

Сдвижков О.А.

НЕПАРАМЕТРИЧЕСКАЯ СТАТИСТИКА

в MS Excel и VBA

УДК 744.4:004.92
ББК 30.11+32.973.26-018.2
С27

Сдвижков О. А.
С27 Непараметрическая статистика в MS Excel и VBA. – М.: ДМК Пресс, 2014. – 172 с.: ил.

ISBN 978-5-94074-917-2

В книгу вошли основные сведения по MS Excel и классическим методам непараметрической статистики, применяемым к независимым выборкам, парным наблюдениям и таблицам сопряженности, реализующие эти методы программы VBA и технологии решения типовых задач в MS Excel. Данные технологии представлены, как пошаговыми решениями (без применения макросов), так и автоматическими, когда задача решается одним макросом, возвращающим значение статистики, критерий принятия основной гипотезы и вывод о том, какую гипотезу следует принять.

Книга ориентирована на студентов вузов, изучающих статистические методы, но будет полезна и более широкому кругу пользователей MS Excel.

УДК 744.4:004.92
ББК 30.11+32.973.26-018.2

Все права защищены. Любая часть этой книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами без письменного разрешения владельцев авторских прав.

Материал, изложенный в данной книге, многократно проверен. Но поскольку вероятность технических ошибок все равно существует, издательство не может гарантировать абсолютную точность и правильность приводимых сведений. В связи с этим издательство не несет ответственности за возможные ошибки, связанные с использованием книги.

© Сдвижков О. А., 2014
© Оформление, издание, ДМК Пресс, 2014
ISBN 978-5-94074-917-2

Содержание

Предисловие	5
Глава 1. Теория вероятностей и математическая статистика в MS Excel	7
§1. Встроенные функции дискретных распределений.....	9
§2. Макросы для дискретных распределений	18
§3. Встроенные функции непрерывных распределений.....	23
§4. Инструмент Гистограмма	26
§5. Инструмент Описательная статистика	34
§6. Встроенные средства корреляционного анализа	38
§7. Макрос Correlation.....	45
§8. Тесты надстройки Анализ данных.....	47
Глава 2. Непараметрические критерии для независимых выборок	58
§1. Критерий Колмогорова–Смирнова	59
§2. Критерий Катценбайссера–Хакли	65
§3. Критерий Вилкоксона	69
§4. Критерий Манна–Уитни	76
§5. Критерий серий Вальда–Вольфовитца	82
§6. Сериальный критерий Рамачандрана–Ранганатана	94
Глава 3. Непараметрические критерии для пар наблюдений	97
§1. Критерий знаков	98
§2. Критерий Фишера.....	104
§3. Знаковый критерий Вилкоксона	107
§4. Модификации критерия Вилкоксона	114
§5. Критерий Спирмена	123
§6. Критерий Кендалла.....	128
§7. Критерий Ван дер Вардена	132
§8. Критерий Ширахате	136
Глава 4. Непараметрические критерии для таблиц сопряженности	141
§1. Четырехклеточный χ^2 критерий	142
§2. Общий случай χ^2 критерия.....	149
§3. Точный критерий Фишера	153

§4. G-критерий Вулфа	156
§5. Критерий Макнимары	159
§6. Биномиальные критерии	163
§7. Критерий Ле Роя	167
Литература	170

Теория вероятностей и математическая статистика поддерживаются в MS Excel встроенными функциями (таблица 1.1) и надстройкой (инструментами) **Анализ данных** (рис. 1.1, 1.2).

Таблица 1.1

ФРАСП	МАКС	СРЗНАЧ
ФРАСПОБР	МАКСА	СРЗНАЧА
ZТЕСТ	МЕДИАНА	СРОТКЛ
БЕТАОБР	МИН	СТАНДОТКЛОН
БЕТАРАСП	МИНА	СТАНДОТКЛОНА
БИНОМРАСП	МОДА	СТАНДОТКЛОНП
ВЕЙБУЛЛ	НАИБОЛЬШИЙ	СТАНДОТКЛОНПА
ВЕРОЯТНОСТЬ	НАИМЕНЬШИЙ	СТОШУХ
ГАММАОБР	НАКЛОН	СТЬЮДРАСП
ГАММАРАСП	НОРМАЛИЗАЦИЯ	СТЬЮДРАСПОБР
ГАММАНЛОГ	НОРМОБР	СЧЕТ
ГИПЕРГЕОМЕТ	НОРМРАСП	СЧЕТЕСЛИ
ДИСП	НОРМСТОБР	СЧЕТЗ
ДИСПА	НОРМСТРАСП	СЧИТАТЬПУСТОТЫ
ДИСПР	ОТБИНОМРАСП	ТЕНДЕНЦИЯ
ДИСПРА	ОТРЕЗОК	ТТЕСТ
ДОВЕРИТ	ПЕРЕСТ	УРЕЗСРЕДНЕЕ
КВАДРОТКЛ	ПЕРСЕНТИЛЬ	ФИШЕР
КВАРТИЛЬ	ПИРСОН	ФИШЕРОБР
КВПИРСОН	ПРЕДСКАЗ	ФТЕСТ
КОВАР	ПРОЦЕНТРАНГ	ХИ2ОБР
КОРРЕЛ	ПУАССОН	ХИ2РАСП
КРИТБИНОМ	РАНГ	ХИ2ТЕСТ
ЛГРФПРИБЛ	РОСТ	ЧАСТОТА
ЛИНЕЙН	СКОС	ЭКСПРАСП
ЛОГНОРМОБР	СРГАРМ	ЭКССЕСС
ЛОГНОМРАСП	СРГЕОМ	

Меню встроенных функций (Мастер функций) открывается кнопкой $f(x)$, по каждой из них в MS Excel можно получить подробную справку. Поэтому специально перечислять их назначения, как и значения параметров, нет необходимости, тем более что многие из них в дальнейшем будут применяться в конкретных задачах.

Пакет **Анализ данных**, если его нет в меню **Сервис**, загружается командами **Сервис** \Rightarrow **Надстройки**, после чего надо поставить галочку в поле **Анализ данных** и подтвердить **ОК**.

Щелчок по опции **Анализ данных** меню **Сервис** открывает список инструментов:

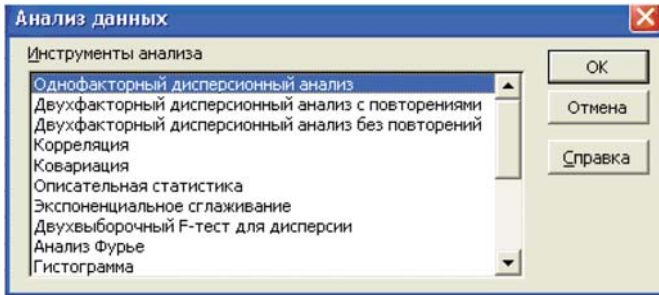


Рис. 1.1

Продолжение списка инструментов на рис. 1.2.

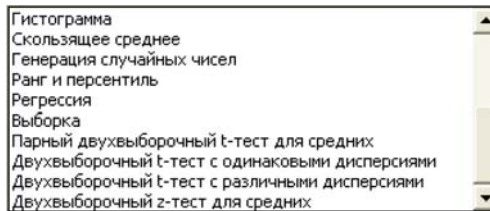


Рис. 1.2

§1. Встроенные функции дискретных распределений

Встроенные функции комбинаторики ПЕРЕСТ, ЧИСЛКОМБ и ФАКТР, причем первая входит в статистические функции, а вторая и третья – в математические.

$\text{ПЕРЕСТ}(n; m)$ – возвращает число выборок из n элементов по m , каждая из которых отличается от остальных или составом элементов, или их порядком.

Пусть требуется подсчитать, сколько вариантов шахматных команд по три игрока (1-я, 2-я, 3-я доски) можно составить из четырех игроков. Открываем диалоговое окно ПЕРЕСТ, вводим данные и получаем результат:

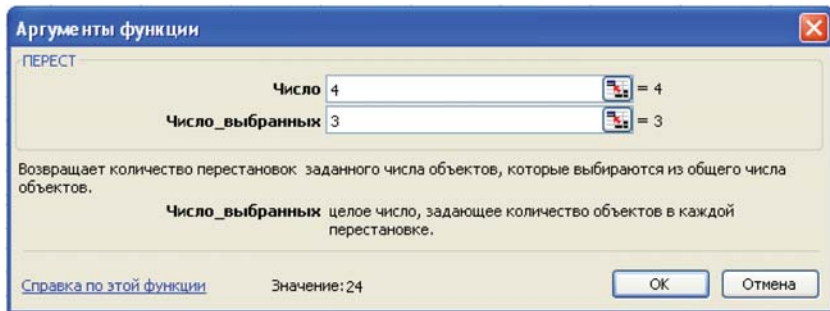


Рис. 1.3

Команда **ОК** вставляет результат на рабочий лист.

Если $m = n$, то встроенная функция ПЕРЕСТ возвращает $n!$, так же как встроенная функция ФАКТР.

ЧИСЛКОМБ(n ; m) – возвращает число сочетаний C_n^m , то есть число выборок из n элементов по m , каждая из которых отличается от остальных хотя бы одним элементом.

Пусть требуется подсчитать, сколько вариантов различных команд по пляжному волейболу (в команде 2 человека) можно составить из пяти человек. Тогда, используя встроенную функцию ЧИСЛКОМБ, получаем 10:

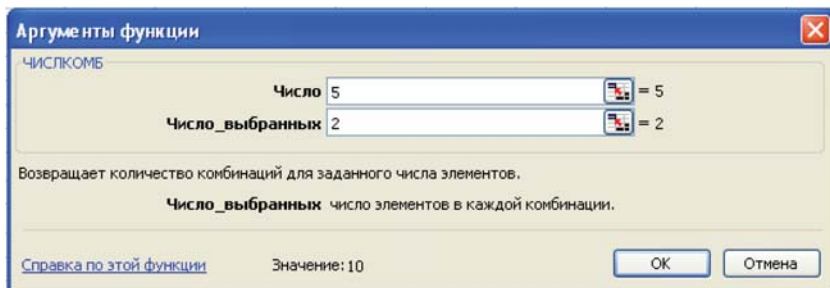


Рис. 1.4

Встроенная функция БИНОМРАСП(m , n , p , ЛОЖЬ) возвращает значение, получаемое по формуле Бернулли

$$P_n(m) = C_n^m p^m (1 - p)^{n-m},$$

встроенная функция ПУАССОН(m , λ , ЛОЖЬ) – значение, получаемое по формуле Пуассона

$$P_n(m) = \frac{\lambda^m}{m!} e^{-\lambda}, \quad \lambda = np.$$

При замене значения ЛОЖЬ (0) на значение ИСТИНА (1) они возвращают значения интегральных функций распределений.

Пусть дискретная случайная величина X имеет, например, биномиальный закон распределения с параметрами $n = 5$, $p = 0,75$. Тогда ее возможные значения 0, 1, 2, 3, 4, 5. Вычислим вероятности, с которыми она их принимает, составим закон распределения, сделаем проверку и построим многоугольник распределения.

1. В диапазон A1:F1 вводим значения X .
2. Выделяем ячейку A2, открываем диалоговое окно БИНОМ-РАСП и заполняем поля (рис. 1.5).

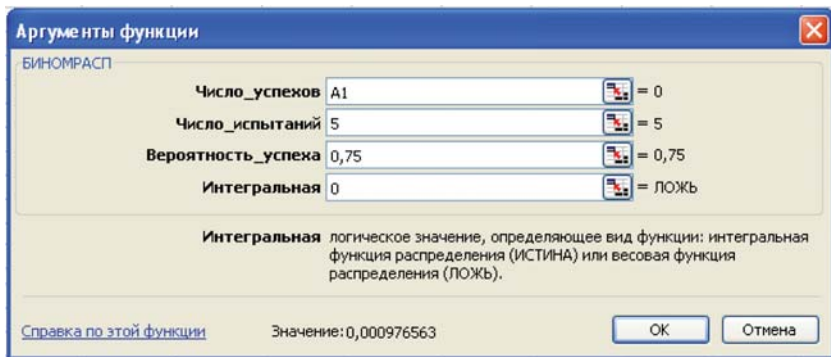


Рис. 1.5

3. Команда **ОК** вставляет в ячейку A2 формулу =БИНОМ-РАСП(A1;5;0,75;0).
4. Методом «протаскивания» маркера заполнения копируем ее в остальные ячейки диапазона A2:F2.
5. Выделяем диапазон A2:F2 и, щелчком по кнопке Σ , убеждаемся в том, что сумма равна 1, то есть в диапазоне A1:F2 получен, действительно, закон распределения дискретной случайной величины:

	A2	fx =БИНОМРАСП(A1;5;0,75;0)						
	A	B	C	D	E	F	G	
1	0	1	2	3	4	5		
2	0,000977	0,014648	0,087891	0,263672	0,395508	0,237305	1	

Рис. 1.6

6. В графическом редакторе **Мастер диаграмм**, выделяя диапазон A1:F2, строим точечную диаграмму:



Рис. 1.7

Вычисление значения интегральной функции данного биномиального распределения при $x = 5$ дает $F(5) = 1$ (рис. 1.8), что показывает, что в MS Excel интегральная функция распределения $F(x) = P(X \leq x)$, хотя принято $F(x) = P(X < x)$ [6].

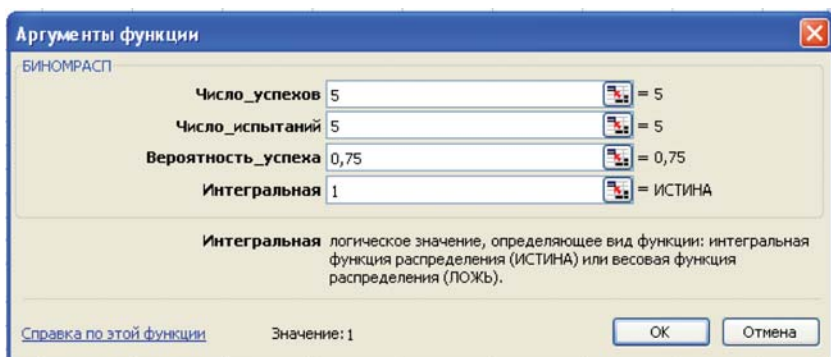


Рис. 1.8

График интегральной функции рассматриваемого распределения на отрезке $[0, 6]$ с шагом $0,2$, построенный с помощью встроенной функции БИНОМРАСП и графического редактора **Мастер диаграмм**, показан на рис. 1.9.

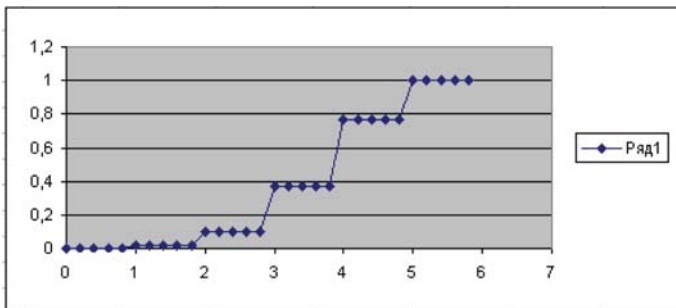


Рис. 1.9

Задача 1.1. Дискретная случайная величина X задана законом распределения

$$\begin{array}{l} x_i \quad 3 \quad 4 \quad 7 \quad 10 \\ p_i \quad 0,2 \quad 0,3 \quad 0,1 \quad 0,4 \end{array}$$

Найти интегральную функцию распределения и построить график.

Технология решения.

1. Вводим в ячейку A1 значение 2,8 и задаем в первом столбце арифметическую прогрессию с шагом 0,2, предельное значение 11.
2. В ячейку B1 вводим формулу интегральной функции заданного распределения

$$=ЕСЛИ(A1 \leq 3; 0; ЕСЛИ(A1 \leq 4; 0,2; ЕСЛИ(A1 \leq 7; 0,5; ЕСЛИ(A1 \leq 10; 0,6; 1))))$$
и копируем ее в ячейки столбца B, соответствующие заполненным ячейкам столбца A.
3. Выделяя полученную таблицу и применяя редактор **Мастер диаграмм**, приходим к графику:

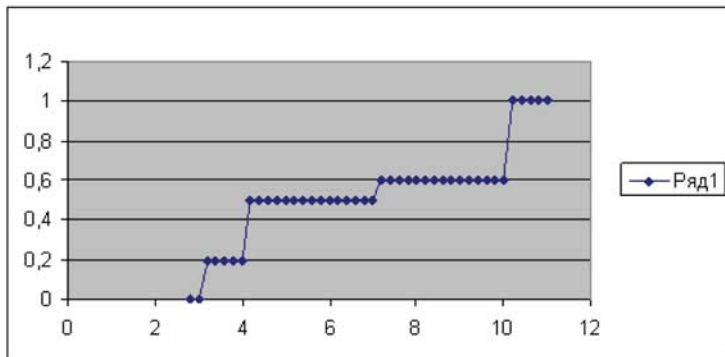


Рис. 1.10

Ответ:

$$F(x) = \begin{cases} 0, & x \leq 3 \\ 0,2, & 3 < x \leq 4 \\ 0,5, & 4 < x \leq 7 \\ 0,6, & 7 < x \leq 10 \\ 1, & x > 10 \end{cases} .$$

Найдем математическое ожидание $M[X] = \sum_{i=1}^n x_i p_i$, дискретной случайной величины X , заданной в задаче 1.1.

1. Введем в диапазон A1:D2 закон распределения (рис. 1.11).
2. В ячейке A3 запишем формулу =A1*A2 и скопируем ее маркером заполнения в остальные ячейки диапазона A3:D3.
3. Выделяя диапазон A3:D3, и делая щелчок по кнопке Σ панели инструментов, получаем в ячейке E3 искомое значение, то есть $M[X] = 6,5$:

	A	B	C	D	E
1	3	4	7	10	
2	0,2	0,3	0,1	0,4	
3	0,6	1,2	0,7	4	6,5

Рис. 1.11

Найдем дисперсию $D[X] = \sum_{i=1}^n (x_i - M[X])^2 \cdot p_i$ случайной величины задачи 1.1. Продолжаем вычисления.

4. В ячейку A4 вводим формулу $= (A1 - \$E\$3)^2 * A2$ и копируем ее маркером заполнения в остальные ячейки диапазона A4:D4.
5. Выделяя диапазон A4:D4, и применяя кнопку Σ , получаем в ячейке E4 значение дисперсии, то есть $D[X] = 9,25$:

	A	B	C	D	E
1	3	4	7	10	
2	0,2	0,3	0,1	0,4	
3	0,6	1,2	0,7	4	6,5
4	2,45	1,875	0,025	4,9	9,25

Рис. 1.12

Сделаем проверку полученного результата, применяя формулу:
 $D[X] = M[X^2] - M[X]^2$.

Продолжаем вычисления.

6. В ячейку A5 вводим формулу $= A1^2 * A2$ и копируем ее в остальные ячейки диапазона A5:D5.
7. Выделяя диапазон A5:D5, и применяя кнопку Σ , получаем $M[X^2] = 51,5$ (ячейка E5).
8. В ячейку F5 вводим формулу $= E5 - E3^2$ Она возвращает значение, совпадающее с полученным по первой формуле в ячейке E4:

	A	B	C	D	E	F
1	3	4	7	10		
2	0,2	0,3	0,1	0,4		
3	0,6	1,2	0,7	4	6,5	
4	2,45	1,875	0,025	4,9	9,25	
5	1,8	4,8	4,9	40	51,5	9,25

Рис. 1.13

По данным диапазона A1:D2 (рис. 1.11) значение $M[X]$ можно получить также функцией СУММПРОИЗВ. Достаточно открыть ее диалоговое окно и ввести ссылки на диапазоны A1:D1, A2:D2 (рис. 1.14).

Задача 1.2. Магазин получил 1000 бутылок минеральной воды. Вероятность того, что при перевозке бутылка окажется разбитой, равна 0,005. Найти вероятности того, что магазин получит разбитых бутылок: а) ровно три; б) менее трех; в) более трех; г) хотя бы одну.

Технология решения. Так как $n=1000$ велико, а $p=0,003$ мало, то применима формула Пуассона, в которой $\lambda = np = 5$. Найдем по ней вероятности, с которыми принимаются значения 0, 1, 2, 3.

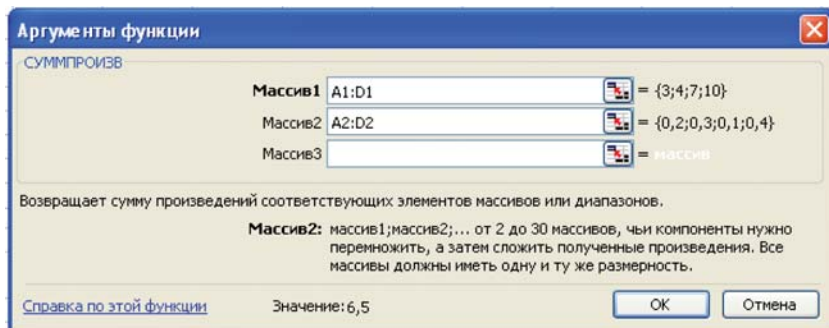


Рис. 1.14

1. Вводим их в диапазон A1:A3 (рис. 1.16).
2. Выделяем ячейку B1, открываем диалоговое окно ПУАССОН и вводим данные, подтверждая командой ОК (рис. 1.15).

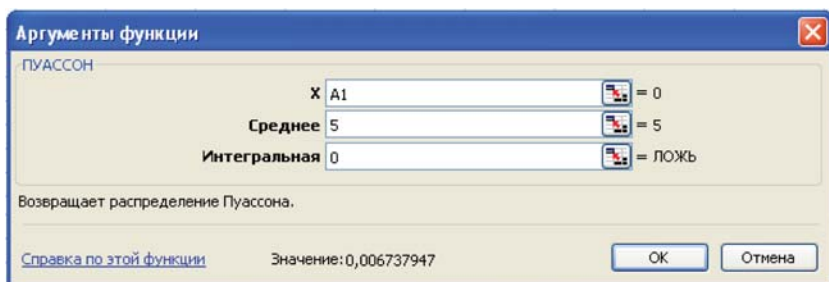


Рис. 1.15

3. Копируем формулу ячейки B1 в ячейки B2:B4. Ответ на первый вопрос находится в ячейке B4.
4. Формула = B1+B2+B3, которую запишем в C1, дает ответ на второй вопрос.
5. Кнопкой Σ находим в ячейке B5 сумму значений диапазона B1:B4. Формулой =1-B5 ячейки C2 получаем ответ на третий вопрос.
6. Последняя величина находится по формуле = 1-B1, ее помещаем в ячейку C3. В результате фрагмент листа Excel с решением задачи принимает вид:

	A	B	C	D
1	0	0,006738	0,124652	
2	1	0,03369	0,734974	
3	2	0,084224	0,993262	
4	3	0,140374		
5		0,265026		

Рис. 1.16

Ответ: а) В4; б) С1; в) С2; г) С3.

Задача 1.3. В партии из 10 деталей имеется 8 стандартных. Наудачу отобраны две детали. Составить закон распределения случайной дискретной величины X – числа стандартных деталей среди отобранных.

Технология решения. Возможные значения случайной величины X : 0, 1, 2. Закон распределения – гипергеометрический.

1. В диапазоне А1:С1 вводим 0, 1, 2 соответственно.
2. Выделяем ячейку А2, открываем диалоговое окно ГИПЕРГЕОМЕТ и вводим данные, подтверждая **ОК**.

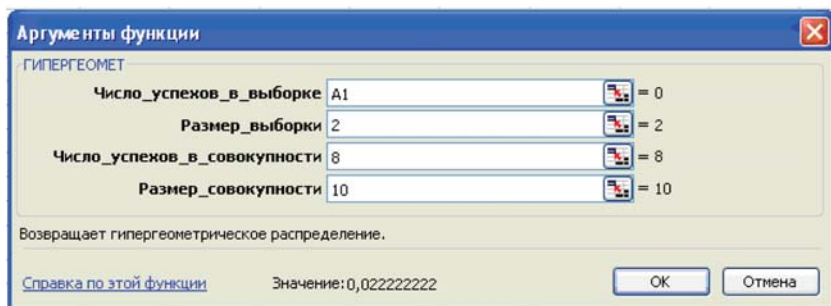


Рис. 1.17

3. Копируем формулу ячейки А2 в ячейки В2:С2 и переходим в диапазоне А2:С2 к формату «Дробный», с дробями до двух цифр, что приводит к следующему закону распределения вероятностей:

	A	B	C
1	0	1	2
2	1/45	16/45	28/45

Рис. 1.18

Ответ: $P(X = 0) = \frac{1}{45}$, $P(X = 1) = \frac{16}{45}$, $P(X = 2) = \frac{28}{45}$.

§2. Макросы для дискретных распределений

1. Макрос `Discrete` вычисляет автоматически математическое ожидание $M[X]$, дисперсию $D[X]$ и среднее квадратическое отклонение $s[X] = \sqrt{D[X]}$ таблично заданной дискретной случайной величины X . Для применения макроса надо ввести данные таблицы в первые две строки рабочего листа, начиная с ячейки `A1`, выделить диапазон с данными и запустить макрос на исполнение. Например, при выделенном диапазоне `A1:D2`, показанном на рис. 1.11, он возвращает полученные ранее результаты:



Рис. 1.19

Код макроса `Discretel`

```
Sub Discrete ()
Dim n As Integer, x() As Single, p() As Single, _
m As Single, d As Single, s As Single
n = Selection.Columns.Count
ReDim x(1 To n)
ReDim p(1 To n)
m = 0: d = 0
For i = 1 To n
    x(i) = Cells(1, i).Value: p(i) = Cells(2, i).Value
    m = m + x(i) * p(i)
Next
For i = 1 To n
    d = d + (x(i) - m) ^ 2 * p(i)
Next
s = Sqr(d)
MsgBox ("1. M[X]=" & m & Chr(13) & "2. D[X]=" & d _
& Chr(13) & "3. s[X]=" & s)
End Sub
```

2. Макрос `Discrete_L` для биномиального, пуассоновского и геометрического законов распределений находит $P(m_1 \leq X \leq m_2)$.

Задача 1.4. Страховой агент в результате каждого визита заключает договор с вероятностью $1/3$. Найти вероятность, что из 10 визитов не менее пяти завершатся заключением договора.

Технология решения. Задача сводится к нахождению вероятности $P_{10}(5 \leq X \leq 10)$ для биномиального закона распределения. Вызываем макрос `Discrete_L` и запускаем его на исполнение, что открывает окно ввода номера выбранного закона распределения:

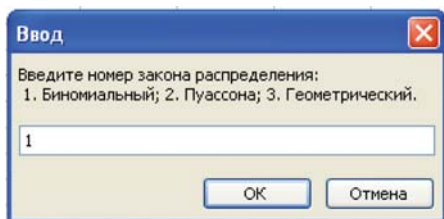


Рис. 1.20

По команде **ОК** открывается следующее окно ввода:

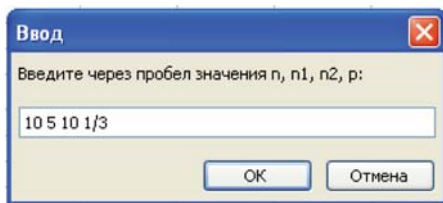


Рис. 1.21

Команда **ОК** окна возвращает искомую вероятность:

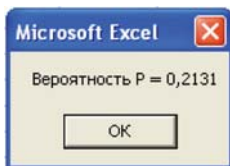


Рис. 1.22

Ответ: 0,2131.

Задача 1.5. Случайная величина X имеет распределение Пуассона с параметром $\lambda = 15$. Найти вероятность $P(12 \leq X \leq 16)$.

Технология решения. Запускаем макрос Discrete_L на исполнение и вводим 2:

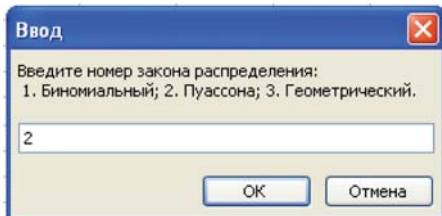


Рис. 1.23

Команда **ОК** открывает следующее окно ввода:

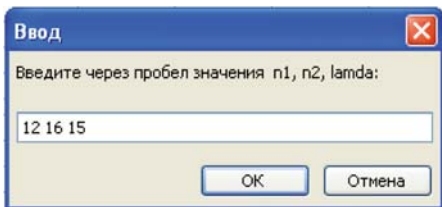


Рис. 1.24

Команда **ОК** возвращает:

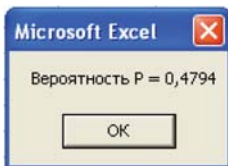


Рис. 1.25

Ответ: 0,4794.

Задача 1.6. Вероятность завести машину на морозе с одной попытки равна 0,8. Найти вероятность, что машина будет заведена не более чем с трех попыток.