

Электронное научное издание

«Международный электронный журнал. Устойчивое развитие: наука и практика»

www.yrazvitie.ru

вып. 2 (17), 2016, ст. 5

Выпуск подготовлен по итогам VI Международной научной конференции по фундаментальным и прикладным проблемам устойчивого развития в системе «природа – общество – человек» (19-20 декабря 2016 г.)

УДК 004.94

ИСПОЛЬЗОВАНИЕ ПРОГРАММЫ MICROSOFT EXCEL В БИОМЕТРИИ

Баяхов Алиби Науханович, член-корреспондент РАЕН, Казахстанский университет инновационных и телекоммуникационных систем

Моисейкина Людмила Гучаевна, доктор биологических наук, профессор

Аннотация

В статье рассматриваются математические методы в биологических исследованиях при изучении групповых свойств биологических объектов. Дан краткий обзор основных операции с данными в Excel. Описан порядок их применения, проанализированы эффективность и ограниченность этих механизмов. На конкретном примере показаны вычисления параметров вариационного ряда, вычисление коэффициента корреляции.

КЛЮЧЕВЫЕ СЛОВА: жизнь, мера, биометрия, Microsoft Excel.

THE USE OF MICROSOFT EXCEL IN BIOMETRICS

Bayakhov Alibi Naukhanovich, corresponding member of the Academy of natural Sciences, Kazakhstan University of innovative and telecommunication systems

Moiseikina Lyudmila Gouchayevna, doctor of biological Sciences, Professor

Abstract

The article deals with mathematical methods in biological research in the study of group properties of biological objects. A brief overview of the main data operations in Excel. Described the manner of their application, effectiveness and limitations of these mechanisms. The given example shows that the computation of the parameters of variational series, calculation of correlation coefficient.

KEYWORDS: life, measure, biometrics, Microsoft Excel.

Биометрия (от греч. bios – жизнь, metron – мера) – наука о применении математических методов в биологических исследованиях при изучении групповых свойств биологических объектов. Содержанием биометрии является обработка данных наблюдений и экспериментов в биологических исследованиях, а ее аппаратом (методом) – теория вероятностей и математическая статистика.

Запуск программы Excel осуществляется так же, как и запуск других приложений пакета Microsoft Office: с помощью команды Microsoft Excel из меню программы или кнопки с таким же именем, расположенной на панели инструментов Microsoft Office.

После запуска Excel открывается окно, изображенное на рис. 1. Меню и панели инструментов Excel очень похожи на соответствующие элементы окна редактора Microsoft Word.

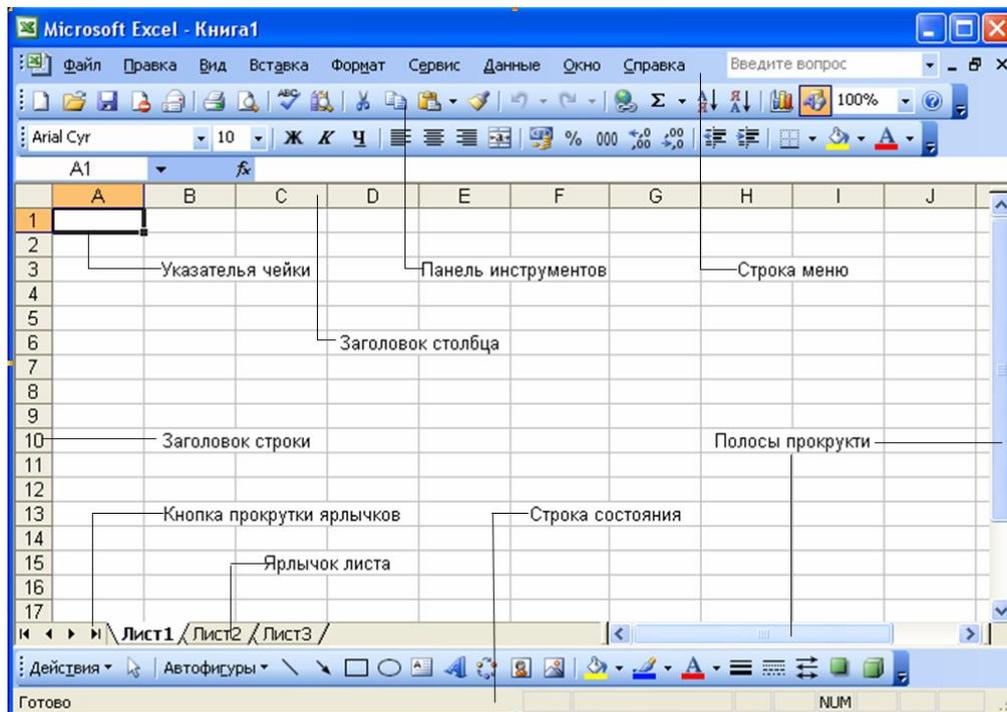


Рис. 1. Главное окно программы Excel

Основные операции с данными в Excel выполняются на рабочем листе. Рабочий лист разделен на строки и столбцы, которые при пересечении образуют ячейки. В ячейки вводится содержимое таблицы. Рабочий лист состоит из 256 столбцов и 65536 строк.

Строка формул находится под панелями инструментов и предназначена для обработки содержимого ячеек и имеет обозначение f_x . Эта строка предназначена для набора формул. Скорость работы с большими таблицами во многом зависит от способа перемещения по рабочему листу. Для перемещения по небольшим диапазонам ячеек достаточно клавиш управления курсором. Чтобы попасть в ту часть таблицы, которая не видна на экране и находится недалеко от видимой области, следует использовать полосы прокрутки. Добиться значительного ускорения при перемещении по рабочему листу можно благодаря специальным комбинациям клавиш. В следующей таблице описаны клавиши и комбинации, предназначенные для перемещения по листам рабочей книги.

Таблица 1. Описание клавиш и комбинации, предназначенные для перемещения по листам рабочей книги

Клавиша	Перемещение
[Home]	В начало текущей строки
[Ctrl+Home]	В ячейку A1
[Ctrl+End]	В последнюю заполненную ячейку таблицы
[↑]	На одну ячейку вверх
[↓]	На одну ячейку вниз
[→]	На одну ячейку вправо

Клавиша	Перемещение
[←]	На одну ячейку влево
[Ctrl+↑]	Вверх до первой заполненной ячейки
[Ctrl+↓]	Вниз до первой заполненной ячейки
[Ctrl+→]	Вправо до первой заполненной ячейки
[Ctrl+←]	Влево до первой заполненной ячейки
[Page Up]	Вверх на один экран
[Page Down]	Вниз на один экран
[Alt+Page Up]	Влево на один экран
[Alt+Page Down]	Вправо на один экран
[Ctrl+Page Up]	К следующему листу рабочей книги
[Ctrl+Page Down]	К предыдущему листу рабочей книги

Ввод данных. Ячейки электронной таблицы могут содержать самую разнообразную информацию: текст, числовые значения и формулы. При вводе данных Excel автоматически распознает их тип. Ввод данных выполняется в активной ячейке или в строке формул. Как только в ячейку будет введен хотя бы один символ, ее содержимое отобразится и в строке формул. Кроме того, и этой строке помнятся три кнопки для обработки содержимого ячейки.

Ввод данных завершается нажатием клавиши **[Enter]** или кнопки с изображением галочки в строке формул. Намного удобнее для завершения ввода использовать клавишу управления курсором. В этом случае после ввода можно сместиться в нужном направлении. Отменить ввод данных позволяют клавиша [Esc] и кнопка с крестиком в строке формул. Введенные числовые значения автоматически выравниваются по правому краю, а текстовые данные — по левому краю. Если длина введенного в ячейку текста превышает ширину этой ячейки, то после завершения ввода текст либо полностью отображается в таблице, закрывая, расположенные справа пустые ячейки, либо урезается по правому краю ячейки, если соседняя ячейка содержит какую-либо информацию. Урезанный текст полностью отображается только в строке формул.

Если числовые значения, полученные в результате вычисления формул, не помещаются в ячейку, вместо них на экране отображается соответствующее число символов **#**.

Если в ячейку вводится формула, то сразу по завершении ввода выполняются вычисления, и в ячейке отображается результат вычислений. Формула в Excel должна начинаться с математического оператора (например, со знака равенства). Так, после ввода формулы =1+6 в ячейке появится число 7, но в строке формул отобразится фактическое содержимое ячейки — формула. Запись 1+6 программа интерпретирует как текст.

Вычисление параметров вариационного ряда рассмотрим на примере данных суточного удоя коров. **Пример.** Имеется выборка (n=100). Требуется разнести по классам

данные суточного удоя. Для построения вариационного ряда, прежде всего следует найти лимиты — минимальное и максимальное значения вариант. В приведенной выборке они выделены.

Таблица 2. Пример выборки

Суточный удой коров									
21,9	21,4	27,7	17,0	12,3	21,7	23,4	25,7	21,2	20,3
23,8	24,1	25,9	21,4	20,7	18,5	22,5	23,0	18,5	25,7
20,1	21,3	15,7	24,8	19,3	22,2	22,9	14,9	26,1	20,5
14,6	27,8	22,4	16,7	22,9	25,3	22,7	19,7	15,2	21,3
22,1	20,5	19,7	24,5	29,6	22,3	19,1	23,5	25,9;	17,2
15,5	18,1	23,9	25,4	20,4	13,2	19,6	24,4	18,2	24,8
24,2	20,9	21,0	16,5	20,9	23,2	27,2	21,1	26,3	18,6
17,2	17,8	31,2	25,0	20,7	18,3	23,7	16,1	16,2	21,6
23,0	20,7	25,3	13,9	17,3	21,8	14,1	19,0	21,9	18,7
28,5	21,2	19,9	24,8	22,7	16,4	20,6	23,5	22,2	19,5

Числовые значения вводим в рабочее поле Excel начиная с ячейки В2 до В11, далее С2-С11 и т.д. до ячейки К11. В итоге получим таблицу состоящую из 10 строк и 10 столбцов. В ячейку В1 можно ввести название данных текстом, например: написать «Суточный удой коров».

В программе Excel имеются стандартные программы вычисления параметров статистики, которые можно использовать указав название этой программы. В данном примере рассчитаем следующие показатели: количество данных, минимальное и максимальное значения ряда, среднее арифметическое, среднее квадратическое отклонение, значение сигмы, ошибку средней, коэффициент вариации и границы доверительного интервала.

Решение:

- 1) количество данных – в ячейку В13 (обозначение ячеек вводят латинским шрифтом) введем = СЧЕТ(В2:К11), затем нажимаем Enter, в ячейке появится цифра 100. Что означает, всего данных сто.

2) вычисление минимального и максимального значений ряда – в ячейку B14 введем =МИН(B2:K11) «Enter», а в ячейку B15 =МАКС(B2:K11) «Enter». В ячейках B14 и B15 появятся цифры 12,3 и 31,2.

3) вычисление остальных величин:

в ячейку B16 введем =СРЗНАЧ (B2:K11) «Enter» – это среднее значение;

среднее квадратическое отклонение – в ячейку B17 введем = ДИСПРА(B2:K11)

«Enter»; значение сигма – в ячейку B18 введем =КОРЕНЬ(B17) «Enter»;

ошибка средней – в ячейку B19 введем =B18/КОРЕНЬ(B13) «Enter»;

коэффициент вариации (в %) – в ячейку B20 – введем =(B19*100)/B16 «Enter».

Определение границ доверительного интервала. Определим доверительный интервал для генерального среднего с уровнем надежности 95%. Исходные данные: объем выборки – 100, среднее – 21,237, стандартное отклонение σ – 3,734.

Решение с помощью функции Лапласа.

1) По таблице значений функции Лапласа найдем значение t , которое удовлетворяет условию

$$2\Phi(t)=\gamma, \quad (1)$$

т.е.

$$\Phi(t) = \frac{\gamma}{2}, \quad (2)$$

$$\Phi(t) = \frac{0,95}{2} = 0,475 \quad (3)$$

Мы увидим, что $t = 1,96$.

Если под рукой нет учебника с таблицей функции Лапласа, то можно воспользоваться программой Excel. Применим для вычисления t формулу первой главы:

$$\Phi(t)=\text{НОРМСТРАСП}(t)-0,5.$$

Мы должны вместо уравнения $\Phi(t)=0,495$ рассмотреть уравнение

$$\text{НОРМСТРАСП}(t)-0,5=0,475$$

или найти t из соотношения

$$\text{НОРМСТРАСП}(t)=0,977.$$

Воспользуемся обратной функцией для НОРМСТРАСП(t).

Введем в произвольную ячейку программы Excel (например, G14) формулу

$$=\text{НОРМСТРАСП}(0,975),$$

получим значение 1,959963985, что совпадает приближенно с табличным значением $t = 1,96$.

2) Теперь найдем границы доверительного интервала

$$\Delta = t \frac{\sigma}{\sqrt{n}} = 1,96 \cdot \frac{3,734}{\sqrt{100}} = 0,731 \quad (4)$$

Введем в ячейку G16 = ДОВЕРИТ(0,05;3,73;100) «Enter», получим 0,731.

Тогда
$$\bar{x} - t \frac{\sigma}{\sqrt{n}} = \bar{x} - \Delta = 21,237 - 0,731 = 20,506 \quad (5)$$

$$\bar{x} + t \frac{\sigma}{\sqrt{n}} = \bar{x} + \Delta = 21,237 + 0,731 = 21,97 \quad (6)$$

	A	B	C	D	E	F	G	H	I	J	K
1											
2		21,9	21,4	27,7	17	12,3	21,7	23,4	25,7	21,2	20,3
3		23,8	24,1	25,9	21,4	20,7	18,5	22,5	23	18,5	25,7
4		20,1	21,3	15,7	24,8	19,3	22,2	22,9	14,9	26,1	20,5
5		14,6	27,8	22,4	16,7	22,9	25,3	22,7	19,7	15,2	21,3
6		22,1	20,5	19,7	24,5	29,6	22,3	19,1	23,5	25,9	17,2
7		15,5	18,1	23,9	25,4	20,4	13,2	19,6	24,4	18,2	24,8
8		24,2	20,9	21	16,5	20,9	23,2	27,2	21,1	26,3	18,6
9		17,2	17,8	31,2	25	20,7	18,3	23,7	16,1	16,2	21,6
10		23	20,7	25,3	13,9	17,3	21,8	14,1	19	21,9	18,7
11		28,5	21,2	19,9	24,8	22,7	16,4	20,6	23,5	22,2	19,5
12											
13		100									
14	Мин.	12,3									
15	Макс.	31,2									
16	Среднее знач.	21,237					Границы доверит интервала Δ	0,731067			
17	Среднее квадратич. Откл.	13,94613					Границы доверит интервала от	20,50593			
18	Сигма	3,734452					Границы доверит интервала до	21,96807			
19	Ошибка средней	0,373445									
20	Кoeffициент вариации, %	1,758465									
21											

Рис. 2. Параметры вариационного ряда

Доверительный интервал имеет вид: $(20,506 < \bar{X} < 21,97)$. В приведенном примере можно в ячейки записывать и пояснения, выбор ячеек для ввода чисел может быть любым. Решенное задание имеет вид (рис. 2).

Вычисление коэффициента корреляции

Для примера возьмем таблицу суточного удоя и жирности молока. В ячейки B3, B4, B5 и до B102 вводим значения суточного удоя в кг, а в ячейки D3 до D102 значения жирности молока (рис. 3).

	A	B	C	D	E	F	G	H
1			Суточный удой (x) и жирность молока (y) коров, кг					
2			x	y				
3			28,8	512				
4			20,2	472				
5			21,4	489		Колич. элем.	100	
6			20,6	482				
7			23,7	468		Коэффиц. уравнения a	6,419735	
8			21	479				
9			25,5	515		Коэффиц. уравнения b	330,9887	
10			21,7	451				
11			20,9	475		Коэффиц. коррел r	0,57179	
12			14,8	402				
13			20,7	473				
14			21	467				
15			23,5	458				
16			26,2	534				
17			16,3	433				
18			24,4	528				
19			20,3	452				
20			12,3	380				

Рис. 3. Вычисление коэффициента корреляции (1-й способ – ввод всех данных в два столбца, на рисунке показаны только первые 20 значений)

Затем в ячейку G11 вводим =КОРРЕЛ(C3:C102;D3:D102) «Enter» получим в этой ячейке значение коэффициента корреляции – 0,572.

Для более компактного ввода данных можно их вводить как на Рис.3

Тогда в ячейку E22 вводим =КОРРЕЛ(C3:N19;I3:N19) «Enter». Получим значение 0,62699.

Для проверки значимости коэффициента корреляции вычислим по формуле значение $T_{набл}$.

Для этого введите в ячейку E24 формулу =E22*КОРЕНЬ((100-2)/(1-E22^2)).

Получим $T_{набл} = -7,967$.

В ячейку E26 введите формулу =СТЮДРАСПОБР(0,01;98), получим значение критерия Стьюдента 2,626.

Так как $T_{набл} = -7,967 > t_{кр}(\alpha; k) = 2,6$, нулевую гипотезу отвергаем, т.е. делаем вывод: коэффициент корреляции значим с уровнем доверия 99% ($1 - 0,01 = 0,99$).

Значение выборочного коэффициента корреляции положительно и близко к единице, что означает очень сильную прямую связь между рассматриваемыми признаками.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1			Суточный удой (x) и жирность молока (y) коров, кг											
2			x	x	x	x	x	x	y	y	y	y	y	y
3			12,3	18,8	15,2	13,9	22,8	29	3,89	4,08	4,59	4,21	4,06	3,66
4			21,4	20,2	23,4	17	21,1	20,7	3,82	3,93	3,48	4,71	4,08	4,01
5			18,9	21,4	24,8	20,9	23,1	17,5	3,68	3,78	4,07	3,85	3,62	4,59
6			21,8	20,6	24,4	15,9	20,2	22,3	3,19	3,47	3,77	4,08	3,78	3,81
7			20,9	13,7	16	17,8	15,2	27	3,18	4,12	4,31	4,24	4,5	3,66
8			21,9	21	23	14,5	20,5	20,9	3,51	3,6	3,56	4,07	3,74	3,61
9			17,8	15,5	24,3	17,6	13,4	29,6	3,75	4,42	3,95	4,16	4,2	3,43
10			20	11,7	19,6	23,8	14,2	25,1	4,03	4,47	4,03	3,68	4,16	3,42
11			21,1	20,9	15,6	15,8	20,5	22,1	4,03	4,02	4,27	4,73	3,78	4,03
12			17,5	14,8	21,6	16,4	20,9	27,4	4,27	3,97	3,9	4,62	3,37	3,22
13			21,6	18,7	14,2	15,6	24,6	16,4	3,58	4,15	4,31	4,62	3,74	4,72
14			14,8	17	20,1	20,1	19,4	22,3	4,34	4,25	4,1	4,07	4,16	4,17
15			21,1	13,5	21,4	14,9	21,2	28,1	4,13	4,08	3,79	4,55	3,17	3,88
16			18,1	16,2	15,2	21,8	27,4	21,7	4,28	4,18	4,72	3,78	3,59	4,03
17			15,8	16,3	21,1	21,3	21,8	12,8	4,08	3,94	3,45	4,08	3,58	4,29
18			21,7	14,4	20,4	12,6	20,8	20,8	4,02	4,54	3,45	4,05	3,38	3,89
19			20,7	20,3	19,8			29,4	3,73	3,89	3,5			3,98
20														
21														
22					Кэф. коррел.									
23						-0,62699								
24					Значение Тнабл.									
25						-7,96758								
26					Кэф. Стьюдента									
27						2,626931								

Рис. 4 Вычисление коэффициента корреляции (2-й способ – ввод всех данных по столбцам, последовательно)

Таким образом, биометрию можно использовать при планировании и обработке лишь тех биологических экспериментов и наблюдений, результаты которых могут быть отнесены к теоретическому понятию статистической совокупности.

Литература

1. Ларцева С.Х., Муксинов М.К. Практикум по генетике. — М.: Агропромиздат, 1985. — 288 с.
2. Фолконер Д.С. Введение в генетику количественных признаков. — М.: Агропромиздат, 1985. — 486 с.
3. Плохинский Н.А. Биометрия. — М.: Московский университет, 1980. — 150 с.
4. Антипов Г.П. Генетика с биометрией. — М.: Изд. МСХА, 1995. — 166 с.
5. Рокицкий П.Ф. Введение в статистическую генетику. — Минск: Высшая школа, 1974. — 447 с.
6. Меркурьева Е.К., Шангин-Березовский Г.Н. Генетика. — М.: Колос, 1991. — 446 с.
7. Баяхов А.Н., Моисейкина Л.Г., Турдуматов Б.М., Мендалиева Ш. Методы математической статистики в биологии. — Уральск: ЦНТИИ, 2016. — 96 с.