

Касымова Тумар Джапашевна  
к.ф.-м.н., доцент,  
кафедра алгебры, геометрии, топологии и  
преподавания высшей математики  
заведующая отделом науки и инноваций  
Кыргызский национальный университет им. Ж. Баласагына  
г. Бишкек, ул. Фрунзе, 547  
E-mail: [tumar2000@mail.ru](mailto:tumar2000@mail.ru)

Ткачман Михаил Леонидович  
старший преподаватель  
кафедры информационных систем и технологий  
факультет инновационных и информационных технологий  
Кыргызский национальный университет им. Ж. Баласагына  
г. Бишкек, ул. Фрунзе 547  
E-mail: [tkachmanmichail@yandex.ru](mailto:tkachmanmichail@yandex.ru)

Рослова Инна Николаевна  
старший преподаватель  
кафедры компьютерных технологий и инженерии  
факультет инновационных и информационных технологий  
Кыргызский национальный университет им. Ж. Баласагына  
г. Бишкек, ул. Фрунзе, 547  
E-mail: [for\\_inna@bk.ru](mailto:for_inna@bk.ru)

### **Разработка WEB - системы на базе PHP MYSQL-технологий для анализа и обработки статистических данных**

**Аннотация.** В статье разработана автоматизированная информационная система для анализа и обработки статистических данных на основе математических моделей. Проведен сравнительный анализ двух методов обработки статистических данных для вероятностной величины: классическим и биномиального распределения определения и сравнения расчетных параметров. Показаны особенности работы в различных средах.

**Ключевые слова:** Биномиальное распределение, формула, математическое ожидание, дискретная случайная величина, график.

Касымова Т.Дж., Ткачман М.Л., Рослова И.Н.

### **PHP MYSQL - технологияларга негизделген статистикалык маалыматтарды талдоо жана иштеп чыгуу үчүн WEB-системаны иштеп өнүктүрүү**

**Аннотация.** Математикалык моделдер боюнча статистикалык маалыматтарды талдоо жана иштеп чыгуу үчүн технологиялар маалыматтык системасы иштелип чыккан. Эки статистикалык ыкмаларын салыштырмалуу талдоо: классикалык жана биномиалдык бөлүштүрүү менен эсептелген параметрлерин аныктоо жүргүзүлгөн. Ар кандай айрым шарттарда өзгөчөлүктөрү көрсөтүлгөн.

**Урунттуу сөздөр:** Биномиалдык бөлүштүрүү, формула, математикалык күтүү, дискреттик кокус сан, график.

Tumar J. Kasymova, Mikhail L. Tkachman, Inna N. Roslova  
**WEB - system development based on PHP MYSQL - technologies  
for analysis and processing of statistical data**

**Summary.** In the paper an automated information system for analyzing and processing statistical data based on mathematical models is developed. A comparative analysis of two methods for processing statistical data for a probability value is carried out: the classical and binomial distribution of the definition and comparison of calculated parameters. Features of work in various environments are shown.

**Key words:** Binomial distribution, formula, mathematical expectation, discrete random variable, graph.

Для статистического анализа и обработки данных существуют различные программные продукты, как встроенные в MS Excel - внутренние, так и внешние. Так, например, в MS Excel с применением надстройки «Анализ данных» можно достаточно успешно и наглядно провести анализ описательной статистики дискретной случайной величины [1]. Полное методическое описание статистической обработки данных посредством свойств и инструментов в среде MS Excel приведено в [1] с наглядным представлением таблиц и графиков.

Напомним основные формулы описательной статистики из [1, 2]

1. Закон распределения случайной величины  $X$  возможно выразим в виде таблицы:

$X$	$\frac{0}{n}$	$\frac{1}{n}$	$\frac{2}{n}$	...	$\frac{m}{n}$	...	...	$\frac{n}{n}$
$P\left(X = \frac{m}{n}\right)$	$1 \cdot q^n$	$C_n^1 \cdot p \cdot q^{n-1}$	$C_n^2 \cdot p^2 \cdot q^{n-2}$	...	$C_n^m \cdot p^m \cdot q^{n-m}$	...	...	$1 \cdot p^n$

Данный закон распределения называется *биномиальным законом*, основываясь на том, что вероятности равняются соответствующим членам разложения выражения:

$$(p + q)^n = \sum_{m=0}^n C_n^m \cdot p^m \cdot q^{n-m}.$$

Математическое ожидание случайной величины  $X$  называется *центром распределения вероятностей* случайной величины  $X$  и определяется нами по формуле:

$$x_C = \sum_{k=1}^n x_k \cdot P_k.$$

Центр масс и математическое ожидание вычисляем по аналогичным формулам.

Составляя план проведения независимых испытаний дискретной величины  $X$  с соответствующим законом распределения: выразим его в виде таблицы, которую так же используем как математическую модель, проводимых испытаний и их отображений в виде статистического исследования случайной величины:

$X$	$x_1$	$x_2$	$\dots$	$x_k$	$\dots$	$x_n$
$P(X = x_k)$	$p_1$	$p_2$	$\dots$	$p_k$	$\dots$	$p_n$

*Математическим ожиданием* дискретной случайной величины  $X$  (будем его обозначать  $M[x]$  или  $m_x$ ) называется сумма произведений всех возможных значений случайной величины на вероятности этих значений:

Математическое ожидание случайной величины, выраженное в виде ряда, где вероятность в каждом значении дискретной случайной величины  $X$  рассчитывается по формуле Бернулли:

$$P(X = x_k) = C_n^k \cdot p^k \cdot q^{n-k},$$

$$M[x] = x_1 \cdot p_1 + x_2 \cdot p_2 + \dots + x_n \cdot p_n = \sum_{k=1}^n x_k \cdot p_k \quad (1),$$

при этом основополагающим является  $\sum_{k=1}^n p_k = 1.$

Будем рассматривать только такие случайные величины, для которых этот ряд сходится. Устанавливая связь математического ожидания случайной величины и его среднего арифметического значения случайной величины при большом числе испытаний, наблюдается, что при неограниченном возрастании числа испытаний среднее значение стремится к ее математическому ожиданию.

Производя  $N$  независимых опытов. Предположим, что

- значение  $x_1$  появилось  $n_1$  раз,
- значение  $x_2$  появилось  $n_2$  раз,
- .....
- значение  $x_v$  появилось  $n_v$  раз.

Случайная величина  $m^*$  принимает значения  $x_1, x_2, \dots, x_v$ . Вычисляя среднее арифметическое полученных значений величины  $x$  (обозначая его через  $M[x]$  или  $m_x$ ), имеем

$$\bar{m}_x = \frac{x_1 n_1 + x_2 n_2 + \dots + x_v n_v}{N} = x_1 \frac{n_1}{N} + x_2 \frac{n_2}{N} + \dots + x_v \frac{n_v}{N} \quad (2).$$

Но так как при большом числе испытаний  $N$  относительная частота стремится к вероятности появления значения  $x_k$ , тогда

$$\sum_{k=1}^v x_k \frac{n_k}{N} \approx \sum_{k=1}^v x_k p_k.$$

При этих предположениях получается:  $\bar{M}[x] \xrightarrow{n \rightarrow \infty} M[x] \quad (3).$

Используя выходные исходные данные независимых случайных событий, будет определено  $n$  — число испытаний дающее математическое ожидание числа наступления событий

$$n = \frac{M[x]}{p}.$$

Данные формулы представляют собой основу математической модели статистической обработки материала, обрабатываемые на ПК.

Нами разработана система для расчета характеристик случайных величин посредством языка программирования PHP и языка баз, данных MySQL, с разбиением информационного материала на классы и объекты исполнения. Для создания алгоритма на PHP обработки статистических данных, применяется та же структура, что и в пособии, но здесь используется технология ООП в виде объектов и классов.

1. Приведем основные фрагменты программного кода для определения параметров характеристик вероятностной случайной величины. Код представлен на алгоритмическом языке программирования PHP, определенный в процедурном решении, избран для реализации язык программирования PHP, так как он прост и оптимален в программировании для обработки процессором, вследствие чего повышается скорость обработки массивов данных вероятностных величин.

```
//Рассчитаем величину мат. ожидания случайной величины:
```

```
For ($j=0; $j<=$int; $j++)
{
    $mj [$j] =$verbern [$j]*$zi [$j];
    echo "<br>";
    echo '$mj['.$j.']='.$mj[$j];
    $k1=$k1+$mj[$j];
}
echo'мат. ожидания случайной величины $mjob[$j]='.$k1;
echo'вычисление дисперсии случайной величины';
for ($j =0; $j <=$int;$j++ )
{
    //$dmj[$j]=$j*$j;
    $dm=pow($zi[$j],2);
    $dmj[$j]=$verbern[$j]*$dm;
    echo "<br>";
    echo'$verbern['.$j.']='.$verbern[$j];
    echo "<br>";
    echo'$dm='.$dm;
    echo "<br>";
    echo'$dmj['.$j.']='.$dmj[$j];
    $k2=$k2+$dmj[$j];
}
echo "<br>";
echo'кв.мат. ожидания случайной величины $disp2='.$k2;
echo "<br>";
$disp2=$k2;
$disp1=pow($k1,2);
echo "<br>";
echo'$disp1='.$disp1;
$disp=$disp2-$disp1;
```

```

        echo"<br>";
        echo'$disp='.$disp;
    $x0=$zi[$k]-$w;
    //Рассчитаем величину моды случайной величины:
    echo'$x0='.$x0;
    $h=$w;
    echo"<br>";
    echo'$h='.$h;
    $Moda=$x0+$h*($in1p1[$k]-$in1p1[$kpm])/((($in1p1[$k]-$in1p1[$kpm])+($in1p1[$k]-$in1p1[$kpp]));
    echo"<br>";
    echo'$Moda='.$Moda;
    //Рассчитаем величину медианы случайной величины:
    echo'$ΣS['.$kpm.']='.$ΣS[$kpm];
    $Σf=$t2;
    $Mediana=$x0+$h*($Σf/2-$ΣS[$kpm])/($in1p1[$k]);
    echo'$Dispersi='.$disp;
    echo"<br>";
        echo'$sigma='.$sigma;
        echo"<br>";
    echo'$Moda='.$Moda;
    echo"<br>";
    echo'$Mediana='.$Mediana;

```

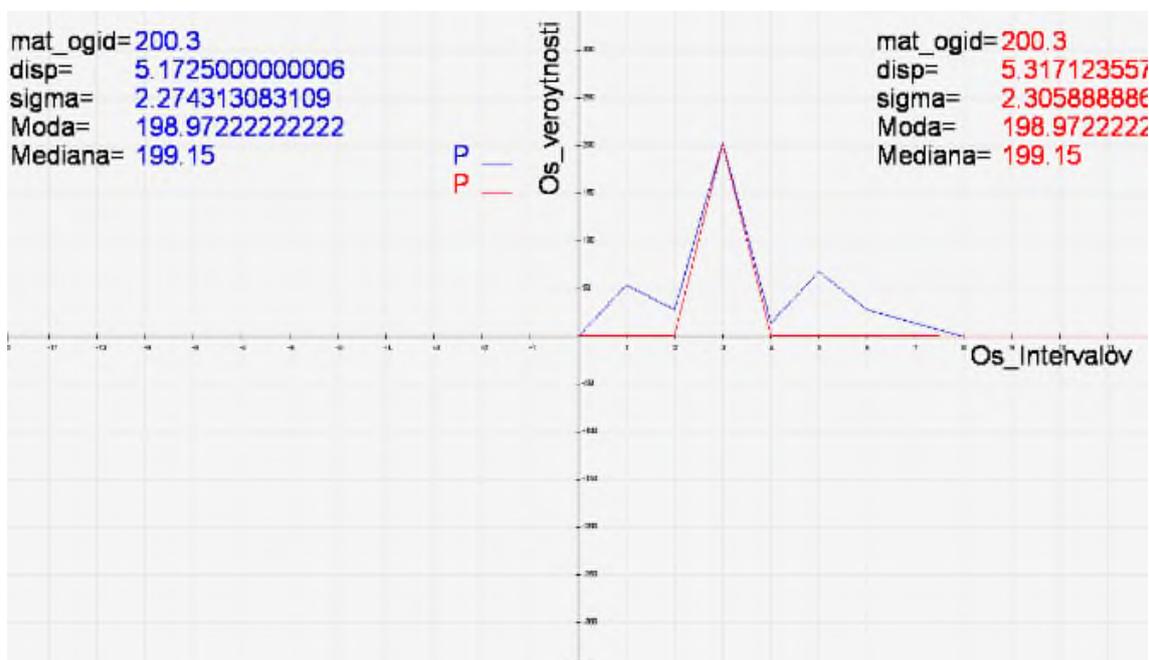
Покажем реализацию вышеприведенного программного кода на примере из [1], решение этой задачи приведено в [1] как вручную классическим методом и средствами MS Excel:

Стоимость книги по математической статистике в тридцати различных интернет-магазинах оказалась (в сомах): 200, 198, 201, 203, 203, 204, 196, 200, 203, 198, 199, 197, 197, 199, 199, 196, 199, 200, 201, 200, 200, 200, 203, 200, 200, 199, 204, 202, 205, 199. Построить таблицу частот, разбив данные на 6 интервалов, график выборочной функции распределения и гистограмму частот. Вычислить числовые характеристики выборки. Объем выборки – количество ее элементов  $n = 30$ . Результаты расчета приведем в сравнительной таблице:

ПАРАМЕТРЫ СЛУЧАЙНОЙ ВЕЛИЧИНЫ, ВЫЧИСЛЕННЫЕ	
классическим методом в MS Excel средствами «Анализа данных»	биномиального распределения в среде PHP и средствами языка баз данных MySQL
<i>Математическое ожидание случайной величины:</i>	
\$mjob[\$j]=200.3	\$mjob[\$j]=200.3
<i>Дисперсия случайной величины:</i>	
\$Dispersi=5.1725000000006	\$Dispersi=5.3171235571062
<i>Среднеквадратическое отклонение случайной величины:</i>	
\$sigma=2.274313083109	\$sigma=2.3058888865481
<i>Мода случайной величины:</i>	
\$Moda=198.97222222222	\$Moda=198.97222222222
<i>Медиана случайной величины:</i>	
\$Mediana***=199.15	\$Mediana***=199.15

Как видно из таблицы полученные результаты имеют незначительное расхождение в дисперсиях и средних квадратических отклонениях. Остальные результаты совпадают. Это означает, что программный код в среде PHP и средствами языка баз данных MySQL выполнен корректно.

Приведем результаты исследования математического прогнозирования случайной величины, выраженные в графическом виде



Несмотря на совпадение расчетов основных параметров описательной статистики в средах MS Excel и PHP, каждая из них имеют свои существенные преимущества, так и недостатки. А именно: MS Excel - мощное консольное приложение с богатым функционалом в виде статистических функций и механизма надстроек, оно является консольным, а не серверным, т.е. результаты обработки данных требуют предварительной подготовки для работы в приложении, например, с мастером функций, результаты нельзя надолго сохранить. В то же время для разработчиков в среде PHP и языка баз данных MySQL есть возможность использования удаленного сервера, а также наглядно интерпретировать полученные результаты выходного продукта. И здесь есть недостатки: сопровождение ПО и периодическая проверка сервера баз данных. Однако, несмотря на это, среда PHP и язык баз данных MySQL позволяет увеличить скорость обработки данных, что влечет оперативность принятия решения.

Таким образом, информационная Web-система на базе MySQL в среде PHP позволит оперативно спрогнозировать возможные решения задач широкого спектра: экологического (выбросов загрязнений окружающей среды), медицинского (распространения эпидемий и уменьшения их основных факторов), экономического (выпуска продукции и уменьшения брака) профиля и других аналогичных систем.

Список литературы:

1. Валеев С. Г. Практикум по прикладной статистике: учебное пособие / С. Г. Валеев, В. Н. Клячкин. – Ульяновск: УлГТУ, 2008. – 129 с.: ил.
2. Пискунов Н.С. Дифференциальное и интегральное исчисления. В 2-х т.
3. Стив Сьюринг PHP, MySQL, JavaScript & HTML5 для начинающих (*PHP, MySQL, JavaScript & HTML5 All-in-One For Dummies*).