов? Подтвердите достоверность выводов

Задание 3.

Сколько популяций махаона обитает на севере Воронежа области, если известна выборка за 1988-1991 гг.:

Нижнедевицкий район: 7, 9, 3, 12, 4, 5.

Новоусманский район: 4, 12, 10, 6, 3, 3.

Верхнехавский район: 3, 2, 2, 5, 3, 4.

Насколько достоверно заключение? Подтвердите достоверность выводов.

Тема 3. Методы обработки большой выборки

Занятие 4. Большая выборка и классирование признаков

В биологических исследованиях приходится нередко сталкиваться с большими массивами данных, т.е. оперировать большой выборкой.

К такого рода выборкам относятся вариационные ряды, насчитывающие более 30 вариант. В этом случае удобнее работать с классированным материалом, т.е. разбитым на классы.

Пример разбивки на классы со средней по классу (Мо), частотами встречаемости (fx), простыми отклонениями от условно среднего класса (а) приведен в таблице 1.

Интервал класса (i) подразумевает все значения вариант, входящие в группу признаков от min до max в пределах одного класса = 0.4, т.е. крайние варианты отличаются от средней арифметической в 2.24 раза.

Таблица 1.

Классы	Мо	fx	a	pa	pa^2
7.1-7.5	7.3	1	-4	-4	16
7.6-8.0	7.8	3	-3	-9	27
8.1-8.5 8.6-9.0	8.3 8.8	6 13	-2 -1	-12 -13	24 13
9.1-9.5	9.3	17	Усл. ср. класс: А=9.3		
9.6-10.0	9.8	15	+1	+15	15
10.1-10.5	10.3	4	+2	+8	16
10.6-11.0	10.8	0	+3	0	0
11.1-11.5	11.3	2	+4	+8	32
11.6-12.0	11.8	2	+5	+10	50
Всего		63		+3	198

ТЕМА 4. Корреляционный анализ

Занятие 5. **Основы корреляционного анализа**

Корреляционный анализ позволяет установить зависимость между вариациями двух или большего количества признаков и определить, изменяются ли эти признаки самостоятельно или зависимо друг от друга. Этот анализ дает возможность определить количественно выраженную связь между переменными и определить ее достоверность. Главным показателем корреляционного анализа является коэффициент корреляции – r. Обязательным условием корреляционного анализа является то, что изменения признаков должны быть зависимыми. Эта зависимость может быть временной, пространственной или носить любой другой характер.

Корреляция бывает:

а) прямая (+r) – когда с увеличением одного параметра другой тоже увеличивается;

б) обратная (-r) – когда с увеличением одного параметра другой уменьшается;

в) прямолинейная – при равных изменениях одного признака другой изменяется на равные доли;

г) криволинейная – при равных изменениях одного признака другой изменяется на неравные доли. Единицы измерения признаков могут быть разными.

Полная положительная корреляция – r = +1.

Полная отрицательная корреляция – r = -1.

В природе полной корреляции практически не наблюдается.

Обозначения корреляции при значениях коэффициента корреляции:

r = 0.7 – 0.9 – сильная корреляция;

r = 0.5 – 0.69 – средняя корреляция;

r = 0.3 – 0.49 – умеренная корреляция;

r ‹ 0.3 – слабая корреляция. В биологических исследованиях не учитывается корреляция с коэффициентом ‹ 0.2.

Пример расчета коэффициента корреляции. Требуется установить взаимосвязь между количеством гемоглобина и эритроцитов в определенном количестве анализов крови. Какова теснота связи и достоверность проведенных исследований? В таблице 2 приведены данные по количеству гемоглобина (%) – x; количеству эритроцитов – y.

Следующий шаг в решении этой задачи – построение первичной матрицы, для чего исходные данные классируются по каждому из признаков, а внутри матрицы разносятся методом конверта (табл.3).

$\Sigma\, p_{xy}\, a_x\, a_y$ = (1x5x-5)+(1x-4x-5)+(1x-3x-3)+(1x-2x-3)+(3x-1x-1)+ 4x1x1)+(4x2x1)+ (2x1x2)+(2x2x2) = 87.

Таблица 2.

Количество гемоглобина (%) и количество эритроцитов в репрезентативной выборке пациентов.

X	y	x	y	X	y	x	У
0.8	22	3.30	82	3.14	83	3.82	87
1.71	45	4.10	81	3.21	73	4.36	94
2.63	61	3.29	82	3.28	82	2.60	50
3.46	77	3.19	66	3.63	78	1.30	27
3.32	80	2.80	72	3.28	79	2.80	63
3.11	82	3.10	71	3.66	84	2.87	70
3.71	97	3.81	87	3.90	75	4.20	87
4.22	96	4.47	90	4.33	82	3.68	72
3.50	92	3.59	76	3.80	79	3.40	71

$$\sigma = \sqrt{\frac{\Sigma pa^2}{n} - \left(\frac{\Sigma pa}{n}\right)^2}; \qquad \sigma_x = 1.6; \quad \sigma_y = 1.68$$

$$r = \frac{\Sigma\, pa_x a_y - \dfrac{\Sigma p_x a_x \quad \Sigma\, p_y a_y}{n}}{n\, \sigma_x\, \sigma_y}; \; r = 0.92.$$

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

В том случае, если взаимосвязь между факторами выражается не линейной, а другой зависимостью, то рассчитывается не коэффициент корреляции, корреляционное отношение.

$$\eta_{x/y} = \sqrt{\frac{\sigma_{x/y}^2}{\sigma_x}}; \qquad \eta_{y/x} = \sqrt{\frac{\sigma_{y/x}^2}{\sigma_y}};$$

Критерий достоверности определяется в основном по критерию Стьюдента:

$$1.\ t_r = \frac{m_r}{r}; \quad \text{при } n \rangle 100;$$

$$2.\ t_r = \frac{r}{1 - r^2}\sqrt{n - 2}; \text{ при } n \langle 100.$$

Таблица 3.

Первичная матрица с попарным распределением частот встречаемости признаков

		0.6-1.0	1.1-1.5	1.6-2.0	2.1-2.5	2.6-3.0	**3.1-3.5**	3.6-4.0	4.1-4.5	p_y	$p_y a_y$	$p_y a_y^2$	$p_{xy} a_x$
	А	-5	-4	-3	-2	-1	**0**	+1	+2				
21-30	-5	1	1							2	-10	50	-9
31-40	-4									0	0	0	0
41-50	-3			1	1					2	-6	18	-5
51-60	-2									0	0	0	0
61-70	-1					3	**1**			4	-4	4	-3
71-80	**0**					**1**	**4**	**5**		**10**	**0**	**0**	**4**
81-90	1						**6**	4	4	14	14	14	12
91-100	2							2	2	4	8	16	6
p_x		1	1	1	1	4	**11**	11	6	36	+2	102	+9
$p_x a_x$		-5	-4	-3	-2	-4	**0**	11	12	+5			
$p_x a_x^2$		25	16	9	4	4	**0**	11	24	93			
$p_{xy} a_y$		-5	-4	-3	-2	-3	**5**	8	8	+2			

Множественная корреляция показывает совокупное влияние группы признаков на изучаемый признак.

$$\mathbf{r_{A.BC}} = \sqrt{\frac{\mathbf{r_{AB}^2 + r_{AC}^2 - r_{AB}\, r_{AC}\, r_{BC}}}{\mathbf{1 - r_{BC}^2}}};$$

Если $\mathbf{r_{A.BC} < 0.3}$**,** то один из факторов нужно убрать.

Задание 1. Какова взаимосвязь между влажностью (%) и численностью злаковых мух, обитающих на пойменных лугах р.Усманка? Достоверны ли Ваши выводы?

Н (%): 72 74 74 76 76 76 78 78 78 80 82 82

n(экз.): 12 13 18 11 19 17 21 18 16 12 15 10

Задание 2. Какова взаимосвязь между температурой (^oC) и относительной влажностью (%) в условиях мезофильных биотопов? Достоверность выводов?

T (^oC): 16.2 16.8 16.9 17.2 17.6 18.2 18.4 18.6 19.0 19.4 19.6 19.8

H (%): 70 72 70 72 74 76 74 72 78 76 78 74.

Тема 5. Основы регрессионного анализа

Занятие 6. **Метод регрессионного анализа и условия его применения в зоологии**

Коэффициент корреляции показывает степень связи и ее направление, но не может показать закономерность, по которой одна величина изменяется относительно другой.

Регрессионный анализ дает возможность показать, насколько изменяется одна величина при изменении другой на единицу измерения, т.е. устанавливает закономерность сопряженного изменения двух признаков.
Главный показатель регрессионного анализа – коэффициент регрессии - $\mathbf{R_{x/y}}$:

$$\mathbf{R_{x/y} = r\frac{\sigma_x}{\sigma_y}}; \qquad \mathbf{R_{y/x} = r\frac{\sigma_y}{\sigma_x}}.$$

При графическом изображении регрессионной кривой $\mathbf{R_{x/y}}$ соответствует tg угла наклона регрессионной кривой.

Уравнение регрессионной кривой для линейной зависимости:
$\mathbf{y - \check{y} = b(x - x)}$, где $\mathbf{\check{y}}$ и $\mathbf{x}$ – средние по ряду.

Уравнение регрессионной кривой для нелинейной зависимости:

$\mathbf{y - \check{y} = f\ (x - x)}$, где функциональная зависимость может быть любой – log, e^x, cos, df и др.

Регрессия в биологии используется:

1) при установлении зависимости между двумя переменными факторами;
2) для анализа материала, включающего эксперимент и его экстраполяцию на природные условия;
3) для анализа качественных факторов, которые выражаются в долях:

$$\mathbf{M = \frac{P}{n}; \quad M = \frac{P \times 100\%}{n}; \quad \sigma = \sqrt{P\,(1 - P)}; \quad m_p = \sqrt{\frac{P\,(1 - P)}{n}}};$$

P – доля признака; **n –** общая выборка.

Задание 1. Какова взаимосвязь между изменениями температуры и численностью двукрылых мезофильных биотопов в пределах выборки и как она будет меняться, если t^oC упадет до 10^oC?

T(^{o}C):	16.2	16.8	17.2	17.4	17.8	18.0	18.2	18.4	18.6	18.8	19.0	19.2
n (экз.):	10	12	12	16	24	27	25	23	21	24	18	17

Тема 6. Дисперсионный анализ

Занятие 7**. Основы однофакторного дисперсионного анализа**

Дисперсионный анализ:

- позволяет установить закономерности пространственного и временного распределения факторов во всей их динамике и разнообразии;
- не зависит от величины выборки;
- признаки в дисперсионном анализе могут быть количественные и качественные, могут быть вероятностные (случайная дисперсия).

Дисперсионный анализ позволяет выявить статистическое влияние одного или многих факторов на результативный признак путем их относительной роли в общей изменчивости изучаемого признака.

В основе дисперсионного анализа лежит представление об общей вариации признаков, которую можно разложить на составные части, отражающие влияние конкретных факторов учтенных и неисследованных, т.е. провести сравнение вариабельности изучаемого фактора по отношению к случайному.

Дисперсия может быть определена как сумма квадратов отклонений отдельных вариант от средней арифметической всего комплекса.

$\mathbf{M_\Sigma}$ – средняя арифметическая по всему вариационному ряду;

$\mathbf{M_x}$ – средняя арифметическая частная (по фактору);

$\mathbf{M_z}$ – средняя арифметическая случайных факторов.

$$\mathbf{M_\Sigma = M_{x\,i\,\dots\,n} + M_z;}$$

$\mathbf{C_y = \Sigma(V- M_\Sigma)^2 = \Sigma\, D_y^{\,2} = \sigma_y^{\,2};}$ где: $\mathbf{C_y}$ - общая дисперсия;

$\mathbf{C_x = \Sigma\, n_x(M_x - M_\Sigma)^2 = \Sigma n_x\, D_x^{\,2} = \sigma_x^{\,2};}$ где: $\mathbf{C_x}$ – частная (факториальная) дисперсия;

$\mathbf{C_z = \Sigma(V- M_x)^2 = \Sigma\, D_z^{\,2} = \sigma_z^{\,2};}$ где: $\mathbf{C_z}$ – случайная дисперсия.

$$\mathbf{C_y = C_x + C_z.}$$

df – число степеней свободы - определяется числом условий вариационного ряда. Для факториальной дисперсии **df** равно числу классов: $\mathbf{df_1 = r - 1;\ df_2 = N - r.}$

$$\sigma_x^2 = \frac{C_x}{r - 1}; \qquad \sigma_z^2 = \frac{C_z}{N - r};$$

где **r** - число классов; **N** - весь комплекс.

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

Достоверность в этом случае определяется по критерию Фишера (F):

$$F = \frac{\sigma_x^2}{\sigma_z^2};$$ Если F › Fst , то полученные данные достоверны.

Занятие 8. **Дисперсионный однофакторный анализ для количественных факторов**

При однофакторном дисперсионном учитывается действие на изучаемый комплекс одного признака (фактора А).

Пример. Определить влияние растворенного вещества на плодовитость дафний. Данные представлены в таблице 4.

Таблица 4.

Выживаемость дафний (количество экземпляров) в среде с разным количеством растворенных веществ

	A_1 Контроль	A_2 Слабая	A_3 Средняя	A_4 Сильная	**r** = 4 – количество градаций А
V_1	6	8	8	9	$\Sigma V_x = 97$
V_2	5	7	8	7	$H_x = \frac{(\Sigma V_x^*)^2}{n_x}$
V_3	5	6	7	8	
V_4	7	6	-	-	
n_x (частоты)	4	4	3	3	$N = \Sigma n_x = 14$
ΣV_x^*	23	27	23	24	$\Sigma V_x = 97$
H_x	132.2	182.2	176.3	192.0	$\Sigma H_x = 682.7$
ΣV_x^2	135	185	177	194	$\Sigma V^2 = 691.0$
M_x	5.75	6.75	7.67	8.00	$M_\Sigma = 6.93$

$$H_\Sigma = \frac{(\Sigma V_x)^2}{N} = \frac{97^2}{14} = 672.1; \quad M_x = \frac{\Sigma V_x^*}{n_x}; \quad M_\Sigma = \frac{\Sigma V_x}{N};$$

$$C_x = \Sigma H_x - H_\Sigma = 10.6; \quad C_z = \Sigma V^2 - \Sigma H_x = 8.3; \quad C_y = \Sigma V^2 - H_\Sigma = 18.9;$$

$$\eta_x^2 = \frac{C_x}{C_y} = \frac{10.6}{18.9} = 0.56.$$

Таким образом, концентрация вещества на 56% определяет плодовитость дафний.

Достоверность проведенных исследований:

$$F = \frac{\sigma_x^2}{\sigma_z^2} = \frac{3.53}{0.83} = 4.1; \ \mathbf{df_1} = 3; \ \mathbf{df_2} = 10; \ F › Fst = 3.7 \text{ при } P = 0.95.$$

Задание 1. Как влияет на численность разных групп фитофагов удаленность от источника загрязнения?

Фитофаги		0.5 км	2 км	5 км	10 км
Карпофаги	Самки	4	5	12	17
	Самцы		3	7	12
Антофаги	Самки	4	7	15	16
	Самцы		3	10	11

Занятие 9. **Двухфакторный дисперсионный анализ количественных факторов**

Двухфакторный дисперсионный анализ представляет собой один из методов обработки биологических данных, когда требуется определить силу влияния фактора на классированный признак с учетом его изменчивости.

Пример. Зависит ли вес рогов (пант) от возраста дальневосточных оленей?

Данные приведены в таблице 5.

$$H_x = \frac{(pa)^2}{n_a}; \quad H_\Sigma = \frac{(\Sigma V_x)^2}{N}; \quad C_x = \Sigma H_x - H_\Sigma; \quad C_z = \Sigma pa^2 - \Sigma H_x; \quad C_y = \Sigma pa^2 - H_\Sigma$$

Таблица 5.

Вес рогов (x) и возраст дальневосточных оленей в одной из выборок.

x y	a	2	3	4	5	6	7	8	9	n_x
90 – 499	0	24								24
500 – 899	1	5	9	6	5	2	1			28
900 - 1299	2		2	1	5	4	2		1	15
1300 1699	3							1	1	2
1700 ›	4								1	1
$V_x = n_a$		29	11	7	10	6	3	1	2	ΣV_x=70
p		2	2	2	2	2	2	1	3	**N**= 70
pa		5	13	8	15	10	5	3	9	**Σpa**=68
pa²		5	17	10	25	18	9	9	29	Σpa^2=122
H_x		0,9	15.4	9.1	12.5	11.7	8.3	9.0	27.0	ΣH_x=108.9

$$\eta_x^2 = \frac{C_x}{C_y} = \frac{42.9}{56.0} = 0.76. \qquad F = \frac{\sigma_x^2}{\sigma_z^2} = \frac{6.1}{0.26} = 6.5 \text{ при } F› Fst = 2.9.$$

Задание 1. Как влияет на общую численность и численность самок мух-журчалок удаленность от источника загрязнений?

n	0	1	5	10	20	самки	0	1	5	10	20
V1	21	12	10	21	24	V1	11	7	6	18	19
V2	5	10	10	12	27	V2	5	8	8	11	21
V3	3	15	16	17	21	V3	3	12	12	13	19
V4	7	12	15	23	27	V4	5	9	11	17	23

Тема 7. Индексы общности и сравнение фаунистических данных

Занятие 10. **Индексы сравнения фаунистических комплексов**

Индексы сравнения фаунистических комплексов бывают разными и основываются или на сходстве фаунистических групп, или на их различии.

Наиболее разработаны индексы сходства. Индексы различия употребляются преимущественно тогда, когда сравниваются биотопы с разнородными природными условиями. Например, степные биоценозы Центрального Черноземья и полупустынные территории Казахстана.

Математический аппарат, лежащий в основе составления индексов сходства, основывается в основном на теории множеств. Под множеством понимается совокупность всех объектов, обладающих одним признаком или свойством. Составная часть множества – элемент. В этом случае при обозначении множества как - А, а элемента как – а, можем иметь два варианта: $a \in A$ (элемент принадлежит множеству) и $a \notin A$ (элемент не принадлежит множеству).

Возможен случай уничтожения одного множества другим $A_0 \varnothing A_1$. Примеров в биологии этому немало: фауна соснового леса, оказавшегося в пригородах города, уничтожена и занята синантропными видами.

Отдельные множества могут перекрываться и способы перекрывания могут быть разными.

1. A_1 полностью включено в A_2 , где A_1 и A_2 – два множества, например, два степных энтомокомплекса, причем один из них существует на заповедной, т.е. нетрансформированной территории, а другой – на соседней, подверженной антропогенной нагрузке.
2. A_1 и A_2 не перекрываются. Часто сравниваемые фаунистические комплексы находятся на территориях, далеко отстоящих друг от друга, и это расстояние не позволяет комплексам иметь точки соприкосновения.
3. A_1 и A_2 частично перекрываются. Этот вариант в биологии довольно часто встречается. Например, сравниваются два фаунистических комплекса, обитающих на сопредельных территориях, но в различающихся условиях: энтомокомплекс опушек и сопредельного с лесом пойменного луга.

Для того чтобы определить, какой из предлагаемых индексов фаунистического сходства наиболее корректен в каждом конкретном случае, необходимо иметь представление о S-M матрице или таксономической таблице 2x2 (табл.6).

Таблица 6.

Таксономическая таблица k-j-списка или 2x2

	k-список		
j-список	a	b	a + b
	c	d	c + d
	a + c	b + d	a + b + c + d = S

Обозначения, принятые в таблице, следующие:
a_i………a_m- j-список;
b_i……….b_n - k-список;
c – количество общих для двух множеств видов;
d – количество разных видов;
M = m + n.

Приводим формулы некоторых индексов фаунистического сходства, наиболее часто употребляющихся в зоологических исследованиях. Условные обозначения приняты во всех индексах одни и те же: **a** – общие для двух фаунистических комплексов виды; **b** и **c** - списки видов, обитающих в каждом биотопе.

Индекс Брауна-Бланквета: $\mathbf{I_{BB} = \dfrac{a}{a + b}}$; где **b ›› c;**

Индекс (**$\mathbf{I_{BB}}$**) применяется в случае неравноценных списков, например, при сравнении фаунистических списков двух биотопов, когда один их них исследовался в течение продолжительного времени, а другой – фрагментарно.

Индекс Симпсона-Сенкевича: $\mathbf{I_{SS} = \dfrac{a}{a + c}}$; где **b ›› c;**

Индекс (I_{SS}) применяется чаще всего для сравнения фаунистических списков по малой составляющей, например, при сравнении списков по присутствию в них редких видов.

Индекс Чекановского- Съеренсена (для качественных параметров):

$$\mathbf{I_{CS} = \frac{2a}{(a + b)+(a + c)}};$$ где: а- общие для двух фаун виды;
b и c – списки видов встречающихся только в одном из биотопов.

Индекс (I_{CS}) применяется для сравнения фаунистических комплексов, обитающих в биотопах, границы которых соприкасаются или возможно взаимное проникновение фаун.

Разновидностью **$\mathbf{I_{CS}}$** является:

$$\mathbf{I_{CS} = \frac{2\,a}{(b + c)}};$$ (обозначения те же, но списки a и b – полные.

Индекс Кульчинского:

$$\mathbf{I_K = \frac{a}{2}\left(\frac{1}{a + b}\quad\frac{1}{a + c}\right)};$$

Индекс (I_K) применяется для сравнения двух неравноценных списков или неравномерного распределения. Неравноценность списков в этом случае подразумевает не методические погрешности, а объективные условия, приведшие фаунистические комплексы к неравноценному состоянию.

Индекс Жаккара:

$$\mathbf{I_J = \frac{a}{a + b + c};}$$

Индекс Жаккара (**I_J**) применяется при сравнении фаунистических комплексов, обитающих на территориях, не сообщающихся друг с другом, или для сравнения комплексов малоподвижных форм, сообщение между популяциями которых невозможно или затруднено.

Задание 1. Каково сходство диптерокомплексов двух островных лесных массивов – Шипов лес и Воронежская нагорная дубрава? Насколько высока общность происхождения этих диптерокомплексов?

А, В, С – количество видов, обитающих в биотопах Шипова леса; Д, Е, Ж – количество видов, обитающих в биотопах Воронежской нагорной дубравы; АВ, ВС, ДЖ и т.д. – количество общих для двух биотопов видов.

А – 27, В – 25, С – 25, Д – 21, Е – 14, Ж – 16; АВ – 23, АС- 21, АД – 12, АЕ – 7, АЖ – 2, ВС – 22, ВД – 16, ВЕ – 7, ВГ – 9, СД – 16, СЕ – 8, СЖ – 7, ДЕ – 12, ДЖ – 14, ЕЖ – 12.

Задание 2. Каково сходство фаун короткоусых двукрылых, обитающих на пойменных лугах двух берегов реки Усманки?

А, Б, В – количество видов, обитающих на правом берегу; Г, Д, Е – количество видов, обитающих на левом берегу; АБ, ВГ, ДЕ и т.д. - количество общих для двух биотопов видов.

А – 21, Б – 20, В – 24, Г – 12, Д – 14, Е – 9; АБ – 18, АВ – 17, АГ – 7, АД – 7, АЕ – 2, БВ – 19, БГ – 8, БД – 9, БЕ – 3, ВГ – 8, ВД – 8, ВЕ – 3, ГД – 10, ГЕ – 6, ДЕ – 6.

Занятие 11. **Индексы сравнения фаунистических комплексов с учетом количественных данных**

$$\mathbf{I_{CS} = \sum \min (piA, piB)} \; –$$

- индекс Чекановского-Съеренсена (для сравнения структур коллекций). Индекс представляет собой сумму минимальных долей численностей видов из обоих списков.

$$I_{CS} = \frac{2 \sum \min (niA, niB)}{\sum niA + \sum niB}$$

- индекс Чекановского-Съеренсена (для сравнения коллекций с учетом различий в их объеме, т.е. когда последний отражает плотность животных).

- индекс Шеннона ($\mathbf{I_{SH}}$):

$$I_{SH} = - \sum_{i=1}^{s} P_i \ln P_i; \quad \text{где } P_i = \frac{n_i}{N};$$ n – число особей вида 1;

N – общее число особей; s – число видов.

Тема 8. Основы кластерного анализа

Занятие 12. Использование политетического метода объединительного иерархического неперекрывающегося кластерного анализа при обработке фаунистических данных

Кластерный анализ позволяет систематизировать группы факторов в зависимости от цели и установить тесноту связи между ними.

Кластер – группа признаков или факторов, объединенная по избранной шкале признаков.

Кластерный анализ в биологии чаще всего проводится на основе индексов фаунистического сходства. Для этого рассчитываются попарные коэффициенты сходства и составляется диагональная матрица. В случае большого количества обрабатываемых комплексов данные матрицы ранжируются по методу четных квадратов.

Рассмотрим кластерный анализ на примере. Известно количество видов злаковых мух в разных биотопах: А – 18 видов; Б – 40 видов; В – 21 вид; Г – 14 видов; Д – 17 видов. Биотопы несоприкасающиеся. Количество видов, общих для двух биотопов, известно. На основе индекса Жаккара рассчитываем коэффициент фаунистического сходства и строим диагональную матрицу (табл.7).

Таблица 7.

Данные индексов фаунистического сходства (индекс Жаккара).

	А	Б	В	Г	Д
А	100				
Б	40	100			
В	31	30	100		
Г	25	21	19	100	
Д	14	15	10	40	100

Существует несколько методов кластерного анализа. В биологии наиболее часто применяется политетический метод объединительного иерархического неперекрывающегося кластерного анализа. При построении кладограммы используется метод «ближайшего соседа».

Кластерный анализ в зоологии часто употребляется для выявления общности между разными фаунистическими группами, установления общности происхождения изучаемых групп беспозвоночных, фауногенеза и зоогеографии.

ОСНОВНАЯ ЛИТЕРАТУРА

1. Математические методы современной биомедицины и экологии / В.И. Афромеев, А.А Протопопов. В.П.Фильчакова, А.А.Яшин - Тула: ТулГУ, 1997. – 223 с.

2. Федоров М.П. Математические основы экологии / М.П. Федоров. – СПб: Изд-во СПбГТУ. – 1999. – 155 с.

3. Недорезов Л.В. Лекции по математической экологии / Л.В. Недорезов. – Новосибирск: Сибирский хронограф. – 1997. – 157 с.

ДОПОЛНИТЕЛНАЯ ЛИТЕРАТУРА

1. Лакин Г.Ф. Биометрия / Г.Ф. Лакин. – М.: Высшая школа, 1974. – 284 с.

2. Урбах В.Ю. Биометрические методы / В.Ю. Урбах. – М.: Наука, 1964. – 415 с.

3. Бейли Н. Математика в биологии и медицине / Н. Бейли. – М.: Мир, 1970. – 326 с.

4. Акчурин И.А. О методологических проблемах математического моделирования в биологии / И.А. Акчурин, М.Ф Веденов., Ю.В. Сачков. - Математическое моделирование жизненных процессов. – М., 1968. – С.7-29.

5. Баженов Л.Б. Некоторые философские вопросы моделирования биологических объектов / Л.Б. Баженов, Б.В. Бирюков. - Математическое моделирование жизненных процессов. – М., 1968. – С. 45-65.

6. Ляпунов А.А. О математическом подходе к изучению жизненных явлений // Математическое моделирование жизненных процессов / А.А. Ляпунов. – М., 1968. – С.65-108.

7. Фомин С.В. Математические проблемы в биологии / С.В Фомин., М.Б. Беркинблит. – М.: Наука, 1973. – 197 с.

8. Ивантер Э.В. Основы практической биометрии: Введение в статистический анализ биологических явлений / Э.В. Ивантер. – Петрозаводск, 1979. – 94 с.

9. Смит Д. Математические идеи в биологии / Д. Смит. – М.: Мир, 1970. – 179 с.

Составитель Пантелеева Наталья Юрьевна
Редактор Тихомирова О.А.