УДК 556.5.114 (075.8)

МЕТОДЫ ОПРЕДЕЛЕНИЯ СТЕПЕНИ ВЛИЯНИЯ И ВЗАИМОВЛИНИЯ УСЛОВИЙ НА ПРОЦЕССЫ И ЯВЛЕНИЯ В РЕЧНОЙ ЭКОСИСТЕМЕ

Доктор техн. наук М.Ж. Бурлибаев

В настоящее время при определении устойчивости речных экосистем вне поля зрения ученых остаются вопросы определения влияния и взаимовлияния процессов, происходящих в речной экосистеме как при естественном, так и в нарушенном гидрологических режимах водотоков. В рассматриваемой статье сделана попытка разработки методов для решения данной проблемы на основе дисперсионного анализа, как основного метода аппарата математической статистики.

В природопользовании, при постановке прикладных комплексных исследований, экспериментальные данные часто требуется разбить на группы, отличающиеся между собой в количественном отношении, и установить сходство (различие) между ними. Например, определить степень влияния географических условий на ход тех или иных природных процессов и явлений. Лучше всего этим требованиям отвечает дисперсионный анализ, который широко используется при решении именно практических задач. Дисперсионный анализ позволяет установить с определенной долей уверенности влияние на изучаемый объект каждого из исследуемых факторов, в отдельности, или в определенных их сочетаниях. Необходимым условием использования дисперсионного анализа является разбивка каждого учитываемого фактора не менее чем на две группы [1, 2, 3]. Если влияние факторов нельзя выразить количественными показателями, то они могут быть представлены качественными, выраженными в виде баллов. Дисперсионный анализ базируется на законе распределения отношения средних квадратов (дисперсий):

$$S_1^2 / S_2^2 = F , \qquad (1)$$

где S_1^2 – среднеквадратическая ошибка средних отдельного опыта; S_2^2 – суммарная среднеквадратическая ошибка средних всех опытов. Использование методов дисперсионного анализа позволяет дать ответ на следующие вопросы: 1) значимо ли влияет изучаемый фактор на воспроизводимости и в целом на результат? 2) если установлено значимое влияние какого-либо фактора на результат, в целом, то начиная с какого уровня фактора это влияние действует; значимо ли различаются выборочные средние между собой? 3) какой количественной мерой можно оценить степень установленного влияния? Сущностью дисперсионного анализа является расчленение общей суммы квадратов отклонений и общего числа степеней свободы на части – компоненты, соответствующие структуре эксперимента, и оценка значимости действия и взаимодействия изучаемых факторов по F-критерию. С этой целью, дисперсия разделяется на независимые слагаемые, которые, затем, сравниваются между собой. Допустим, в результате измерения величины *M* получено значение *X* и пусть на процесс измерения влияют случайные независимые факторы *A* и *B*. Тогда, отклонение – $M - X = \alpha + \beta + \gamma$, где α – отклонение под влиянием фактора *A*, β – то же под влиянием фактора *B*, а γ – под влиянием остальных, неучтенных факторов, причем, α, β и γ – независимы. В этом случае, дисперсия – $D(M - X) = D(\alpha + \beta + \gamma)$, а $DX = (D\alpha + D\beta + D\gamma)$, где $D\alpha$ – характеризует влияние фактора *A*, $D\beta$ – влияние фактора *B*, $D\gamma$ – влияние остальных, неучтенных факторов. Дисперсия $D\gamma$ называется *остаточной дисперсией.* Для оценки значимости факторов *A* и *B* сравниваются соответствующие дисперсии $D\alpha$ и $D\beta$ с остаточной дисперсией $D\gamma$. Если исследуется влияние одного фактора, то говорят об однофакторном дисперсионном анализе, при исследовании влияния двух факторов – о двухфакторном дисперсионном анализе и т.п. Рассмотрим простейшие способы применения дисперсионного анализа.

Однофакторный дисперсионный анализ

При однофакторном дисперсионном анализе обычно изучается действие одного фактора на *m* – уровнях *k* > 2, при равном числе определений (измерений) на каждом уровне *n*. Пусть фактор имеет *m* – уровней. Из каждого уровня делается выборка из *n* – элементов. Общее количество выбранных элементов обозначается как – $N = m \cdot n$. Вся выборка представляет собой матрицу:

$$X = \begin{Vmatrix} X_{11} & X_{12} & \dots & X_{1m} \\ X_{21} & X_{22} & \dots & X_{2m} \\ \dots & \dots & \dots & \dots \\ X_{n1} & X_{n2} & \dots & X_{nm} \end{Vmatrix}. \qquad (2)$$

Полагая, что данная выборка сделана из нормально распределенной генеральной совокупности, и задавая уровень значимости α, необходимо проверить гипотезу H_0 о равенстве средних значений на всех уровнях фактора – $\mu_1 = \mu_2 = ... = \mu_m$. При альтернативной гипотезе H_1 не все средние значения μ_1 должны быть равными. В качестве статистики используется величина F, определяемая по аналогичной с (1) зависимости:

$$F = S_1^2 / S_2^2, \tag{3}$$

в которой S_1^2 – дисперсия, характеризующая влияние исследуемого фактора (факторная дисперсия); S_2^2 – дисперсия, характеризующая влияние остальных факторов (остаточная дисперсия). Если гипотеза H_0 верна, то случайная величина F имеет F-распределение со степенями свободы $m – 1$ и $N – m = m·(n – 1)$. При проверке гипотезы H_0 используется правосторонняя критическая область, исходя из условия:

$$P(F > f_a) = \alpha . \tag{4}$$

Если значение статистики входит в критическую область, то гипотеза H_0 о равенстве средних значений на всех уровнях фактора отвергается, т. е. считается значимым влияние исследуемого фактора. В противном случае, принимается гипотеза H_0, т. е. считается, что значимость влияния фактора не установлена. При определении F находится сумма квадратов отклонений элементов выборки относительно общего среднего арифметического:

$$Q = \sum_{j=1}^{m} \cdot \sum_{i=1}^{n} (X_{ij} - \overline{X})^2 , \tag{5}$$

где $\overline{X} = \frac{1}{N} \cdot \sum_{j=1}^{m} \cdot \sum_{i=1}^{n} X_{ij}$, которая, в свою очередь, может быть разделена на два независимых слагаемых –

$$Q_1 = n \cdot \sum_{j=1}^{m} (\overline{X}_j - \overline{X})^2 , \tag{6}$$

$$Q_2 = \sum_{j=1}^{m} \cdot \sum_{i=1}^{n} (X_{ij} - \overline{X}_j)^2 \tag{7}$$

так, чтобы выполнялось равенство $Q = Q_1 + Q_2$; здесь $\overline{X}_j = \frac{1}{n} \cdot \sum_{j=1}^{n} X_{ij}$, при $j = 1, 2, ..., m$ – групповое среднее. Суммой квадратов межгрупповых отклонений, характеризующих влияние исследуемого фактора, является сумма квадратов отклонений групповых средних относительно общей средней (сумма Q_1). С другой стороны, Q_2 представляет собой сумму квадратов отклонений значений выборки относительно групповых средних, так называемую сумму квадратов внутри групповых отклонений. Эта сумма характеризует влияние остальных, неучтенных факторов. Имея суммы Q, Q_1, Q_2, можно вычислить соответствующие дисперсии:

$$S^2 = \frac{Q}{m \cdot n - 1}; \quad S_1^2 = \frac{Q_1}{m-1}; \quad S_2^2 = \frac{Q_2}{m \cdot (n-1)}. \tag{8}$$

Две последние дисперсии используются при вычислении F. При практических вычислениях, величины Q и Q_1 находятся, обычно, по выборке, а Q_2 – определяется как разность:

$$Q_2 = Q - Q_1. \tag{9}$$

В простейших случаях, на каждом уровне фактора, выбирается одинаковое количество объектов исследований, но нахождение сумм по (5) – (7) достаточно сложно, и более приемлемые формулы можно получить, преобразуя выражения (5) и (6), когда:

$$Q = \sum_{j=1}^{m}\left(\sum_{i=1}^{n} X_{ij}^2\right) - \frac{1}{m \cdot n} \cdot \left(\sum_{j=1}^{m}\left(\sum_{i=1}^{n} X_{ij}\right)\right)^2; \tag{10}$$

$$\boldsymbol{Q_1 = \frac{1}{n} \cdot \sum_{j=1}^{m}\left(\sum_{i=1}^{n} X_{ij}\right)^2 - \frac{1}{m \cdot n} \cdot \left(\sum_{j=1}^{m}\left(\sum_{i=1}^{n} X_{ij}\right)\right)^2}. \tag{11}$$

Тогда, обозначив $R_j = \sum_{i=1}^{n} X_{ij}^2$ и $L_j = \sum_{i=1}^{n} X_{ij}$, имеем:

$$Q = \sum_{j=1}^{m} R_j - \frac{1}{m \cdot n} \cdot \left(\sum_{j=1}^{m} L_j\right)^2; \tag{12}$$

$$Q_i = \frac{1}{n} \cdot \sum_{j=1}^{m} L_j^2 - \frac{1}{m \cdot n} \cdot \left(\sum_{j=1}^{m} L_j\right)^2. \tag{13}$$

По исходной матрице (выборке) вычисляются суммы элементов и их квадратов по столбцам L_j и R_j, при $j = 1, 2, ..., m$. Далее производится замена переменных $Y_{ij} = X_{ij} - C$. Целесообразно их выбрать близкими к общему среднему. В результате замены (для Y_{ij}) получаются следующие формулы:

$$Q = \sum_{j=1}^{n} P_j - \frac{1}{N} \cdot \left(\sum_{j=1}^{m} T_j\right)^2; \tag{14}$$

$$Q_1 = \frac{1}{n} \cdot \sum_{j=1}^{n} T_j^2 - \frac{1}{N} \cdot \left(\sum_{j=1}^{m} T_j \right)^2, \qquad (15)$$

где $N = m \cdot n$; $P_j = \sum_{i=1}^{n} Y_{ij}^2$ и $T_j = \sum_{i=1}^{n} Y_{ij}$, при $j = 1, 2, ..., m$.

На практике не всегда удается гарантировать одинаковое количество элементов на каждом уровне фактора. Если количество элементов на j-м уровне, обозначенное через n_j, при $j = 1, 2, ..., m$, то объем выборки составит:

$$N = \sum_{j=1}^{m} n_j . \qquad (16)$$

Формула (15) может быть записана в виде:

$$Q_1 = \sum_{j=1}^{m} \frac{1}{n_j} \cdot T_j^2 - \frac{1}{N} \cdot \left(\sum_{j=1}^{m} T_j \right)^2, \qquad (17)$$

а суммы P_j и T_j как – $P_j = \sum_{i=1}^{n_j} Y_{ij}^2$ и $T_j = \sum_{i=1}^{n_j} Y_{ij}$. Тогда дисперсии определяются по зависимостям:

$$\sigma^2 = \frac{Q}{N-1}; \quad \sigma_1^2 = \frac{Q_1}{m-1}; \quad \sigma_2^2 = \frac{Q_2}{N-m}. \qquad (18)$$

Остальные формулы не изменяются.

На примере материалов полевого опыта, в котором сравнивается урожайность озимой пшеницы при пяти вариантах технологий внесения удобрений и обработки почвы (табл. 1), проследим порядок выполнения дисперсионного анализа экспериментальных данных.

Таблица 1

Средняя урожайность озимой пшеницы по вариантам опыта, ц/га

Вариант	Урожайность по повторениям				Суммы по вариантам	Средняя урожайность
	I	II	III	IV		
A (контроль)	47,8	46,9	45,4	44,1	184,2	46,0
B	53,7	50,3	50,6	48,0	202,6	50,6
C	46,7	42,0	43,4	40,7	172,8	43,2
D	48,0	47,0	45,9	45,7	186,6	46,6
E	41,8	40,0	43,0	41,6	166,4	41,6
Суммы по повторениям	238,0	226,2	228,3	220,1	912,6 = $\Sigma X_i = \Sigma V_i = \Sigma P_i$	45,6 = $\overline{X}$

Алгоритм обработки опытных данных следующий:

1) в исходной таблице подсчитываются суммы урожаев – по вариантам (строки – V_i), по повторениям (столбцы – P_i) и определяются средние урожайности по вариантам (последний столбец);

2) проверяется правильность вычислений, для чего подсчитывается сумма сумм урожаев по вариантам и повторениям: ΣV_i = 184,2 + 202,6 + 172,8 + 186,6 + 166,4 = 912,6 (ц/га), $\sum P_i$ = 238,0 + 226,2 + 228,3 + 220,1 = 912,6 (ц/га), и проверяется равенство – $\sum X_i = \sum V_i = \sum P_i$ = 912,6 (ц/га);

3) вычисляются средние значения урожаев по вариантам путем деления сумм по вариантам V_i на число повторений (в рассматриваемом примере, n = 4);

4) определяется средняя урожайность озимой пшеницы, в целом по опыту, делением общей суммы урожаев $\sum X_i$ на общее число делянок Ni в опыте (20 дел.);

5) преобразуются исходные данные по соотношению $X'_{ij} = X_{ij} - C$, когда приняв за условное среднее число 45, близкое к среднему урожаю по опыту, можно облегчить вычисления сумм квадратов. Для варианта A (повторение I), при урожае 47,8 (ц/га) значение X'_{1I} = 47,8 – 45 = 2,8 и т. д. Преобразования значительно упрощают все последующие вычисления и не оказывают влияния на величину сумм квадратов отклонений. Преобразованные данные записываются в табл. 2, определяются суммы по повторениям (графам) и вариантам (строкам) и проверяется правильность расчетов по равенству – $\sum P'_{ij} = \sum V'_{ij} = \sum X'_{ij} = 12{,}6$ (ц/га);

Таблица 2

Преобразованные данные

Вариант	$X'_{ij} = X_{ij} - C = X_{ij} - 45$ (ц/га)				$\sum V'_{ij}$
	I	II	III	IV	V
А(контроль)	2,8	1,9	0,4	-0,9	4,2
B	8,7	5,3	5,6	3,0	22,6
C	1,7	-3,0	-1,6	-4,3	-7,2
D	3,0	2,0	0,9	0,7	6,6
E	-3,2	-5,0	-2,0	-3,4	-13,6
$\sum P'_{ij}$	13,0	1,2	3,3	-4,9	$\sum X'_{ij}$ =12,6

6) вычисляются суммы квадратов отклонений для различных компонентов варьирования в следующей последовательности –

а) общее число наблюдений – $N = m \cdot n = 5 \cdot 4 = 20$;

б) корректирующий фактор – S $=\sum(x'_{ij})^2$/N = 12,6^2/20 = 7,94;

в) суммы квадратов:

общая – $S_Y = \sum(X'_{ij})^2 - S$ = (2,8^2+1,9^2+...+3,4^2) - 7,94 = 246,67,

повторений – $S_P = \sum(P'_{ij})^2$/m – S = (13,0^2+1,2^2+2,3^2+4,9^2)/5 - 7,94 = 33,13,

вариантов – $S_V = \sum(V'_{ij})^2$/n – S = (4,2^2+22,6^2+...+13,6^2)/4 – 7,94 = 194,25,

ошибки: $S_z = S_y - S_p - S_V$ = 246,67 – 33,13 – 194,25 = 19,29;

7) заполняется таблица дисперсионного анализа (табл. 3).

Таблица 3

Результаты дисперсионного анализа

Варианты	Сумма квадратов отклонений	Степень свободы	Дисперсия	Критерий Фишера	
				$F_Ф$	$F^T_{0,05}$
Общая	246,67	19	–	–	–
Повторений	33,13	3	–	–	–
Вариантов	194,25	4	478,56	30,25	3,26
Остаток (ошибки)	19,29	12	1,6	–	–

Значение критерия $F^T_{0,05}$ находится по Приложению [1] для 4-х степеней свободы дисперсии вариантов (числитель) и для 12 степеней – дисперсии ошибки (знаменатель). В опыте есть существенные различия между вариантами и H_0: d = 0 отвергается $F_ф > F^T_{0,05}$;

8) для оценки существенности частных различий и группировки вариантов вычисляются – ошибка опыта, ошибка разности средних и наименьшая существенная разность ($НСР_{0,05}$) в абсолютных и относительных величинах:

$$S_{\bar{X}} = \sqrt{\frac{S^2}{n}} = \sqrt{\frac{1,60}{4}} \approx 0,64 \text{ (ц/га)}, \quad S_d = \sqrt{\frac{2 \cdot S^2}{n}} = \sqrt{\frac{2 \cdot 1,60}{4}} \approx 0,90 \text{ (ц/га)},$$

$$НСР_{0,05} = t_{0,05} \cdot S_d = 2,18 \cdot 0,90 = 1,96 \approx 2,0 \text{ (ц/га)},$$

$$НСР_{0,05} = \frac{t_{0,05} \cdot S_d}{\overline{X}} \cdot 100\% = \frac{2,18 \cdot 0,90}{45,6} \cdot 100\% = 4,3 \text{ (\%)}.$$

Для 12 степеней свободы ошибки, которые находятся как $(m-1)(n-1) = (5-1)(4-1) = 12$, по Приложению [1] отыскивается $F^T_{0,05} = 2,18$;

9) результаты эксперимента и статистической обработки записываются в итоговую таблицу; на основе $НСР_{0,05}$ распределяются варианты по группам и делаются выводы.

Таблица 4

Сравнение урожайности озимой пшеницы с контролем

Вариант	Средняя урожайность, (ц/га)	Разность с контролем		Группа	Заключение о существенности разности
		ц/га	%		
A(контроль)	46,0	–	–	–	–
B	50,6	4,6	10,0	I	Существенна
C	43,2	-2,8	-6,1	III	Существенна
D	46,6	0,6	1,3	II	Не существенна
E	41,6	-4,4	-9,6	II	Существенна
$НСР_{0,05}$	–	2,0	4,3		

При распределении вариантов по величине $НСР_{0,05}$ на три группы, руководствуются следующими положениями:

I группа – отклонения средних урожаев от контроля с положительным знаком больше $НСР_{0,05}$ – существенное повышение урожая;

II группа – отклонение не выходит за пределы ± $НСР_{0,05}$ – разность несущественная;

III группа – отклонения с отрицательным знаком больше по абсолютной величине $НСР_{0,05}$ – существенное снижение урожая.

Исходя из подобной группировки, вариант *B* (группа I) существенно превышает по урожаю контрольный вариант, а варианты *C* и *E* (III группа) существенно уступают ему. Вариант *D* (II группа) несущественно отличается от стандарта (контроля). Следовательно, на основе статистической обработки данных по урожайности озимой ржи, полученной в результате полевого опыта с пятью вариантами технологий внесения удобрений и обработки почвы, можно считать, что в варианте *B* в данных условиях получены более высокие урожаи, чем при традиционной технологии в контрольном варианте *A*.

А сейчас рассмотрим случай, когда используется критерий Фишера для установления различия влияния фактора на конечный результат анализа. Известно, что оптимальным условиям питания культурных растений соответствует среда достаточно увлажненной дерновой легкосуглинистой гумусированной нейтральной почвы. Ее можно создать путем внесения в пахотный горизонт добавок минерального грунта с определенным механическим составом. Формирование антропогенного почвенного слоя требует предварительных полевых экспериментов. В связи с этим, поставлена задача: определить влияние на урожай зерна ячменя различных доз гумуса (200, 300, 400, тонн абсолютно сухого вещества на гектар) при внесении его на фоне органоминеральных, органических удобрений и доломитовой муки. Исходная почва – дерново-подзолистая, глеевая, связносупесчаная, мелиорированная. Сведения об урожайности зерна ячменя в названных условиях приведены в табл. 5, куда занесена исходная информация по группам влияющего фактора (вариантам опыта) и некоторые результаты расчетов (для удобства сделано округление по урожайности до целых чисел).

Таблица 5

Однофакторный дисперсионный комплекс

Варианты опыта	Урожай ячменя по повторностям, ц/га*				$\frac{\sum X_i}{\sum X_i^2}$	$(\sum X_i)^2$	M_i
	I	II	III	IV			
1	2	3	4	5	6	7	8
Контроль (фон)	$\frac{20}{400}$	$\frac{21}{441}$	$\frac{22}{484}$	$\frac{20}{400}$	$\frac{83}{1725}$	6889	20,75
Фон + 200 т/га	$\frac{30}{900}$	$\frac{32}{1024}$	$\frac{32}{1032}$	$\frac{31}{961}$	$\frac{125}{3909}$	15625	31,25
Фон + 300 т/га	$\frac{35}{1225}$	$\frac{36}{1296}$	$\frac{35}{1225}$	$\frac{36}{1296}$	$\frac{142}{5032}$	20164	35,50
Фон + 400 т/га	$\frac{36}{1296}$	$\frac{35}{1225}$	$\frac{37}{1369}$	$\frac{37}{1369}$	$\frac{145}{5254}$	21025	36,25
$\sum X_k$	121	124	126	124	$\sum\sum X_{i,k} = 495$	$\sum(\sum X_i)^2 =$ $= 63703$	
$\sum X_k^2$	3816	3981	4102	4021	$\sum\sum X_{ik}^2 = 15920$		
$(\sum X_k)^2$	14641	15376	15876	15376	$\sum(\sum X_k)^2 = 61269$		
$\overline{X}_k$	30,25	31,00	31,50	31,00		$\overline{X}_{общ} = 30,93$	

*Примечание:** В числителе – опытные данные, в знаменателе – квадраты этих показателей.

Производятся расчеты по вариантам опыта (строкам), с разноской данных по столбцам: суммарный по повторностям урожай ячменя $\sum X_i$ по каждому варианту опыта вносится в столбец 6 (числитель); затем, в столбце 7, приводятся квадраты суммарного урожая ячменя по повторностям $\left(\sum X_i\right)^2$; среднее арифметическое $\overline{X}_i$ по общему варианту опыта, заносится в столбец 8; определяется общее среднее $\overline{X}_{общ}$.

После получения данных по вариантам опыта производятся расчеты необходимых показателей по повторностям X_k. Сначала суммируются данные по урожайностям ячменя и приводятся в строке $\sum X_k$ по повторностям. Суммы урожайности ячменя по вариантам опыта и повторностям должны совпасть и дать сумму всех вариант $\Sigma\Sigma X_{i,k}$ = 495. Аналогично суммируются квадраты этих показателей (знаменатель) по повторностям $\sum X_k^2$. Суммы сумм квадратов по вариантам и повторностям опыта должны совпадать и дать сумму квадратов всех вариант $\sum X_i^2 = \sum X_k^2$ = 15920. Ниже записываются результаты возведения в квадрат сумм вариант по каждой повторности $\left(\sum X_k\right)^2$ и их сумма – $\sum\left(\sum X_k\right)^2$ = 61269. Вычисляется среднее арифметическое по каждой повторности опыта $\overline{X}_k$. Общее среднее арифметическое всех вариант опыта составляет: $\overline{X}_{общ}$ = ($\Sigma X_{i,k}$)/N = 495/16 = 30,93.

Следующий этап расчетов – нахождение сумм квадратов отклонений, т. е. расчленение общего варьирования признака на составные части, исходя из равенства $Q = Q_1 + Q_2 + Q_3$, где Q – сумма квадратов отклонений по общему варьированию данных; Q_1 – то же по группам фактора (варианты опыта); Q_2 – то же по повторностям опыта; Q_3 – то же по остаточному варьированию. Общая сумма квадратов отклонений вычисляется следующим образом:

$$Q = \left(\sum\sum X_{i,k}^2 - \left(\sum\sum X_{i,k}\right)^2\right)/N = (15920 - (495^2)/16) = 620{,}94.$$

Затем находится сумма квадратов отклонений по группам фактора опыта по формуле: $Q_1 = \left(\sum\left(X_i\right)^2 - \left(\sum\sum X_{i,k}\right)^2/k\right)/i$ = 63703 – 495²/4/4 = 611,69, где k = 4 – число групп фактора; i = 4 – число слагаемых в сумме по вариантам опыта (равное количеству повторностей). В данном случае, должно соблюдаться равенство $N = k \cdot i = 4 \cdot 4 = 16$.

Сумма квадратов отклонений по повторностям опыта находится по формуле $Q_2 = \left(\sum\left(X_k\right)^2 - \left(\sum\sum X_{i,k}\right)^2/i\right)/k$ = 61269 – 495²/4/4 = 3,19.

Сумма квадратов отклонений по остаточному варьированию определяется из равенства $Q_3 = Q - Q_1 - Q_2$ = 620,94 – 611,68 – 3,18 = 6,08.

Результаты дисперсионного анализа данных урожая ячменя приведены в табл. 6. В таблицу вносятся рассчитанные суммы квадратов отклонений Q, Q_1, Q_2, Q_3. Число степеней свободы определяется следующим образом: по общей сумме квадратов отклонений $\nu = N - 1 = 16 - 1 = 15$; по вариантам опыта $\nu_1 = n_1 - 1 = 4 - 1 = 3$; по повторностям – $\nu_2 = n_2 - 1 = 4 - 1 = 3$; по остаточной сумме – $\nu_3 = \nu - \nu_1 - \nu_2 = 15 - 3 - 3 = 9$.

Таблица 6

Результаты однофакторного дисперсионного анализа

Варьирование данных	Сумма квадратов отклонений, Q	Степень свободы, ν	Дисперсия, σ^2	Критерий Фишера, F	
				$F_Ф$	F_T
Общее по опыту	620,94	15	41,39	–	–
По вариантам опыта	611,68	3	203,89	304,31	8,81
По повторностям	3,18	3	1,05	1,56	8,81
Случайное (остаточное)	6,08	9	0,67	–	–

Дисперсия определяется путем деления сумм квадратов отклонений Q, Q_1, Q_2, Q_3 на соответствующие им числа степеней свободы ν, ν_1, ν_2, ν_3, что можно выразить в общем виде формулой: $\sigma^2 = Q/\nu$. Фактический критерий Фишера $F_ф$ определяется путем деления каждой из величин дисперсий на значение остаточной дисперсии. Критическое значение критерия Фишера $F_ф$ находится по Приложению [1] на пересечении значений большей и меньшей степеней свободы, которые устанавливаются по величине сравниваемых дисперсий. Так как $F_Ф > F_T$, то внесение добавок минерального грунта положительно влияет на величину урожая ячменя в исследуемых условиях.

Двухфакторный дисперсионный анализ

Если в дисперсионный анализ включить несколько факторов, влияющих на результативный признак, то они должны быть независимыми друг от друга. При обработке данных исходной информации, алгоритм расчетов аналогичен однофакторному дисперсионному анализу.

Определим влияние метеорологических условий (фактор I) и водохозяйственных мероприятий (фактор II) на урожай биомассы трав в различных агроландшафтах.

Здесь имеет место обработка данных с двумя факторами, каждый из которых делится на две группы. Для этого составляется комбинационный (двухфакторный) дисперсионный комплекс (табл. 7). Каждый фактор характеризуется тремя наблюдениями (повторностями). Аналогичную схему можно использовать для двухфакторного анализа с большим числом групп и повторностей в каждом факторе.

Таблица 7

Двухфакторный дисперсионный комплекс

Повторяемость опыта по фактору (II)	Биомасса, кг/м^2		$\frac{\sum Y_i}{\sum Y_i^2}$	$\left(\sum Y_i\right)^2$	$\overline{Y}$
	Группы по фактору (I)				
	1992 год (сухой)	1994 год (влажный)			
1	*2*	*3*	*4*	*5*	*6*
Группа фактора (II) (агроландшафт)					
Первая	5/25	4/16	9/41		
Вторая	6/36	5/25	11/61		
Третья	5/25	6/36	11/61		
∑/∑	16/86	15/77	31/163	961	5,16
Группа фактора (II) (антропогенный агроландшафт)					
Первая	3/9	5/25	8/34		
Вторая	4/16	6/36	10/52		
Третья	4/16	6/36	10/52		
∑/∑	11/41	17/97	28/138	784	4,66
$\frac{\sum X_i}{\sum X_i^2}$	27/127	32/174	59/301	$\left(\sum\sum Y_i\right)^2 = 1745$	
$\left(\sum X_i\right)^2$	729	1024		$\left(\sum\sum X_i\right)^2 = 1753$	
$\overline{M}$	4,50	5,33		$\overline{M}_{общ} = 4{,}90$	

Примечание: X_i – варианты опыта по фактору (I), Y_i – то же по фактору (II). В числителе – урожай биомассы, в знаменателе – квадрат чисел.

Суть двухфакторного дисперсионного анализа можно представить равенством:

$$Q = Q_1 + Q_2 + Q_3 + Q_4 + Q_5, \qquad (19)$$

где Q – общая сумма квадратов; Q_1 и Q_2 – соответственно, сумма квадратов отклонений для факторов (I) и (II); Q_3 – сумма квадратов отклонений, имеющих место при взаимодействии факторов (I) и (II); Q_4 – сумма квадратов отклонений по повторностям; Q_5 – остаточная сумма квадратов отклонений неучтенных факторов. Общая сумма квадратов отклонений находится как:

$$Q = \sum\sum X^2 Y^2 - \left(\left(\sum\sum X_i Y_i\right)^2 \Big/ N\right) = (301 - (59^2/12)) = 10{,}92,$$

где $N = 12$ – общий объем выборки; сумма квадратов отклонений по фактору (I) как:

$$Q_1 = \sum\left(\sum X_i\right)^2 - \left(\sum\sum X_i Y_i\right)\Big/ n_x \Big/ k_x = (1753 - 59^2/2)/6 = 2{,}08,$$

где $n_x = 2$ – число групп фактора (I); $k_x = 6$ – число вариант в каждой отдельной сумме; сумма квадратов отклонений по фактору (II) вычисляется аналогично:

$$Q_2=(\sum(\sum Y_i)^2-(\sum\sum X_i,Yi/n_y))/k_y=(1745-59^2/2)/6=0,75;$$

сумма квадратов отклонений, вызываемых взаимодействием факторов (I) и (II), определяется следующим образом:

$$Q_3=(\sum(\sum Z_i^2)-(\sum\sum X_i,Y_i)^2/n_z)/k_z-Q_1-Q_2=(891-59^2/4)/3-2,08-0,75=4,08,$$

где $\sum(\sum Z_i^2)=(16^2+15^2+11^2+17^2)=891$ – сумма квадратов сумм значений вариант по группам выборки комбинационной таблицы; $n_z=4$ - число сумм вариант по группам; $k_z=3$ - число слагаемых вариант в каждой группе выборки; сумма квадратов отклонений по повторностям Q_4 определяется как - $Q_4=(\sum(\sum X_i)^2-(\sum\sum X_i,Y_i)/n_{x,y})/k_{x,y}=(1171-59^2/3)4=2,67$, где $n_{x,y} = 3$ – число сумм по повторностям; $k_{x,y} = 4$ – число слагаемых в каждой сумме; сумма квадратов сумм X_i, вычисленная как:

$$\sum(\sum X_i)^2 = ((5+4)+(3+5))^2+((6+5)+(4+6))^2+((5+6)+(4+6))^2=1171;$$

сумма квадратов отклонений по остаточному варьированию составляет:

$$Q_5 = Q - Q_1 - Q_2 - Q_3 - Q_4 = 10,92 - 2,08 - 0,75 - 4,08 - 2,67 = 1,14;$$

число степеней свободы для Q будет – $\nu = N - 1 = 11$; для Q_1 и Q_2 – соответственно, равно числу градаций фактора минус единица – $\nu_1 = n_1 - 1 = 2 - 1 = 1$, $\nu_2 = n_2 - 1 = 2 - 1 = 1$; для Q_3 определится как – $\nu_3 = \nu_1 \cdot \nu_2 = 1\cdot1 = 1$; для Q_4 – равно числу повторностей минус единица – $\nu_4 = 3 - 1=2$; для Q_5 этот показатель определяется следующим образом – $\nu_5 = \nu - \nu_2 - \nu_3 - \nu_4 = 11 - 1 - 1 - 1 - 2) = 6$.

Полученные расчетным путем характеристики сведены в табл. 8. Показатели дисперсии (табл. 8) вычисляются путем деления значений сумм квадратов отклонений на соответствующие значения степеней свободы. Фактический критерий Фишера определяется путем деления каждой из величин дисперсий на значение остаточной дисперсии. Критическое значение критерия Фишера найдено по Приложению[1].

Анализируя критерии Фишера, можно заключить, что влияние исследуемых параметров на биомассу существенно во всех рассмотренных случаях при вариации опытных данных $V = \sqrt{\sigma^2_{общ}} / \overline{M}_{общ} \cdot 100\% = \sqrt{0,99} / 4,9 \cdot 100\% = 20,0\%$ и $F_{тм} > F^T_{0,05}$; действие группы факторов (II) на исследованных агроландшафтах не доказано $F_{\text{тм}} < F^{T}_{0,05}$.

Таблица 8

Результаты двухфакторного дисперсионного анализа

Варьирование данных	Сумма квадратов отклонений, Q	Степень свободы, ν	Дисперсия, σ^2	Критерий Фишера	
				F_{ϕ}	$F^{T}_{0,05}$
Общие по опыту	10,92	11	0,99	5,21	4,03
По фактору (I)	2,08	1	2,08	10,94	5,99
По фактору (II)	0,75	1	0,75	3,94	5,99
По взаимодействию факторов (I) и (II)	4,08	1	4,08	21,47	5,99
По повторностям	2,67	2	1,34	7,05	5,14
Остаточное	1,14	6	0,19	1,00	–

Оценку результатов эксперимента можно сделать и по критериям наименьшей существенной разности и Стьюдента. Для вычисления *НСР* и *t* находится ошибка среднего арифметического m_M всего опыта и ошибка разности средних по формулам:

$$m_{м} = \sqrt{\sigma^2_{ост.} / N} = \sqrt{0{,}19/12} = 0{,}1258;$$

$$m_d = \sqrt{2 \cdot \sigma^2_{ост} / n} = \sqrt{2 \cdot 0{,}19/6} = 0{,}25;$$

$$\text{НСР} = m_d \cdot t^{T}_{0,05} = 0{,}25 \cdot 2{,}45 = 0{,}61,$$

в которых *n* – численность меньшей из сравниваемых частных групп. По критерию Стьюдента сравниваются средние арифметические данные по естественному и антропогенному агроландшафтам:

$$t = (M_{y,1} - M_{y,2}) / m_{d_y} = (5{,}16 - 4{,}66)/0{,}25 = 2{,}00.$$

По Приложению [1] критерий Стьюдента – $t^{T}_{0,05}$ = 2,45, при P = 0,95 для ν = 6. Таким образом, на биомассу трав в агроландшафтах не влияет мелиорация (т. е. фактор II), так как $t_{Ф}$ = 2,0 < $t^{T}_{0,05}$ = 2,45, при P = 0,95; метеорологические условия (фактор I) достоверно влияют на биомассу трав, при P = 0,95. Выводы, сделанные с использованием критериев Фишера и Стьюдента, совпадают. В заключении обычно определяется точность опыта, которая составляет:

$$P = \left(m_M / M_{общ}\right) \cdot 100\ \% = 0{,}1258/4{,}9 \cdot 100\ \% = 2{,}56\ \%.$$

Точность опыта признается достаточно высокой, поскольку $P < 3\ \%$. Коэффициент варьирования опытных данных: $V = \frac{\sqrt{\sigma_{об}^2}}{\overline{M}_{общ}} \cdot 100\,\% = \frac{\sqrt{0{,}99}}{4{,}9} \cdot 100\,\% = 20{,}0\,\%,$ также незначителен, что удовлетворяет требованиям опыта.

Рассмотрим двухфакторный дисперсионный анализ биологической проблемы, т. е. когда требуется установить, значимо ли различие в действии форм азотных удобрений на урожай овсяницы луговой (табл. 9). Нулевая гипотеза H_0: $d = 0$. Особенностью повариантной обработки данных вегетационного опыта с разной повторностью является необходимость вычисления нескольких значений наименьшей существенной разности, так как не все средние равноточны. В примере варианты (1–2) имеют четыре, а варианты (3–4) – шесть наблюдений. В установленном порядке, выполняются следующие вычислительные операции:

1) определяются суммы урожаев и средние по вариантам, общая сумма и средний урожай по опыту (табл. 9);

Таблица 9

Урожай овсяницы (*г* на сосуд)

Варианты (формы азота)	Урожай, X						Число наблюдений, n	Суммы V	Средние
1	16,0	17,2	14,4	15,8	–	–	4	63,4	15,85
2	29,4	30,4	30,3	28,1	–	–	4	118,2	29,55
3	26,0	29,2	26,7	27,1	26,0	28,1	6	163,1	27,18
4	25,3	24,8	26,1	23,2	25,7	24,0	6	149,1	24,85
Общая сумма							$20 = \Sigma n = N$	$493{,}8 = \Sigma X$	$24{,}69 = \bar{x}$

2) преобразуются исходные даты по соотношению $X_1 = X - A$, при условной средней A, принимаемой равной числу 25, близкому к среднему урожаю по опыту $\overline{X}$ = 24,69 (табл. 10), с последующим вычислением сумм квадратов отклонений.

При повариантном вычислении сумм квадратов отклонений необходимо иметь ввиду, что в суммы V входит разное число наблюдений n. Далее, последовательно определяются:

а) общее число наблюдений $N = \Sigma n = 20$;

б) корректирующий фактор $C = (\Sigma X_1)^2/N = (-6{,}2)^2/20 = 0{,}07$;

в) суммы квадратов отклонений

$$C_y = \Sigma X_1^2 - C = (9{,}0^2 + 7{,}8^2 + \ldots + 1{,}0^2 - 1{,}92 = 465{,}70;$$

$$C_v = \Sigma\left(\frac{V_1^2}{n_1} + \frac{V_2^2}{n_2} + \ldots + \frac{V_l}{n_l}\right) - C = \left(\frac{36{,}6^2}{4} + \frac{18{,}2^2}{4} + \frac{13{,}1^2}{6} + \frac{0{,}9^2}{6}\right) - 1{,}92 = 444{,}51;$$

$$C_z = C_y - C_v = 465{,}70 - 444{,}51 = 21{,}19.$$

Таблица 10

Преобразованные данные

Варианты	$X_1 = X - 25$						Суммы V
1	–9,0	–7,8	–10,6	–9,2	–	–	–36,6
2	4,4	5,4	5,3	3,1	–	–	18,2
3	1,0	4,2	1,7	2,1	1,0	3,1	13,1
4	0,3	–0,2	1,1	–1,8	0,7	–1,0	–0,9
Общая сумма							$-6{,}2 = \Sigma X_1$

Вычисленные суммы квадратов отклонений вносятся в табл. 11, наряду с другими составляющими дисперсионного анализа.

Таблица 11

Результаты дисперсионного анализа

Дисперсия	Сумма квадратов	Степени свободы	Средний квадрат	$F_ф$	F_{05}
Общая	465,70	19	–	–	–
Вариантов	444,51	3	148,80	112,7	3,24
Остаток (ошибки)	21,19	16	1,32	–	–

Значение F_{05} принимается по Приложению[1] для 3-х степеней свободы дисперсии вариантов (числитель) и 16-ти степеней свободы остатка (знаменатель). Так как $F_ф > F_{05}$, то между вариантами опыта имеются существенные различия на 5%-ном уровне значимости и гипотеза H_0 отвергается;

3) при оценке существенности частных различий в опыте с разной повторностью необходимо учесть неравноточность сравнения средних. Ошибки средних первых двух вариантов $\bar{x}_1$ и $\bar{x}_2$ предопределяются числом наблюдений $n_1 = n_2 = 4$, а двух последних $\bar{x}_3$ и $\bar{x}_4$, соответственно, $n_3 = n_4 = 6$ наблюдений. Поэтому, ошибку разности между средними нужно определять по формуле, учитывающей разную повторность по (1-2) и (3-4) вариантам: $s_d = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{s^2 \frac{n_1 + n_2}{n_1 n_2}}$.

Тогда:

а) ошибка разности средних, при сравнении $\bar{x}_1$ с $\bar{x}_2$ $n_1 = n_2 = 4$, будет

$$s'_d = \sqrt{2s^2/n} = \sqrt{2 \cdot 1{,}32/4} = 0{,}81$$

при сравнении $\bar{x}_1$ и $\bar{x}_2$ с $\bar{x}_3$ и $\bar{x}_4$ $n_1 = 4$ и $n_2 = 6$, будет

$$s''_d = \sqrt{s^2 \frac{n_1 + n_2}{n_1 n_2}} = \sqrt{1{,}32 \cdot \frac{4+6}{4 \cdot 6}} = 0{,}74$$

при сравнении $\bar{x}_3$ с $\bar{x}_4$ $n_3 = n_4 = 6$, будет

$$s'''_d = \sqrt{2s^2/n} = \sqrt{2 \cdot 1{,}32/6} = 0{,}66$$

б) наименьшая существенная разность для 5%-го (или 1%-го) уровня значимости определяется как

$$HCP'_{05} = t_{05} s'_d = 2{,}12 \cdot 0{,}81 = 1{,}72$$

$$HCP''_{05} = t_{05} s''_d = 2{,}12 \cdot 0{,}74 = 1{,}57$$

$$HCP'''_{05} = t_{05} s'''_d = 2{,}12 \cdot 0{,}66 = 1{,}40 \ .$$

Значения критерия $t_{05} = 2{,}12$ принимаются из Приложений [1] для 16-ти степеней свободы дисперсии ошибки (остатка). Результаты опыта и статистической обработки заносятся в табл. 12.

Таблица 12

Различия в действии форм азотных удобрений на урожай овсяницы луговой (г на сосуд)

Варианты	Урожай	Сравнение с контролем		Сравнение с аммиачной селитрой	
		разность	НСР$_{05}$	разность	НСР$_{05}$
Без удобрений (контроль)	15	–	–	–11,4	1,57
Сульфат аммония	29,6	13,8	1,72	2,4	1,57
Аммиачная селитра	27,2	11,4	1,57	–	–
Мочевина	24,8	9,0	1,57	–2,4	1,40

Результаты анализа показывают, что все формы азотных удобрений существенно повышают урожай овсяницы луговой. Аммиачная селитра и мочевина примерно равноценны по эффективности; сульфат аммония обеспечивает статистически значимый на 5%-ом уровне эффект в сравнении с аммиачной селитрой.

А сейчас рассмотрим особенности применения двухфакторного дисперсионного анализа генетической проблемы, т.е. когда по данным учета числа зерен в колосе у гибридов ячменя (табл. 13) необходимо вычислить коэффициент генетической наследуемости.

Таблица 13

Число зерен в колосе у гибридов ячменя

Материнская форма (A)	Отцовская форма (B)	Повторения (X)				Суммы V	Средние
		1	2	3	4		
a_1	b_1	20	20	22	21	83	20,8
	b_2	18	19	20	19	76	19,0
a_2	b_1	23	23	20	22	88	22,0
	b_2	22	23	24	24	93	23,2
a_3	b_1	29	28	27	28	112	28,0
	b_2	29	30	32	32	123	30,8
Суммы P		141	143	145	146	$575 = \Sigma X$	

1) проводится дисперсионный анализ для двухфакторного комплекса:

$$N = l_A \cdot l_B \cdot n = 3 \cdot 2 \cdot 4 = 24;\ C = (\Sigma X)^2 : N = 575^2 : 24 = 13776{,}04;$$

$$C_Y = \Sigma X^2 - C = (20^2 + 20^2 + ... + 32^2) - 13776{,}04 = 428{,}96;$$

$$C_P = \Sigma P^2 : l_A \cdot l_B - C = (141^2 + 143^2 + 145^2 + 146^2) : 3 \cdot 2 - 13776{,}04 = 2{,}46;$$

$$C_V = \Sigma V^2 : n - C = (83^2 + 76^2 + ... + 123^2) : 4 - 13776{,}04 = 406{,}71;$$

$$C_Z = C_Y - C_P - C_V = 428{,}96 - 2{,}46 - 406{,}71 = 19{,}79;$$

2) для выполнения оценки существенности действия материнских и отцовских форм и их взаимодействия на результативный признак гибридов составляется табл. 14, в которую вписываются суммы V для каждого гибрида и находятся суммы по факторам A и B.

Таблица 14

Вычисление сумм для определения эффектов A, B и взаимодействия AB

Фактор A	Фактор B		Суммы A
	b_1	b_2	
a_1	83	76	159
a_2	88	93	181
a_3	112	123	235
Суммы B	283	292	$575 = \Sigma X$

3) по результатам дисперсионного анализа (табл. 15) имеем: общее варьирование C_{A+B+AB} – внутренняя часть таблицы (численно $C_{A+B+AB} = C_v = 406{,}71$ и вычислено нами ранее), варьирование факторов A и B. Взаимодействие AB находится по разности: $C_A = \Sigma A^2 / l_B \cdot n - C = (159^2+181^2+235^2)/2 \cdot 4 - 13776{,}04 = 382{,}34$, при $l_A - 1 = (3 - 1) = 2$ – степенях свободы; $C_B = \Sigma B^2 / l_A\, n - C = (283^2+292^2) : 3 \cdot 4 - 13776{,}04 = 3{,}38$, при $l_B - 1 = 2 - 1 = 1$ – степени свободы; $C_{AB} = C_{A+B+AB} - C_A - C_B = 406{,}71 - 382{,}34 - 3{,}38 = 20{,}99$, при $(l_A - 1) \cdot (l_B - 1) = (3-1)\,(2-1) = 2$ – степенях свободы; существенность

действия и взаимодействия факторов оценивается по критерию F (Приложение [1]).

Таблица 15

Результаты дисперсионного анализа

Дисперсия	Сумма квадратов	Степени свободы	Средний квадрат	$F_ф$	F_{05}
Общая	428,96	23	-	-	-
Повторений	2,46	3	-	-	-
Материнских форм (A)	382,34	2	191,17	144,83	3,60
Отцовских форм (B)	3,38	1	3,38	2,56	4,54
Взаимодействия (AB)	20,99	2	10,50	7,95	3,60
Остаток	19,79	15	1,32	-	-

В рассматриваемом примере существенным оказалось действие А (материнских форм) и взаимодействие АВ (материнская форма×отцовская форма). Следовательно, имеет смысл вычислить коэффициенты наследуемости, характеризующие силу генетического влияния материнских форм и взаимодействия; вариабельность числа зерен в колосе у гибридов ячменя не зависит существенно от отцовских форм $F_ф < F_{05}$;

4) в двухфакторном комплексе дисперсия групповых средних имеет более сложную природу, чем в однофакторном, и определяется как генетической изменчивостью, обусловленной генотипами материнских и отцовских форм и их взаимодействием, так случайной изменчивостью (остаток). Общий коэффициент наследуемости в этом случае равен: $h^2 = h^2_A + h^2_B + h^2_{AB}$.

Для выяснения сущности вычислительных операций при определении дисперсий, характеризующих влияние на фенотипическую изменчивость генотипов материнских форм s^2_A, отцовских форм s^2_B и их взаимодействия s^2_{AB}, целесообразно дополнительно рассмотреть схему компонентного двухфакторного анализа результатов эксперимента. В настоящем примере, существенным оказалось влияние материнских форм A и взаимодействия AB. Однако, при подборе пар для скрещивания, необходимо иметь в виду, что проявление результативного признака в гибридах зависит в основном (на 86 %) от материнского растения.

СПИСОК ЛИТЕРАТУРЫ

1. Бурлибаев М.Ж., Волчек А.А., Шведовский П.В. Проблемы оптимизации природопользования и природообустройства в математических моделях и методах. – Алматы: Изд-во: «Каганат», 2003. – 532 с.

2. Бурлибаев М.Ж., Нурмаганбетов Д.Ш., Волчек А.А. Теоретические и прикладные основы проблем планирования и управления природопользованием и охраной природы. – Алматы, Изд-во: «Каганат», 2007. – 360 с.
3. Бурлибаев М.Ж. Теоретические основы устойчивости экосистем трансзональных рек Казахстана. – Алматы, Изд-во: «Каганат», 2007. – 516 с.

Казахстанское Агентство Прикладной Экологии (КАПЭ), г. Алматы

ӨЗЕН ЭКОСИСТЕМАСЫНДАҒЫ ПРОЦЕСТЕР МЕН ҚҰБЫЛЫСТАРҒА ӘСЕР ЕТУШІ ЖӘНЕ БІРІГІП ӘСЕР ЕТУШІ ЖАҒДАЙЛАРДЫҢ ДӘРЕЖЕСІН АНЫҚТАУ ӘДІСТЕРІ

Техн. ғылымд. докторы М.Ж. Бүрлібаев

Қазіргі уақытта суағындарының табиғи және өзгерген гидрологиялық режимдерде өзен экосистемасындағы болатын процестерге әсер етуші және бірігіп әсер етуші өзен экосистемасының орнықтылығын анықтау сұрақтары ғалымдардың ескеруінсіз қалып отыр. Қарастырылып отырған мақалада математикалық статистика аппаратының негізгі әдісі мен дисперсиялық талдау негізінде осы мәселені шешудің өңдеу әдістемелері әрекеті жасалған.