МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ  
федеральное государственное АВТОНОМНОЕ образовательное учреждение высшего профессионального образования  
«Национальный исследовательский ядерный университет «МИФИ»  
  
  
**Димитровградский инженерно-технологический институт –**филиал федерального государственного автономного образовательного учреждения высшего  
профессионального образования «Национальный исследовательский ядерный университет «МИФИ»  
**(ДИТИ НИЯУ МИФИ)**

**Лабораторная работа №3  
по курсу «Основы математической статистики и   
планирование эксперимента»**

Составил: доцент кафедры   
 высшей математики   
 канд. экон. наук   
 Кожухова В.Н.

Димитровград 2017

**Лабораторная работа №3  
по курсу «Основы математической статистики и   
планирование эксперимента»**

**Критерии согласия**

**Задание**

Каждый вариант представляет собой три выборки по 100 значений каждая. Необходимо на уровне значимости 0,05 проверить первую выборку на соответствие биномиальному закону распределения (число испытаний *N* в каждом варианте известно), вторую – закону Пуассона, третью – нормальному закону распределения.

**Ход выполнения лабораторной работы**

Зная плотность распределения *f(х),* или функцию распределения *F(х)* случайной величины, можно вывести все свойства этой величины.

Поэтому при обработке статистического материала часто приходится решать вопрос о подборе для данной выборки аналитического выражения для *f(x)* или *F(x).* Речь идет о проверке «непараметрических статистических гипотез». К методам решения данной задачи относят критерии согласия Пирсона χ2, Колмогорова, ω2 и другие.

Нулевая гипотеза при применении общих критериев согласия записывается в форме

*Н0: Fn(x) = F(x)*,

где Fn(x)—эмпирическая функция распределения вероятностей; F(x)—гипотетическая функция распределения вероятностей.

Все известные общие критерии согласия можно разбить на три основные группы:

— критерии, основанные на изучении разницы между теоретической плотностью распределения и эмпирической гистограммой;

— критерии, основанные на расстоянии между теоретической и эмпирической функциями распределения вероятностей;

— корреляционно-регрессионные критерии, основанные на изучении корреляционных и регрессионных связей между эмпирическими и теоретическими порядковыми статистиками.

**1. Критерии согласия для биномиального распределения**

* 1. **Классический критерий χ2 для биномиального распределения**

Проверим первую выборку на соответствие биномиальному закону распределения. Объем выборки – число проведенных экспериментов. Каждый эксперимент состоит из N испытаний. В каждом испытании 2 исхода: успех и неудача. Каждое число в выборке – это число получившихся успехов в N испытаниях.

Необходимо составить ряд эмпирического распределения для данной выборки – число успехов принимает значения от 0 до N, – и рассчитать, сколько раз появляется в выборке каждое число успехов. Таким образом получим ряд распределения:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Число успехов xi | 0 | 1 | 2 | … | N |
| Эмпирическая частота ni | n0 | n1 | n2 | … | nN |

Проверим гипотезу о том, что выборка принадлежит биномиальному закону распределения:

*Н0: Fn(x) = FB(x),*

где *Fn(x)* – эмпирическая функция распределения по выборке; *FB(x)* – функция теоретического биномиального распределения.

Введем обозначения: N – число испытаний (дано каждому в исходных данных варианта); n – объем выборки (100 значений); p – параметр биномиального закона (вероятность успеха в одном испытании), вместо которого мы будем использовать оценку по методу моментов;  – теоретические вероятности числа успехов по схеме Бернулли.

Критерий хи-квадрат Пирсона имеет статистику:

,

где  – эмпирические частоты распределения,  – теоретические частоты,  – ероятность числа успехов *x*i из *N* испытаний, рассчитывается по схеме Бернулли:

.

Оценка ** параметра *p* биномиального распределения рассчитывается методом моментов:  – математическое ожидание биномиального закона; :

.

Критическая область – правосторонняя: если , то нулевая гипотеза отвергается.

**Число степеней свободы распределения хи-квадрат равно: (число групп выборки) минус (число определяемых параметров закона распределения) минус (1).**

Число групп выборки от 0 до N = N+1.

Число определяемых параметров закона распределения = 1 (определяем только параметр *p*).

Таким образом, число степеней свободы = (N+1) – (1) – (1) = N – 1.

Значит, критическое значение – это квантиль хи-квадрат распределения порядка   
(1 – *α*) с (*N* – 1) степенями свободы:.

**1.2 Критерий согласия Колмогорова-Смирнова для биномиального закона распределения**

Расчет проводят по негруппированным данным. Упорядочим выборку по возрастанию.

Статистики критерия Колмогорова-Смирнова:

; ; .

Здесь * –* значение интегральной функции биномиального распределения в точке с параметром *p*. Параметр *p* рассчитывается как в предыдущем пункте. В качестве  берутся значения исходной (негруппированной) выборки.

Вычисляем величину **. 3акон распределения определенной таким образом случайной величины при *n → ∞* не зависит от закона распределения случайной величины *Х,* этот закон табулирован.

Задаемся уровнем значимости α и определим критическоезначение*,* для которой вероятность того, что при справедливости гипотезы *Н0* наблюденное отклонение  функции эмпирического распределения от теоретического за счет случайных факторов примет значение большее, чем  равна α. Критическая область – правосторонняя. Если наблюденное значение ** окажется меньше , приведенного в табл. 1, то нет оснований отклонять .

Таблица 1

**Критические точки распределения Колмогорова**

|  |  |  |  |
| --- | --- | --- | --- |
| γ=1-α |  | γ=1-α |  |
| 0,01 | 0,44 | 0,60 | 0,89 |
| 0,05 | 0,52 | 0,70 | 0,97 |
| 0,10 | 0,57 | 0,80 | 1,07 |
| 0,15 | 0,61 | 0,90 | 1,22 |
| 0,20 | 0,65 | 0,95 | 1,36 |
| 0,30 | 0,71 | 0,98 | 1,52 |
| 0,40 | 0,77 | 0,99 | 1,63 |

**1.3 Критерий согласия Смирнова-Крамера-фон Мизеса**  **для биномиального распределения**

Расчет производится по **негруппированным данным**.

Статистика критерия *nω*2:

.

Здесь * –* значение интегральной функции биномиального распределения в точке с параметром *p*. Параметр *p* рассчитывается как в предыдущем пункте. В качестве  берутся значения исходной (негруппированной) выборки.

Таблица 2

Критические точки распределения *nω*2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| γ | 0,900 | 0,950 | 0,990 | 0,995 | 0,999 |
|  | 0,3473 | 0,4614 | 0,7435 | 0,8694 | 1,1679 |

Критическая область также правосторонняя.

**2. Критерии согласия для распределения Пуассона**

* 1. **Классический критерий χ2 для распределения Пуассона**

Проверим вторую выборку на соответствие закону распределения Пуассона. Объем выборки – число проведенных экспериментов.

Необходимо составить группированный ряд эмпирического распределения для данной выборки – рассчитать, сколько раз встречается каждое наблюдение в выборке, составив ряд от 0 до N c эмпирическими частотами. Здесь N – максимальное число в выборке. Таким образом получим ряд распределения:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| xi | 0 | 1 | 2 | … | N |
| Эмпирическая частота ni | n0 | n1 | n2 | … | nN |

Проверим гипотезу о том, что выборка принадлежит закону распределения Пуассона:

*Н0: Fn(x) = FP(x),*

где *Fn(x)* – эмпирическая функция распределения по выборке; *FP(x)* – функция теоретического распределения Пуассона.

Критерий хи-квадрат Пирсона имеет статистику:

,

где  – эмпирические частоты распределения,  – теоретические частоты,  – теоретические вероятности, рассчитываются по формуле Пуассона:

.

Оценка ** параметра распределения Пуассона рассчитывается методом максимального правдоподобия:

; .

Критическая область – правосторонняя: если , то нулевая гипотеза отвергается.

**Число степеней свободы распределения хи-квадрат равно: (число групп выборки) минус (число определяемых параметров закона распределения) минус (1).**

Число групп выборки от 0 до N = N+1.

Число определяемых параметров закона распределения = 1 (определяем только параметр оценкой **).

Таким образом, число степеней свободы = (N+1) – (1) – (1) = N – 1.

Значит, критическое значение – это квантиль хи-квадрат распределения порядка   
(1 – *α*) с (*N* – 1) степенями свободы:.

**2.2 Критерий согласия Колмогорова-Смирнова для закона распределения Пуассона**

Расчет производится по негруппированным данным.

Критерий Колмогорова-Смирнова для закона:

; ; .

Здесь * –* значение интегральной функции распределения Пуассона в точке с параметром ****. Здесь  – исходные значения в выборке.

Вычисляем величину **. Определяем критическоезначение*,* для которой вероятность того, что при справедливости гипотезы *Н0* наблюденное отклонение  функции эмпирического распределения от теоретического за счет случайных факторов примет значение большее, чем  равна α. Критическая область – правосторонняя. Если наблюденное значение ** окажется меньше , приведенного в табл. 1, то нет оснований отклонять .

**2.3 Критерий согласия Смирнова-Крамера-фон Мизеса**  **для распределения Пуассона**

Расчет производится по **негруппированным данным**.

Статистика критерия *nω*2:

.

Здесь * –* значение интегральной функции распределения Пуассона в точке с параметром ****. Критические точки распределения *nω*2 представлены в таблице 2.

Критическая область также правосторонняя.

**3. Критерии согласия для нормального распределения**

**3.1 Предварительное исследование на нормальность**

1. **Графический критерий**

В графическом способе результаты испытаний располагают в вариационном ряду. Затем для каждого результата xi рассчитывают накопленную частость:

*W i =i/(n+1)*

Здесь *i*– номер результата в вариационном ряду, *n*– объём испытаний. По значениям накопленных частостей находят соответствующие значения квантиля стандартного нормального распределения  *z(Wi)*. Далее строят график в координатах *xi* по оси абсцисс*, z* – по оси ординат. Если значения *xi* являются квантилями нормального распределения, как и *z(W i)*, между ними должна быть линейная зависимость. Тогда, если точки на графике укладываются вдоль прямой линии лишь с небольшими отклонениями, то результаты испытаний достаточно хорошо описываются нормальным распределением. При больших отклонениях хотя бы некоторых точек от прямой распределение результатов испытаний нельзя считать соответствующим нормальному.

1. **Оценка вида распределения по асимметрии и эксцессу**

Асимметрия показывает меру несимметричности кривой плотности реального распределения относительно нормального распределения. Эксцесс показывает меру вытянутости кривой плотности реального распределения относительно нормального распределения.

По асимметрии и эксцессу можно провести приближенную проверку нормальности распределения результатов испытаний. Асимметрию (А) и эксцесс (Е) рассчитывают так:



В Excel вместо этих формул можно использовать статистические функции СКОС (для расчета А) и ЭКСЦЕСС (для расчёта Е).

Дисперсии асимметрии и эксцесса рассчитывают так:

.

Если , то результаты испытаний считают распределёнными нормально.

**3.2 Классический критерий χ2 и модифицированный для проверки гипотезы нормальности распределения критерий χ2**

1) Традиционное применение критерия Пирсона включает в себя построение гистограммы. Вычисляется размах |xn - x1| вариационного ряда и образуем «r» равных интервалов шириной



Число интервалов «*r*» выбираем в зависимости от объема выборки.

Таблица 3

**Рекомендуемые числа интервалов в зависимости от числа результатов измерений согласно ГОСТ Р 8.736-2011 ГСИ**

|  |  |
| --- | --- |
| Число результатов измерений | Рекомендуемое число интервалов |
| 40-100 | 7-9 |
| 100-500 | 8-12 |
| 500-1000 | 10-16 |
| 1000-10000 | 12-22 |

Можно рекомендовать и формулу Стерджеса для определения минимального числа групп:

*r* = 1 + 3,322lg *n.*

Результаты наблюдений *xi* группируем по интервалам, подсчитываем частоты *mj* величин *xi ,* попавших в *j* - тыеинтервалы. Если внутрь *j* - того интервала попало *т′j* точек, а внутрь (*j* *+* 1) - того - *m'j+1* точек, причем на границу этих интервалов попало *ν* точек выборки, то рекомендуется полагать следующее:



Таким образом, из точки на границе интервалов в смеж­ные интервалы относят по «1/2 точки».

Через *x'j-1 , xj'* будем обозначать границы *j -* тогоинтервала (j*=*1, 2,…, *r*).

Строим на первом чертеже гистограмму относительных частот, принимая высоту прямоугольников равной  .

При этом площадь *j* - того прямоугольника равна – относительной частоте наблюдений, попавших в *j* - тый интервал.

Площадь гистограммы относительных частот равна сумме всех относительных частот, т. е. единице. Гистограмма выступает в роли эмпирической аппроксимации генерального распределения.

Визуальный анализ вида гистограммы помогает выдвинуть гипотезу *Но* о виде кривой распределения *f(x)*, которой может быть описано эмпирическое распределение.

При этом гипотеза *Н*о в общем случае не дает численной информации о параметрах гипотетического распределения.

В качестве числовых характеристик распределения можно использовать значения, рассчитанные по методу «группированных данных».

В «методе группированных данных» выборка объемом «*n*»заменяется группированной выборкой объемом «*r*» из равноотстоящих вариант , где *r* – количество интервалов, а - середина *j* - того интервала, называемая «представителем j– того интервала».

Найдем и S для полученной выборки по формулам:

,

.

Т.е. расчет оценок параметров производится по **группированным данным**.

Добавим, что методом «группированных данных» часто пользуются для нахождения оценок *, S* при достаточно большой выборке *(п >50).*

Используя группированную выборку (объемом *r* значений *),* строим на втором чертеже эмпирическую функцию распределения *Fэ().* Значения *Fэ()* в точках, соответствующих середине *j*-того интервала, вычисляем по формуле



т. е. рассматриваем их как сумму относительных частот  попадания *Х* в интервалы, предшествующие .

Получаем ступенчатую функцию, скачки которой происходят в точках .

Переходим к нормированным и центрированным случайным величинам:



для которых табулируются значения функции распределения и плотности вероятностей.

Для выдвинутой гипотезы о плотности вероятностей (функции распределения) определяем теоретические вероятности попадания опытных данных в *j*-тый интервал. Для этого вычисляем значения *npj.*

Вычисляем меру расхождения эмпирического (статистического) и предложенного гипотетического распределений



Закон распределения, определенной таким образом случайной величины χ*2* при

*n → ∞* не зависит от параметров закона распределения случайной величины *Х* (именно такими мерами расхождения и пользуются в математической статистике), а определяется только числом степеней свободы *k* и в предположении о справедливости гипотезы *Но* соответствует χ2 – распределению, зависящему от числа степеней свободы «k».

При применении критерия согласия Пирсона важен правильный подсчет числа степеней свободы *k.* Если параметры гипотетического закона распределения известны, то *k* = *r*–1*.* В этом случае число наложенных связей *s\* =* 1*,* т.к. сумма наблюденных относительных частот равна 1. Эта наложенная связь существует при применении критерия согласия Пирсона всегда.

Чаще *l* параметров распределения устанавливают по выборочным значениям.

Поэтому из общего числа разрядов следует дополнительно вычесть число наложенных таким образом связей, т.е. *s\* = l* + 1,а *k = r – l -* l*.*

Для проверки гипотезы зададимся уровнем значимости *α,* которая равна вероятности того, что при справедливости нулевой гипотезы *Н0* о теоретической плотности вероятности *f(х)* наблюденная мера расхождения *χ2* за счет случайных факторов примет значение больше критического значения *χ2КР(α,k)* при данных *α* и *k:*



Имеем правостороннюю критическую область. Если наблюденное *χ2* < *χ2*кр(α, *k),* то нет основания отвергать нулевую гипотезу. Если наблюденное *χ2*  *χ2*кр(α, *k),* то нулевая гипотеза отвергается и следует выдвинуть гипотезу о другом виде теоретического распределения.

Чтобы графически проверить гипотезу отеоретическом распределении, нанесем на первый чертеж с гистограммой точки (; *p*j/*h*),а на второй чертеж с эмпирической функцией распределения – точки ; F()и соединим их плавными линиями.

Для того, чтобы подобрать плотность нормального распределения

****

следует принять  *mх =* *;*  *= S;* – полученные по методу максимального правдоподобия и *s*\* = 3 – число наложенных связей, при этом вероятность попадания в *i*-тый интервал



где - значение функции Лапласа для нормированной величины *xjн=**.*

2) Рассмотрим модифицированный критерий согласия хи-квадрат, улучшенный специально для проверки нормальности.

Параметры распределения **оцениваются по негруппированной выборке в отличие от классического критерия χ2**.

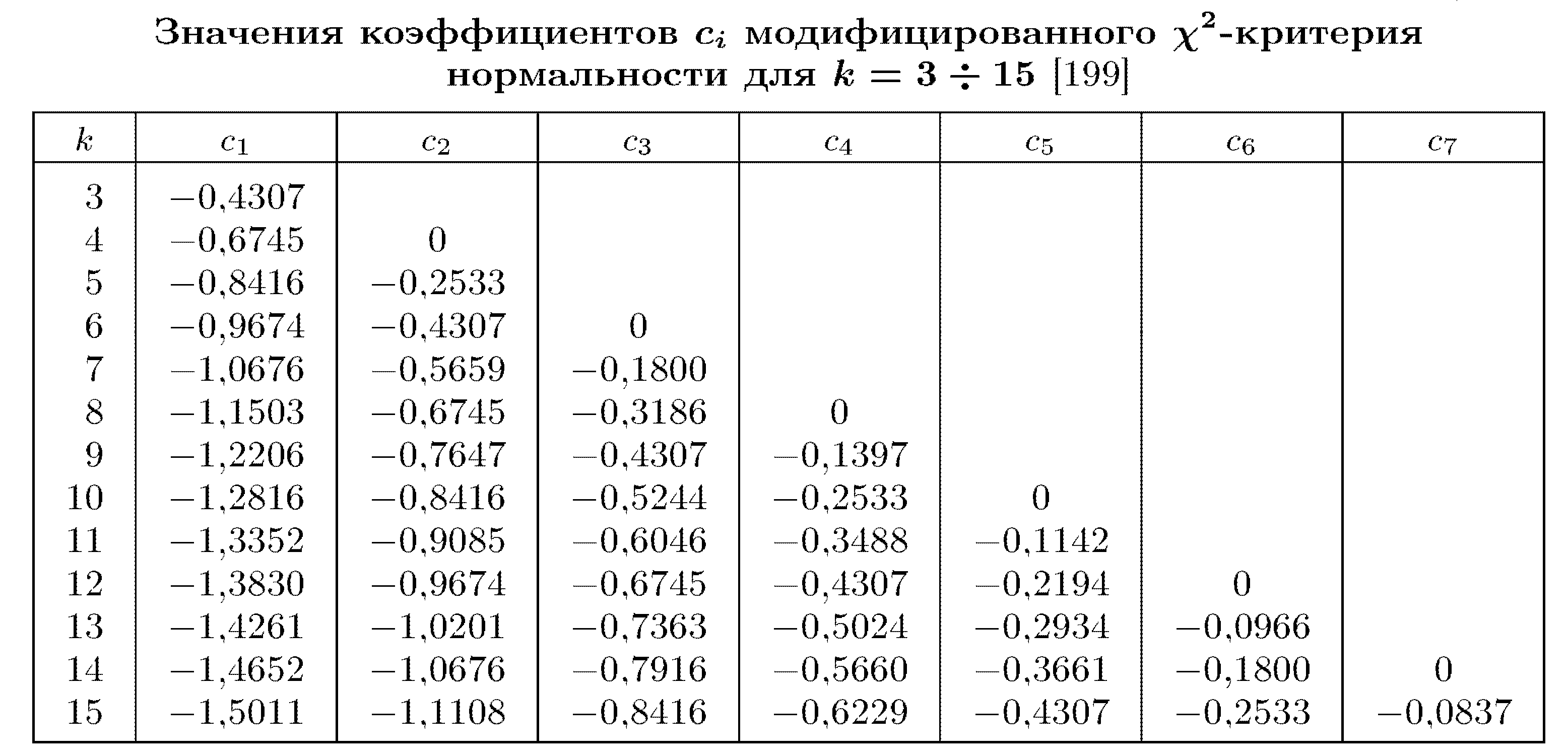
Далее выборка разбивается на ***k*** равновероятных интервалов – так, чтобы вероятность попадания в каждый из них была одинакова (*pi* = 1/***k*** = const). Статистика критерия рассчитывается по формуле

,

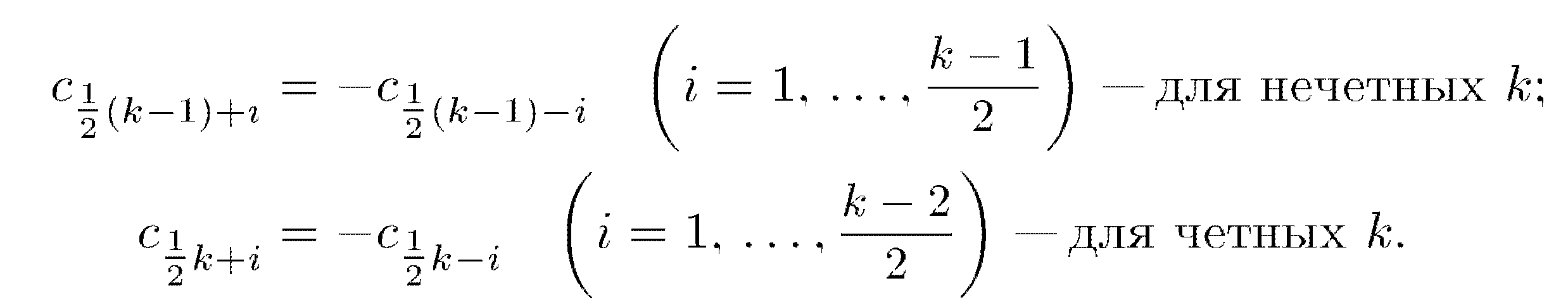
где *n* — объем выборки; *mi* — количество членов выборки, попавшее в *i*-й интервал.

Границы интервалов определяются как . Значения коэффициентов *c* приведены в табл. 4.

Таблица 4

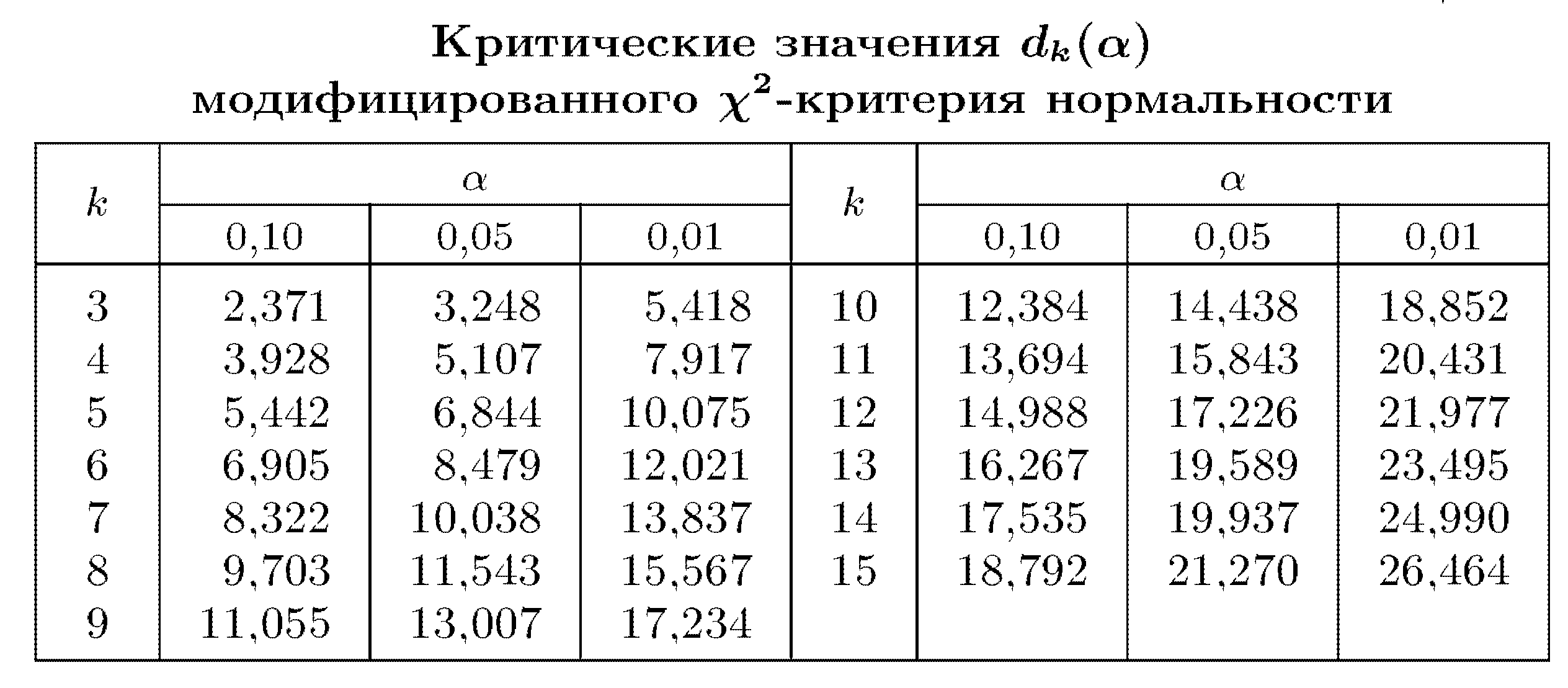


Следует учесть, что *c*0 = –∞ и *c*k = ∞. Так как *с*i симметричны относительно нуля, то недостающие значения можно найти из соотношений



Если , где  – критическое значение статистики критерия на уровне значимости *α*, то гипотеза нормальности отклоняется. Критические значения приведены в табл. 5.

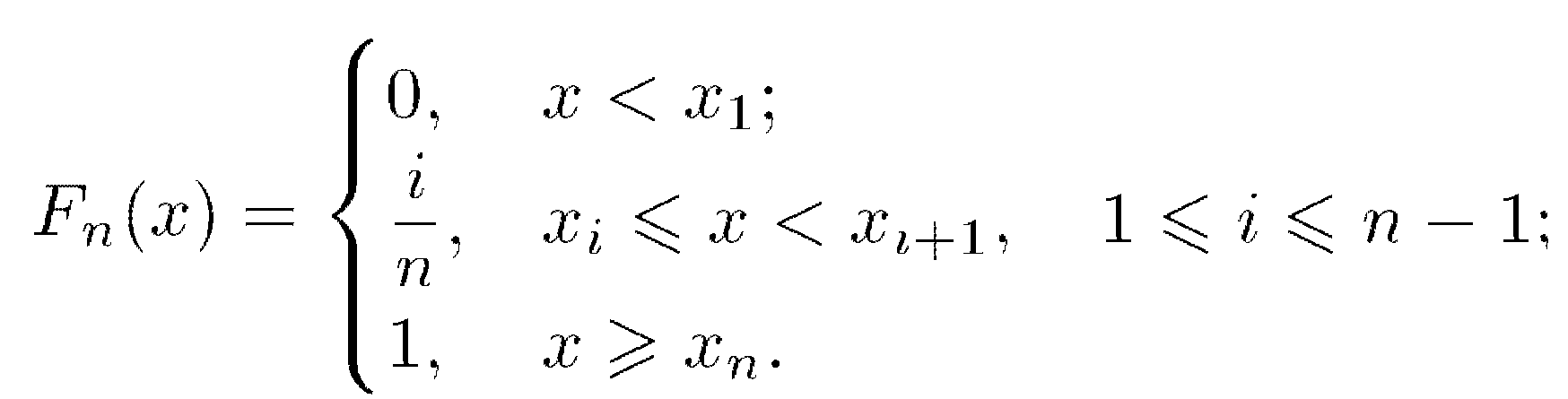
Таблица 5



**3.3 Критерий согласия Колмогорова-Смирнова для нормального распределения**

Расчет производится по **негруппированным данным**.

В этом случае значения эмпирической функции распределения Fn(x) вычисляются как



Для проверки нулевой гипотезы *H0*: *Fn(x) = Ф(x)*, где *Ф(x)* – полностью определенная (с точностью до параметров) теоретическая функция распределения, рассматривается расстояние между эмпирической и теоретической функциями распределения:

; ; .

где  – значение интегральной функции стандартного нормального распределения, рассчитанной в точке . Оценки параметров нормального распределения берутся методом максимального правдоподобия как выборочное среднее и выборочное СКО по негруппированным данным соответственно.

Модифицированная для проверки нормальности статистика критерия имеет вид:

.

Таблица 6

Критические точки распределения Колмогорова-Смирнова

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| α | 0,15 | 0,10 | 0,05 | 0,03 | 0,01 |
|  | 0,775 | 0,819 | 0,895 | 0,955 | 1,035 |

Критическая область – правосторонняя.

**3.4 Критерий согласия Смирнова-Крамера-фон Мизеса  для нормального распределения**

Расчет производится по **негруппированным данным**.

Нулевая гипотеза имеет тот же вид, что и в предыдущем критерии. Статистика критерия *nω*2:

.

Таблица 7

Критические точки распределения *nω*2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| γ | 0,900 | 0,950 | 0,990 | 0,995 | 0,999 |
|  | 0,1035 | 0,1260 | 0,1788 | 0,2018 | 0,2559 |

Критическая область также правосторонняя.

**3.5 Критерий Шапиро-Уилка для проверки гипотезы о нормальности распределения**

Расчет производится по **негруппированным данным**.

Критерий Шапиро-Уилка основан на отношении оптимальной линейной несмещенной оценки дисперсии к ее обычной оценке методом максимального правдоподобия. Изучение мощности критерия Шапиро-Уилка показало, что это – **один из наиболее эффективных критериев проверки нормальности распределения случайных величин.**

Статистика критерия имеет вид:

, где .

Коэффициенты  берутся из справочных таблиц. Критические значения статистики W(α) также находятся таблично. Если W < W(α), то нулевая гипотеза о нормальности распределения отклоняется при уровне значимости α.

Однако, эти таблицы довольно громоздки и неудобны в применении. Потому была выведена полезная аппроксимация, позволяющая применить критерий Шапиро-Уилка без помощи таблиц. Для α = 0.05 предлагается статистика

,

где



Если *W1* < 1, то нулевая гипотеза нормальности распределения случайных величин отклоняется. [ ] – это целая часть числа.

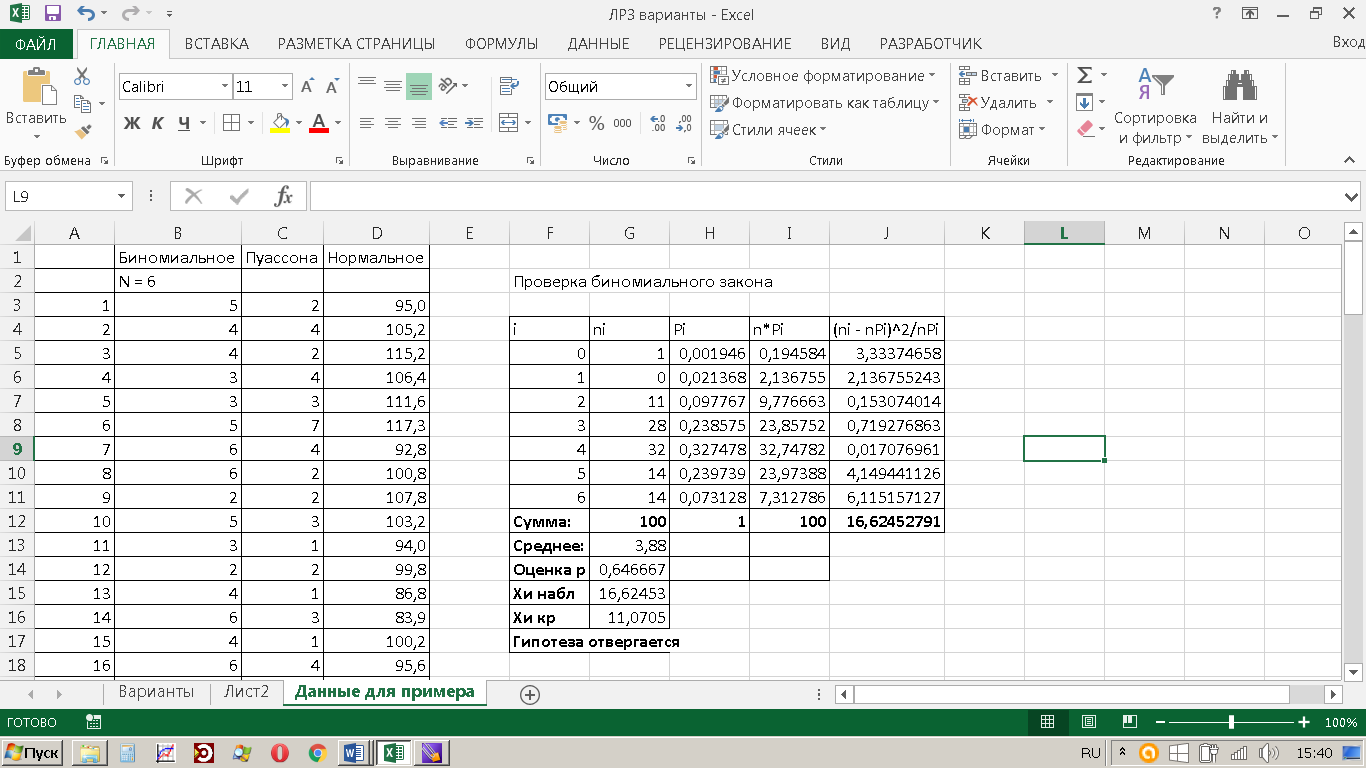
Приведем ранжирование указанных выше критериев нормальности от лучшего к худшему по совокупной мощности критериев:

1. Критерий Шапиро-Уилка
2. Критерий χ2
3. Критерий Колмогорова-Смирнова
4. Критерий Смирнова-Крамера-фон Мизеса

**Пример выполнения лабораторной работы**

Этот пример создан в помощь студенту. Он не означает, что выполнять отчет по лабораторной работе нужно так же, как здесь.

Даны три выборки по 100 значений. Необходимо проверить первую выборку на соответствие биномиальному закону, вторую – закону Пуассона, третью – нормальному закону.



**1. Критерии согласия для биномиального закона**

**1.1 Критерий хи-квадрат**

Расчет в Excel можно выполнить как на рисунке 1:

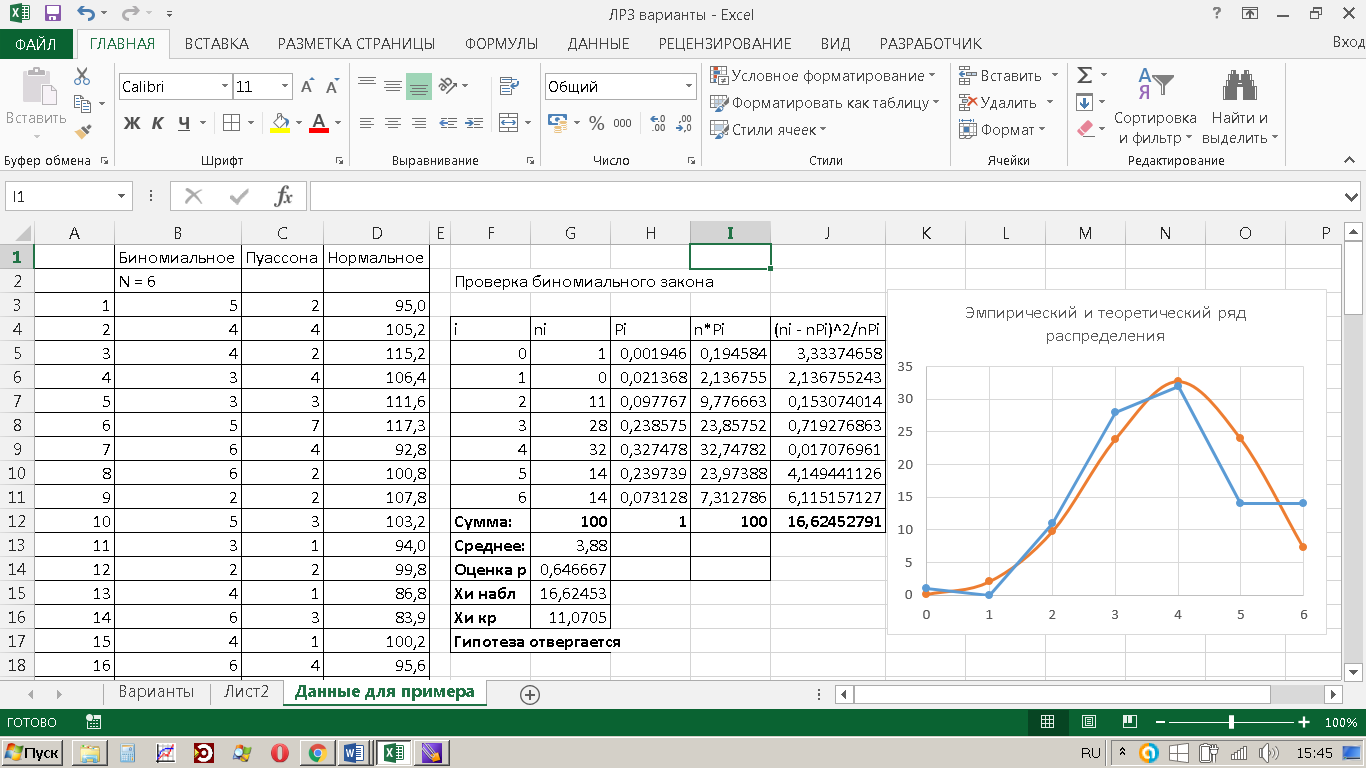


Рисунок 1 – Расчет критерия хи-квадрат для биномиального распределения

Сгруппируем выборку по числу успехов, рассчитав частоты с помощью функции СЧЕТЕСЛИ:

|  |  |
| --- | --- |
| i | ni |
| 0 | 1 |
| 1 | 0 |
| 2 | 11 |
| 3 | 28 |
| 4 | 32 |
| 5 | 14 |
| 6 | 14 |
| **Сумма:** | **100** |

Рассчитаем среднее значение:  и оценку параметра *p*: 

Рассчитаем , , :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| i | ni | Pi | n\*Pi | (ni - nPi)^2/nPi |
| 0 | 1 | 0,001946 | 0,194584 | 3,33374658 |
| 1 | 0 | 0,021368 | 2,136755 | 2,136755243 |
| 2 | 11 | 0,097767 | 9,776663 | 0,153074014 |
| 3 | 28 | 0,238575 | 23,85752 | 0,719276863 |
| 4 | 32 | 0,327478 | 32,74782 | 0,017076961 |
| 5 | 14 | 0,239739 | 23,97388 | 4,149441126 |
| 6 | 14 | 0,073128 | 7,312786 | 6,115157127 |
| **Сумма:** | **100** | **1** | **100** | **16,62452791** |

Критерий хи-квадрат Пирсона имеет статистику:

,



Критическая область – правосторонняя: если  (наш случай), то нулевая гипотеза отвергается.

Также построим график эмпирических и теоретических частот (рисунок 2).

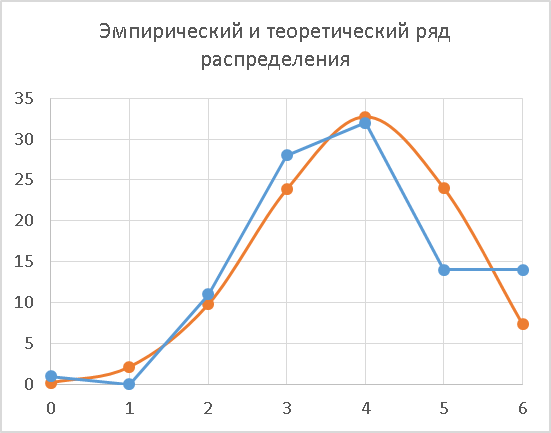


Рисунок 2 – График эмпирических и теоретических частот

Решение в формулах выглядит следующим образом (рисунок 3):

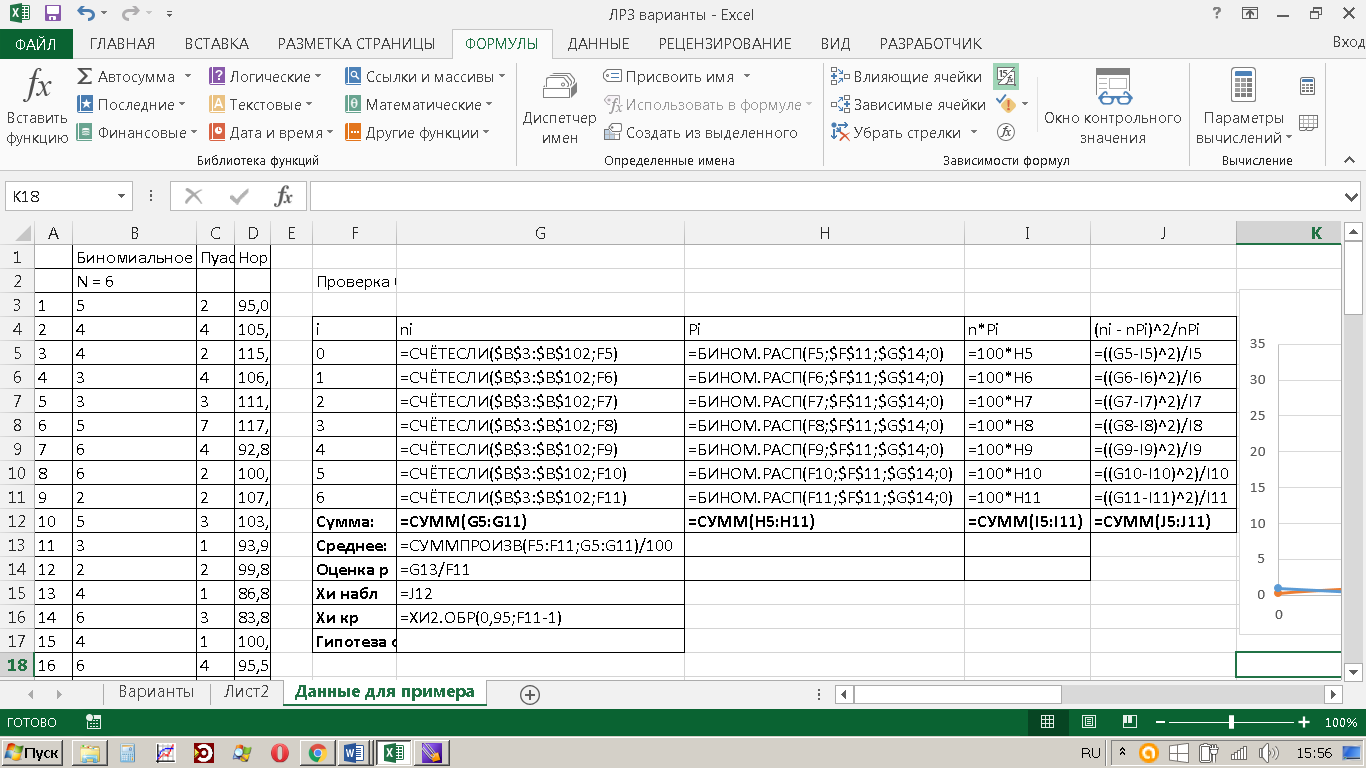


Рисунок 3 – Расчет критерия хи-квадрат в режиме отображения формул

**1.2 Критерий Колмогорова-Смирнова**

В силу простоты критерия приведем результат (рисунок 4):

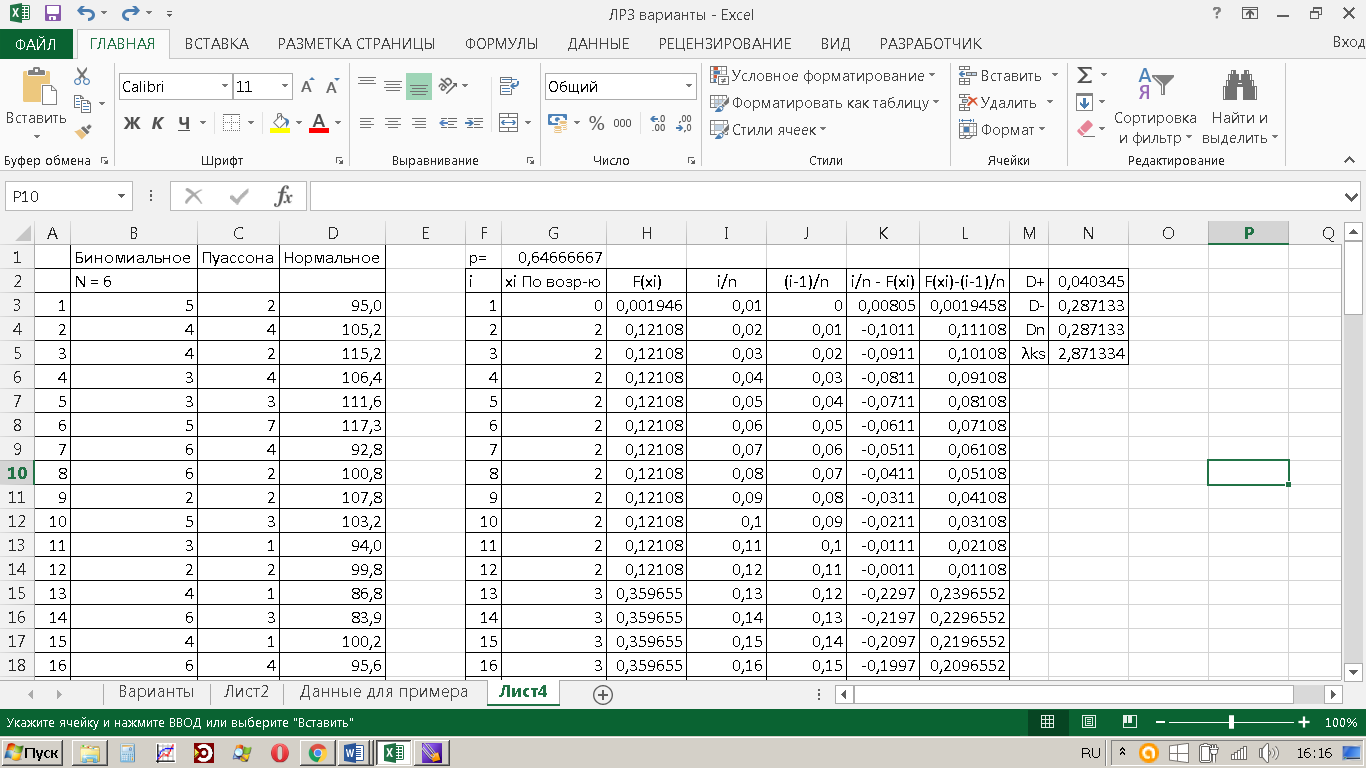


Рисунок 4 – Расчет критерия Колмогорова-Смирнова

В формулах решение приведено на рисунке 5:

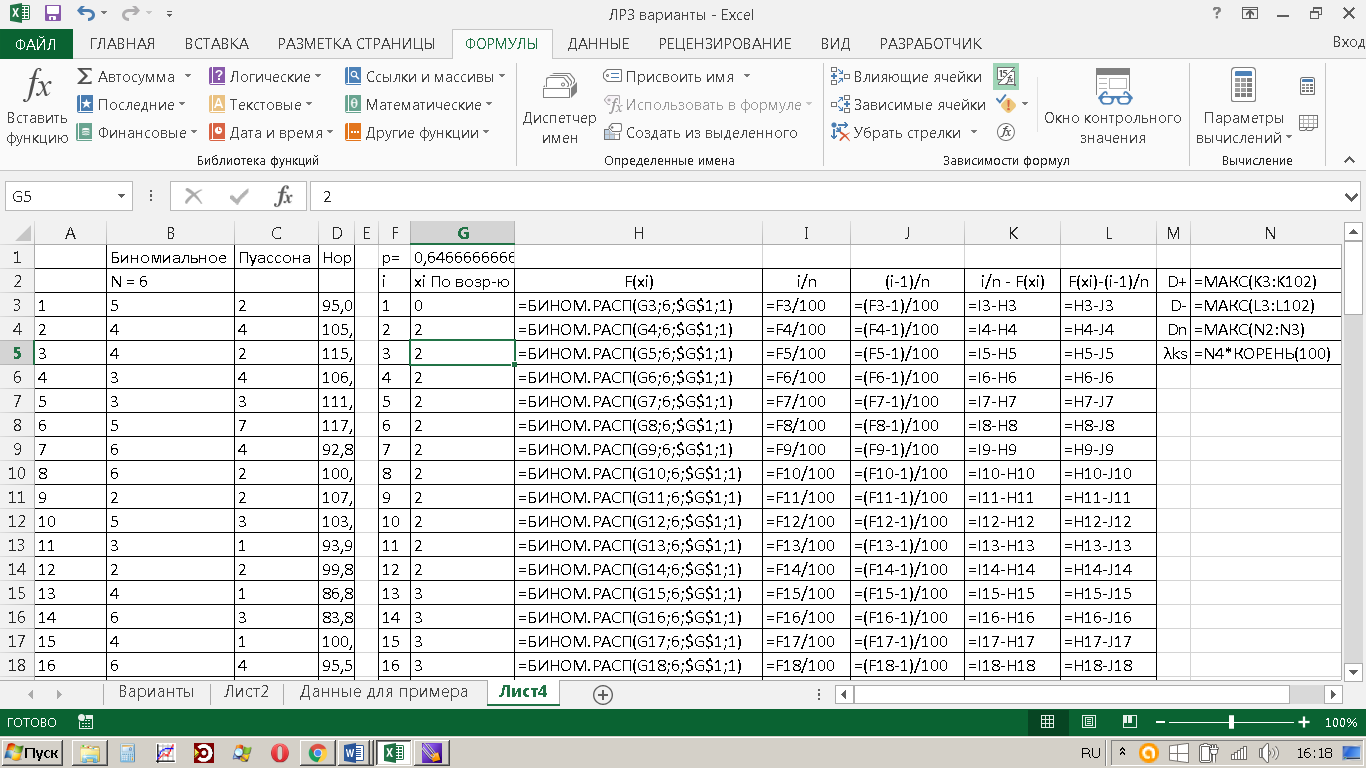


Рисунок 5 – Расчет критерия Колмогорова-Смирнова в режиме отображения формул

Наблюдаемое значение: 2,87

Критическое значение: 1,36

Так как наблюдаемое больше критического, то гипотеза о биномиальном распределении отвергается.

Также по рисунку 6 видно, что эмпирическая функция не попала между двумя прямыми, как должно было быть, если бы гипотеза была принята:

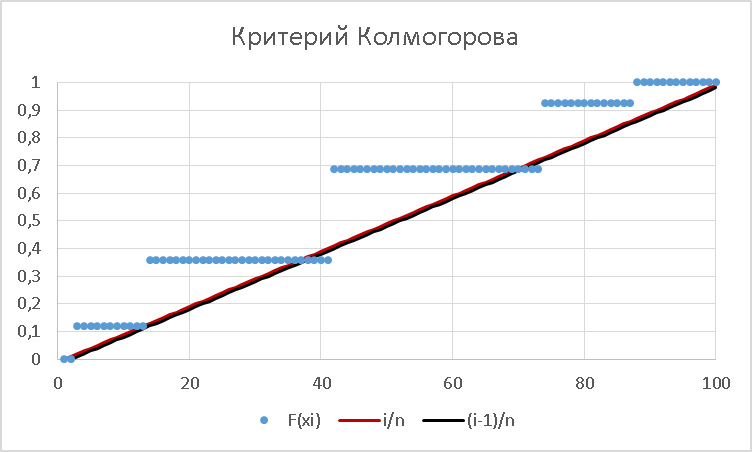


Рисунок 6 – Теоретические границы распределения и эмпирическая функция распределения при расчете критерия Колмогорова-Смирнова

**1.3 Критерий Смирнова-Крамера-фон Мизеса**

Расчет критерия представлен на рисунках 7, 8.

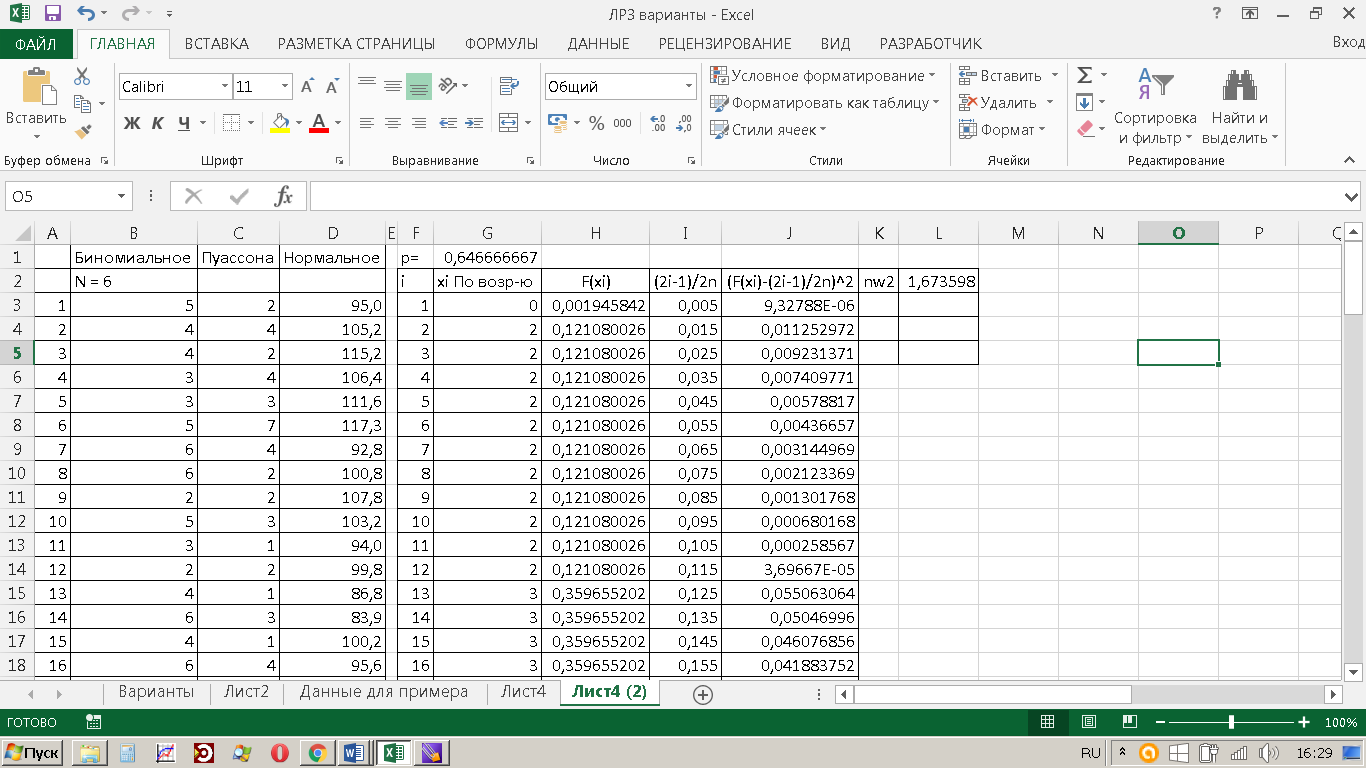


Рисунок 7 – Расчет критерия Смирнова-Крамера-фон Мизеса

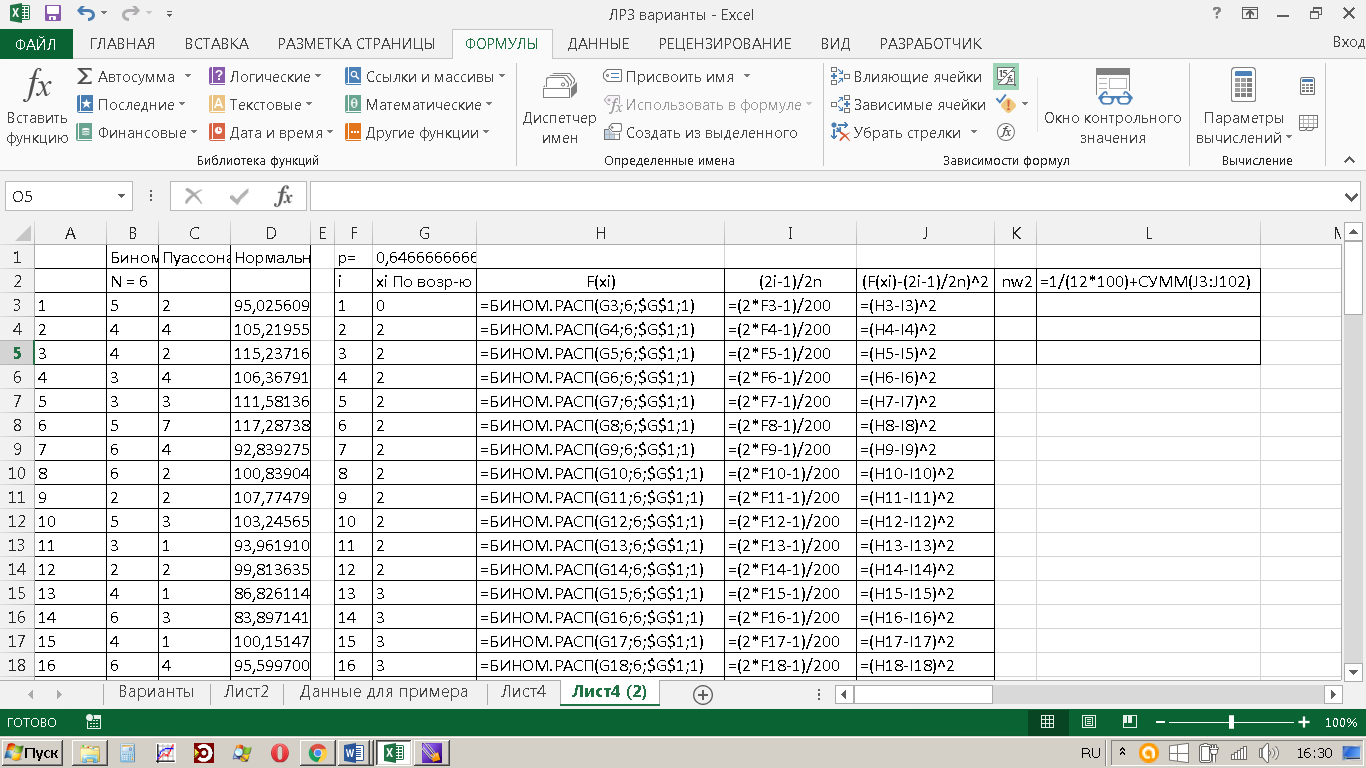


Рисунок 8 – Расчет критерия Смирнова-Крамера-фон Мизеса в режиме отображения формул

Наблюдаемое значение: 1,67

Критическое значение: 0,46

Так как наблюдаемое больше критического, то нулевая гипотеза отвергается. Также по рисунку 9 видно, что распределение прошло выше теоретического и плохо им описывается:

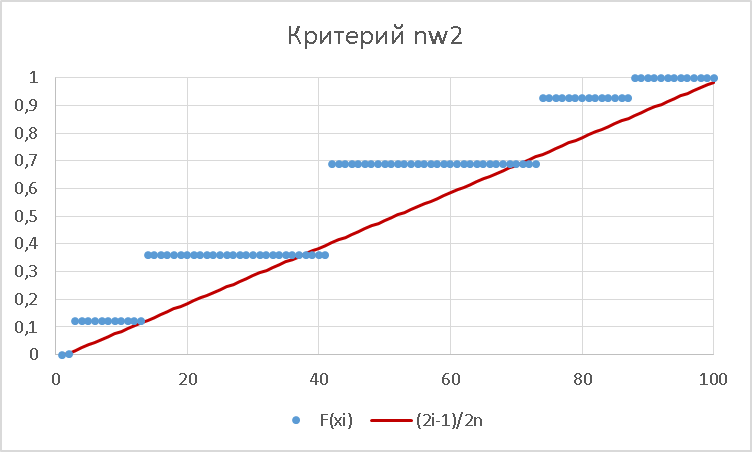


Рисунок 8 – Эмпирическая и теоретическая функции распределения при расчете по критерию Смирнова-Крамера-фон Мизеса

**2. Критерии согласия для закона Пуассона**

* 1. **Классический критерий χ2 для распределения Пуассона**

Проверим вторую выборку на соответствие закону распределения Пуассона. Объем выборки – число проведенных экспериментов = 100.

Необходимо составить группированный ряд эмпирического распределения для данной выборки – рассчитать, сколько раз встречается каждое наблюдение в выборке, составив ряд от 0 до N c эмпирическими частотами. Здесь N – максимальное число в выборке. В нашем примере N = 9. Таким образом получим ряд распределения:

|  |  |
| --- | --- |
| i | ni |
| 0 | 6 |
| 1 | 11 |
| 2 | 21 |
| 3 | 20 |
| 4 | 22 |
| 5 | 7 |
| 6 | 8 |
| 7 | 3 |
| 8 | 1 |
| 9 | 1 |
| **Сумма:** | **100** |

Проверим гипотезу о том, что выборка принадлежит закону распределения Пуассона. Оценка ** параметра распределения Пуассона рассчитывается методом максимального правдоподобия:

; .

Критерий хи-квадрат Пирсона имеет статистику:

,

где  – эмпирические частоты распределения,  – теоретические частоты,  – теоретические вероятности, рассчитываются по формуле Пуассона:

.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| i | ni | Pi | n\*Pi | (ni - nPi)^2/nPi |
| 0 | 6 | 0,039955058 | 3,995505826 | 1,005629091 |
| 1 | 11 | 0,128655288 | 12,86552876 | 0,270505598 |
| 2 | 21 | 0,207135013 | 20,7135013 | 0,003962705 |
| 3 | 20 | 0,222324914 | 22,2324914 | 0,224177208 |
| 4 | 22 | 0,178971556 | 17,89715558 | 0,94055909 |
| 5 | 7 | 0,115257682 | 11,52576819 | 1,777111719 |
| 6 | 8 | 0,061854956 | 6,185495596 | 0,53228172 |
| 7 | 3 | 0,02845328 | 2,845327974 | 0,008407971 |
| 8 | 1 | 0,011452445 | 1,14524451 | 0,018420492 |
| 9 | 1 | 0,00409743 | 0,409743036 | 0,850297024 |
| **Сумма:** | **100** | **0,998157622** | **99,81576217** | **5,631352619** |
| **Среднее:** | 3,22 |  |  |  |
| **Оценка ** | 3,22 |  |  |  |
| **Хи набл** | 5,631352619 |  |  |  |
| **Хи кр** | 15,50731306 |  |  |  |

Критическая область – правосторонняя: если , то нулевая гипотеза отвергается. В нашем случае гипотеза принимается с вероятностью 0,95. Расчет критерия представлен рисунками 10, 11, 12.

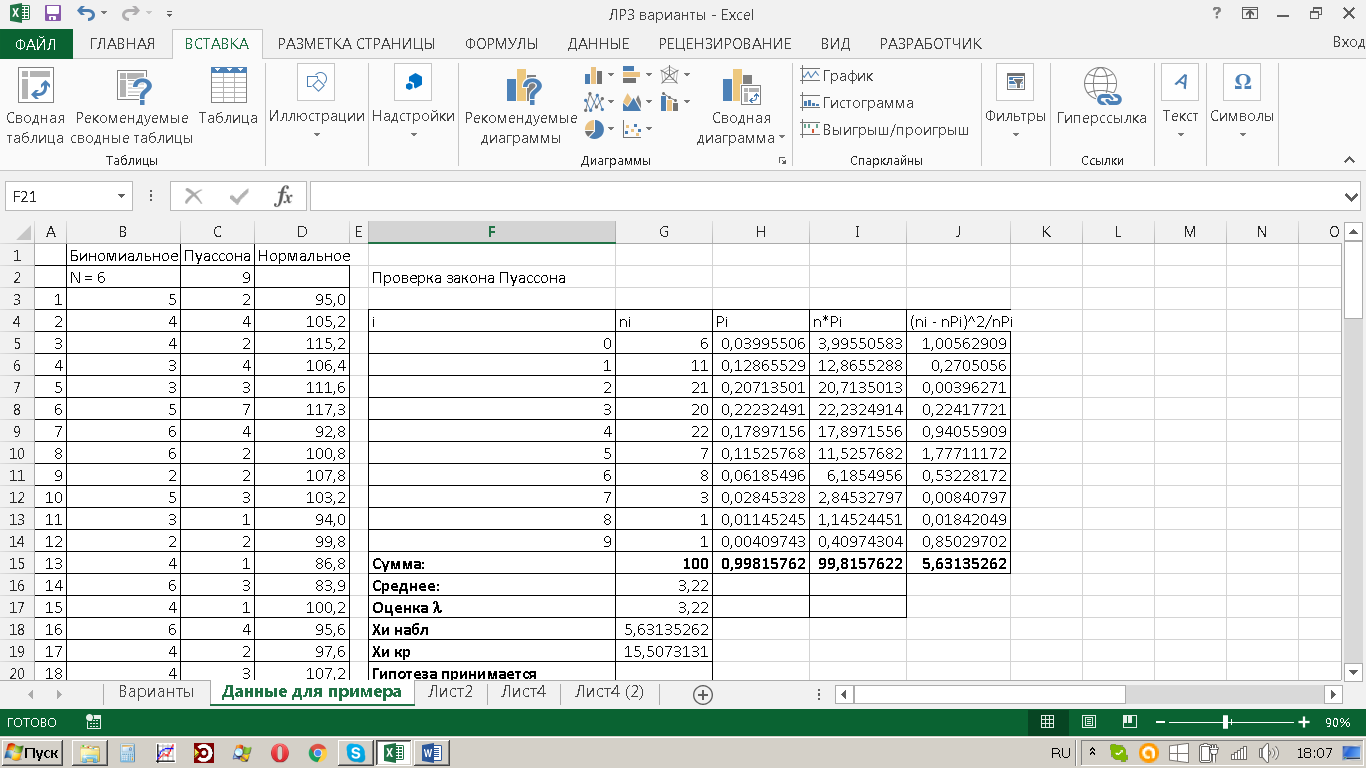


Рисунок 10 – Расчет критерия хи-квадрат

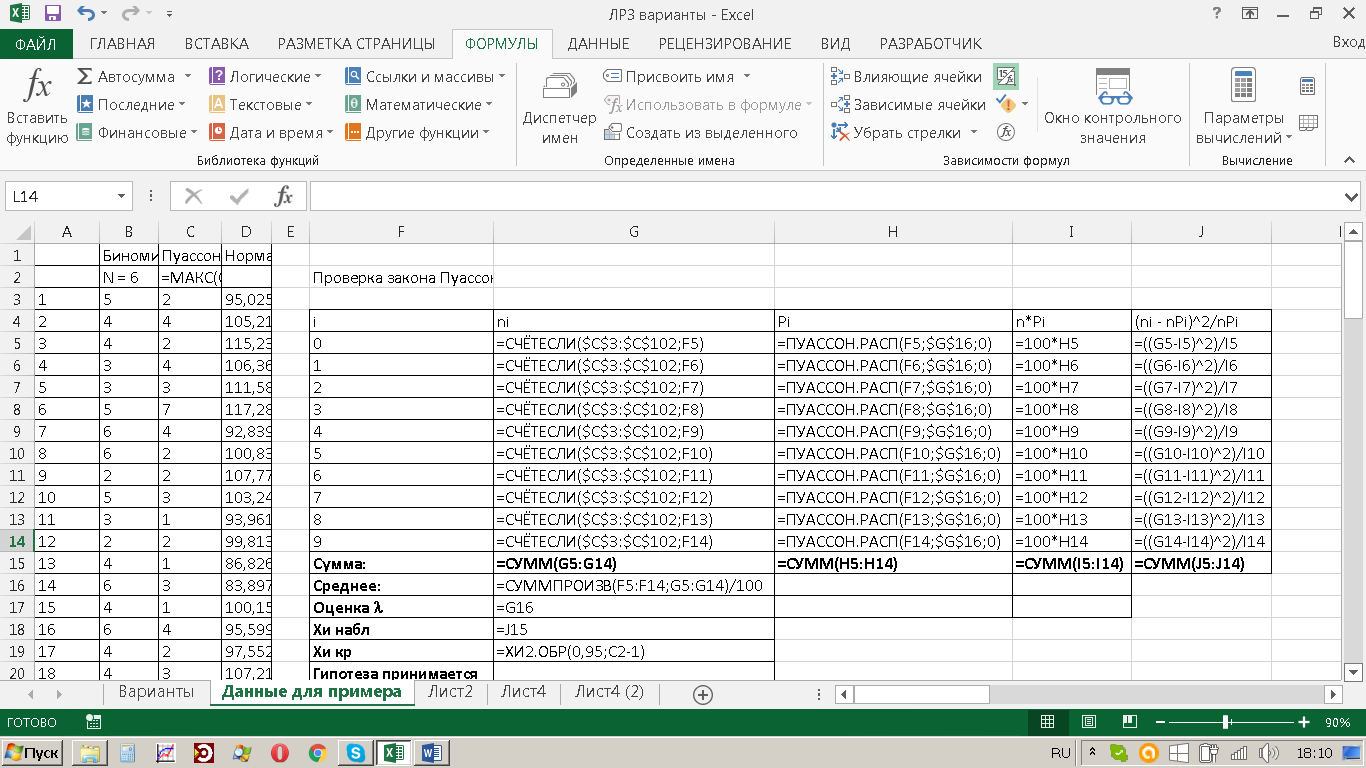


Рисунок 11 – Расчет критерия хи-квадрат в режиме отображения формул

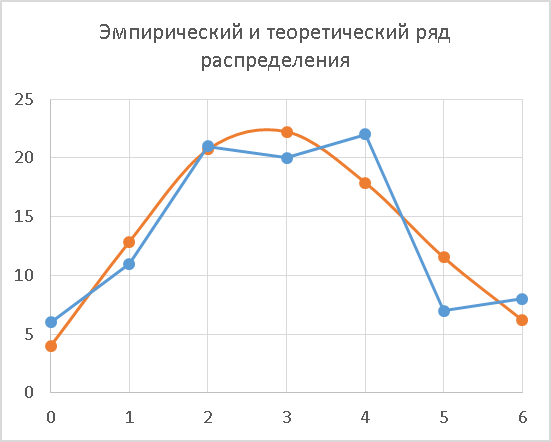


Рисунок 12 – График эмпирических и теоретических частот

**2.2 Критерий согласия Колмогорова-Смирнова для закона распределения Пуассона**

В силу простоты критерия приведем результат (рисунок 13):

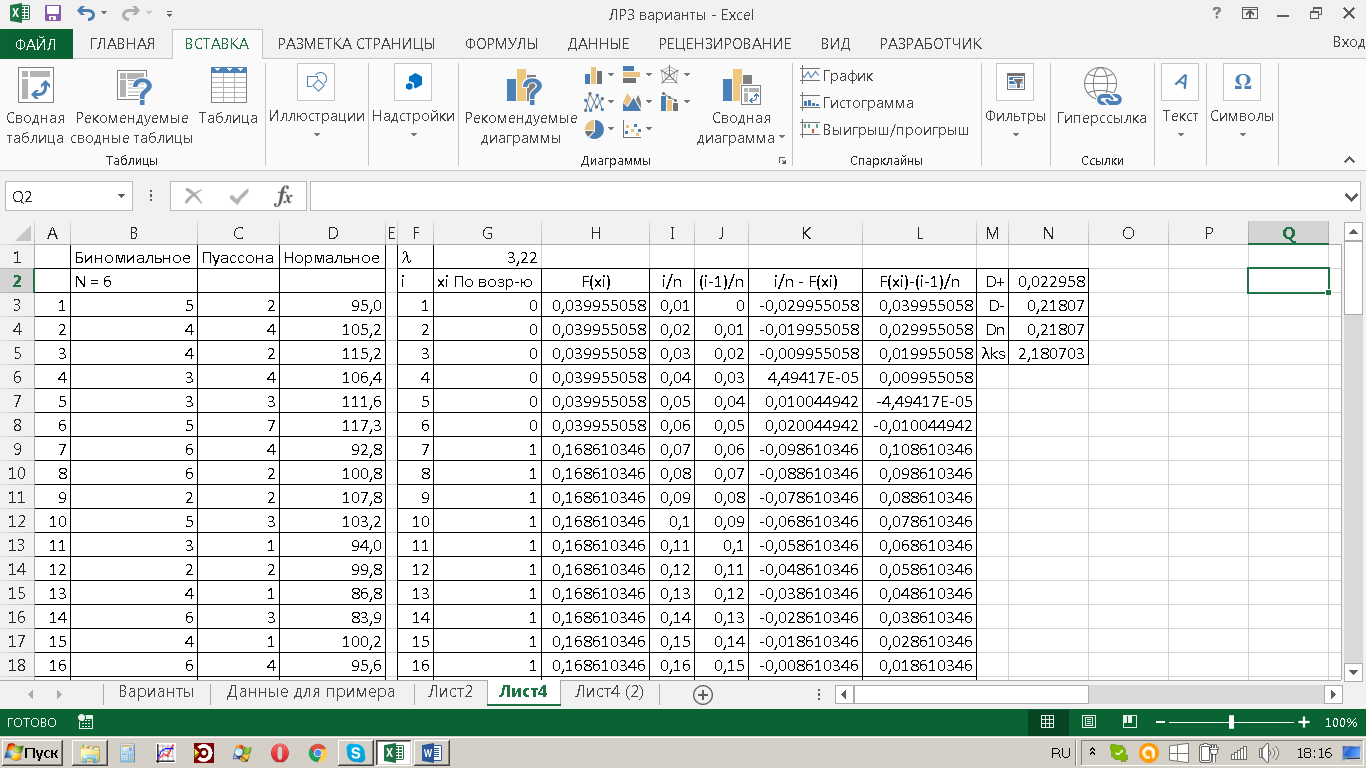


Рисунок 13 – Расчет критерия Колмогорова-Смирнова

В формулах решение приведено на рисунке 14:

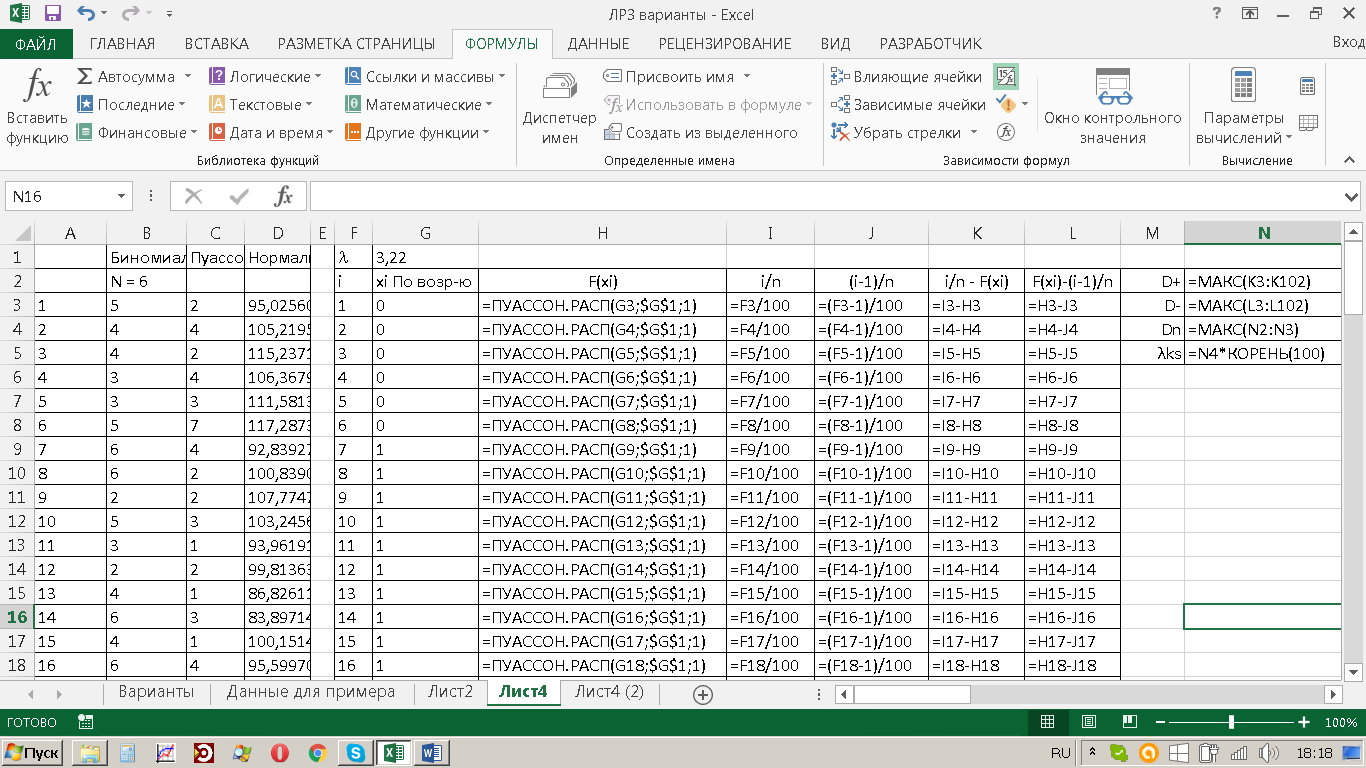


Рисунок 14 – Расчет критерия Колмогорова-Смирнова в режиме отображения формул

Наблюдаемое значение: 2,18

Критическое значение: 1,36

Так как наблюдаемое больше критического, то гипотеза о распределении Пуассона отвергается.

Также по рисунку 15 видно, что эмпирическая функция не попала между двумя прямыми, как должно было быть, если бы гипотеза была принята:



Рисунок 15 – Теоретические границы распределения и эмпирическая функция распределения при расчете критерия Колмогорова-Смирнова

**2.3 Критерий согласия Смирнова-Крамера-фон Мизеса**  **для распределения Пуассона**

Расчет критерия представлен на рисунках 16, 17.

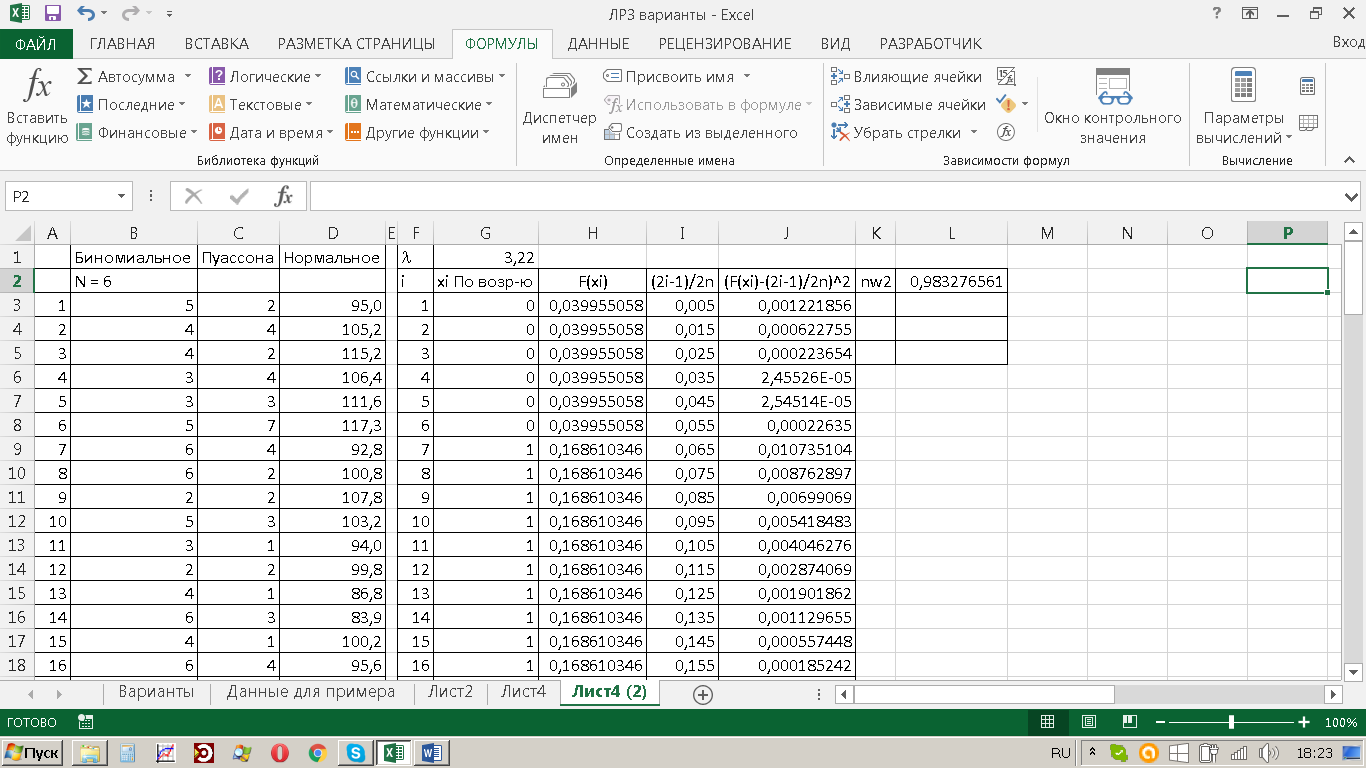


Рисунок 16 – Расчет критерия Смирнова-Крамера-фон Мизеса

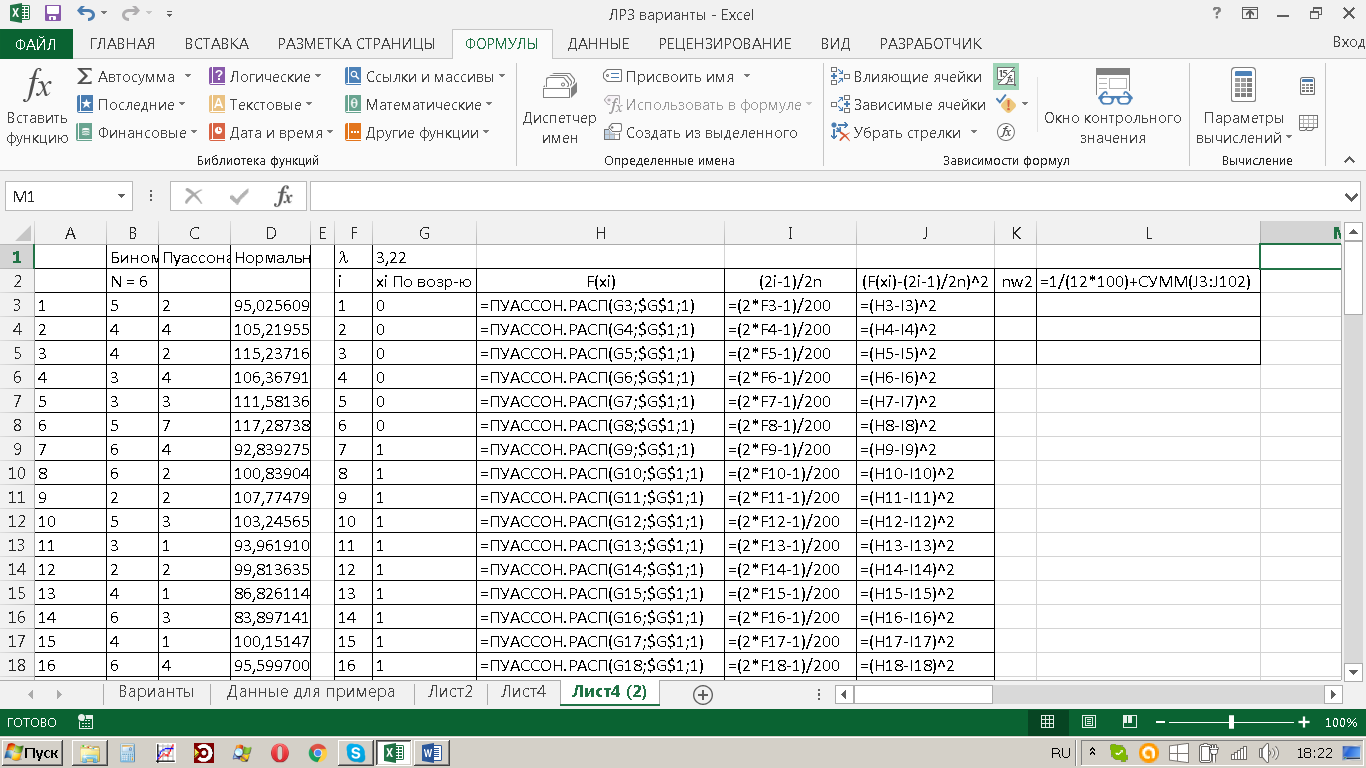


Рисунок 17 – Расчет критерия Смирнова-Крамера-фон Мизеса в режиме отображения формул

Наблюдаемое значение: 0,98

Критическое значение: 0,46

Так как наблюдаемое больше критического, то нулевая гипотеза отвергается. Также по рисунку 18 видно, что распределение прошло выше теоретического и плохо им описывается:

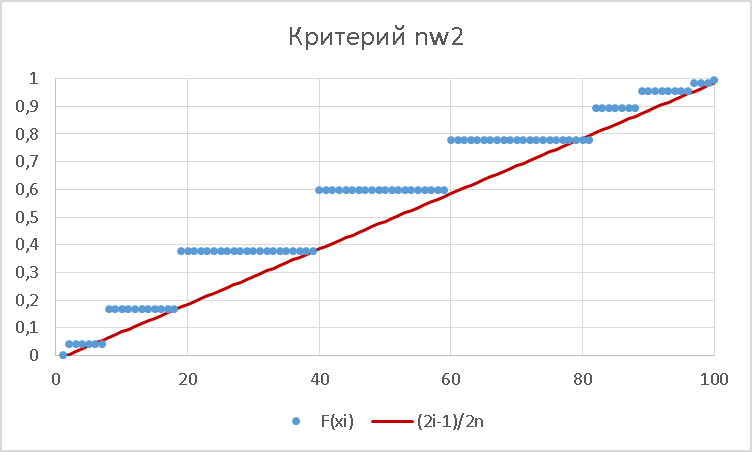


Рисунок 18 – Эмпирическая и теоретическая функции распределения при расчете по критерию Смирнова-Крамера-фон Мизеса

**3. Критерии согласия для нормального распределения**

**3.1 Предварительное исследование на нормальность**

1. **Графический критерий**

Упорядочим третью выборку по возрастанию. Рассчитаем затем частости *W i*  и квантили стандартного нормального распределения *z(Wi)*. Затем построим график в координатах:  *xi* по оси абсцисс*, zi* – по оси ординат. Можно выделить получившийся график, щелкнуть правой кнопкой мыши, выбрать пункт меню «Добавить линию тренда» - Линейная, и убедиться, что график хорошо описывается прямой (рисунок 19, 20):

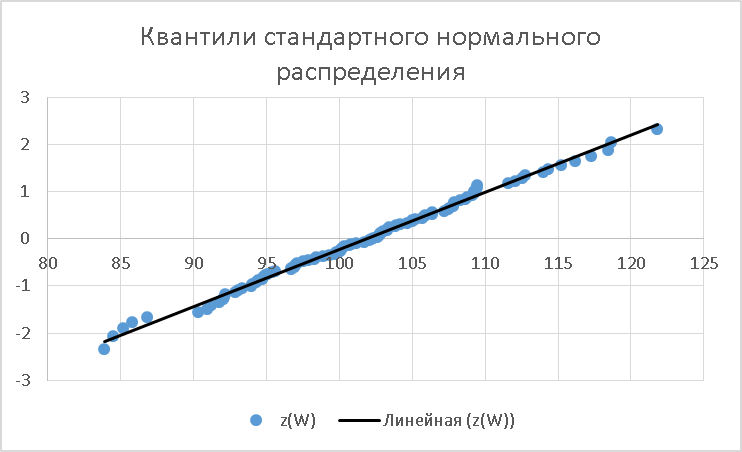
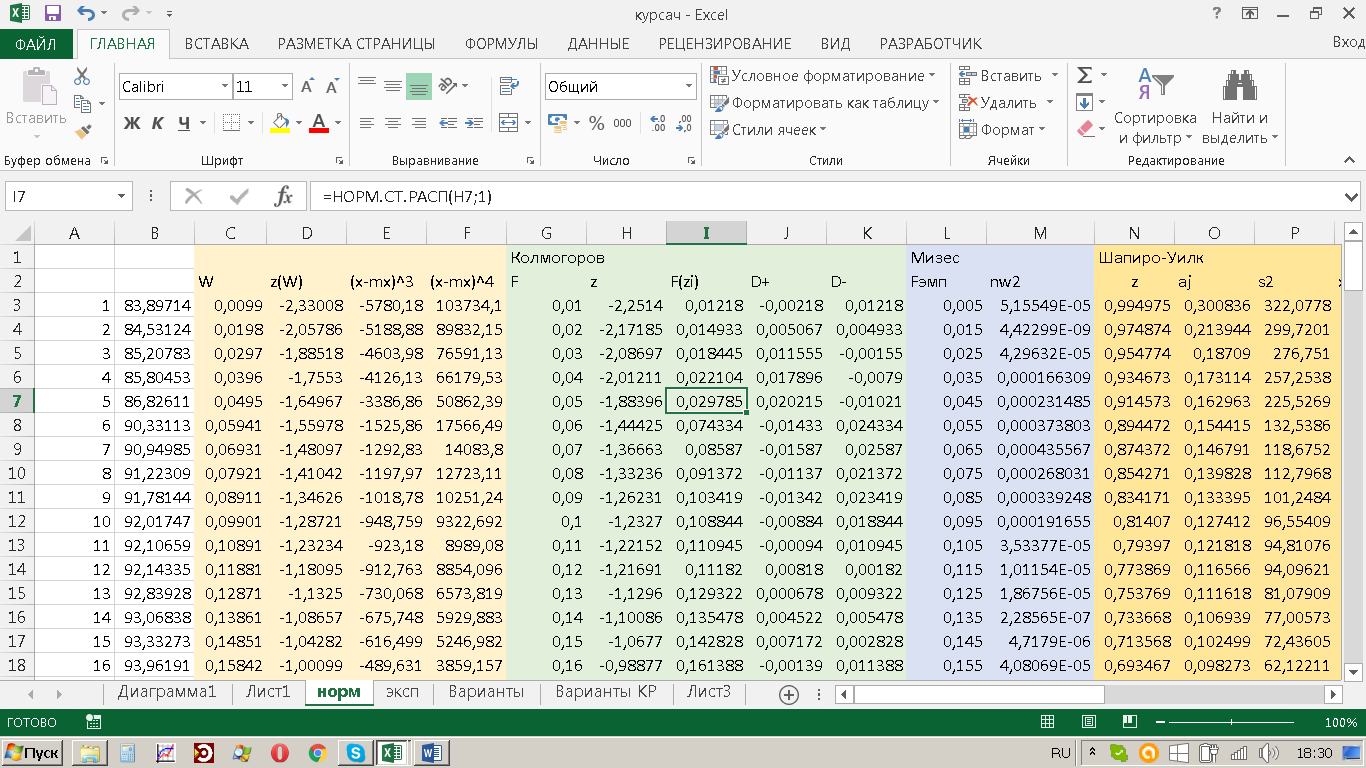


Рисунок 19 – Графический критерий нормальности

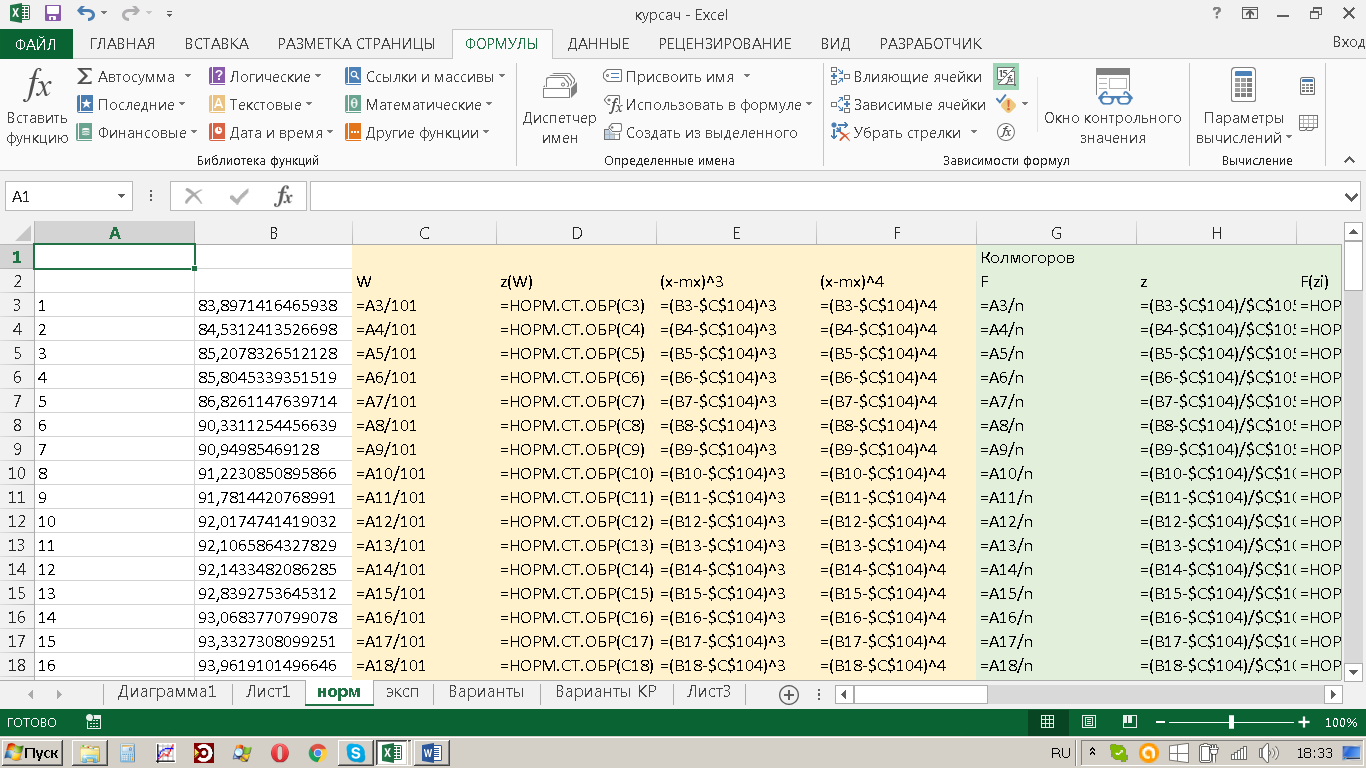


Рисунок 20 – Графический критерий нормальности в режиме отображения формул

Таким образом, нельзя отвергнуть гипотезу о нормальности распределения

1. **Оценка вида распределения по асимметрии и эксцессу**

Рассчитаем асимметрию (А) и эксцесс (Е):



В Excel вместо этих формул можно использовать статистические функции СКОС (для расчета А) и ЭКСЦЕСС (для расчёта Е).

Дисперсии асимметрии и эксцесса рассчитывают так:

.

Для удобства можно переименовать ячейку с цифрой 100, чтобы использовать название *n* в формулах (рис. 21):

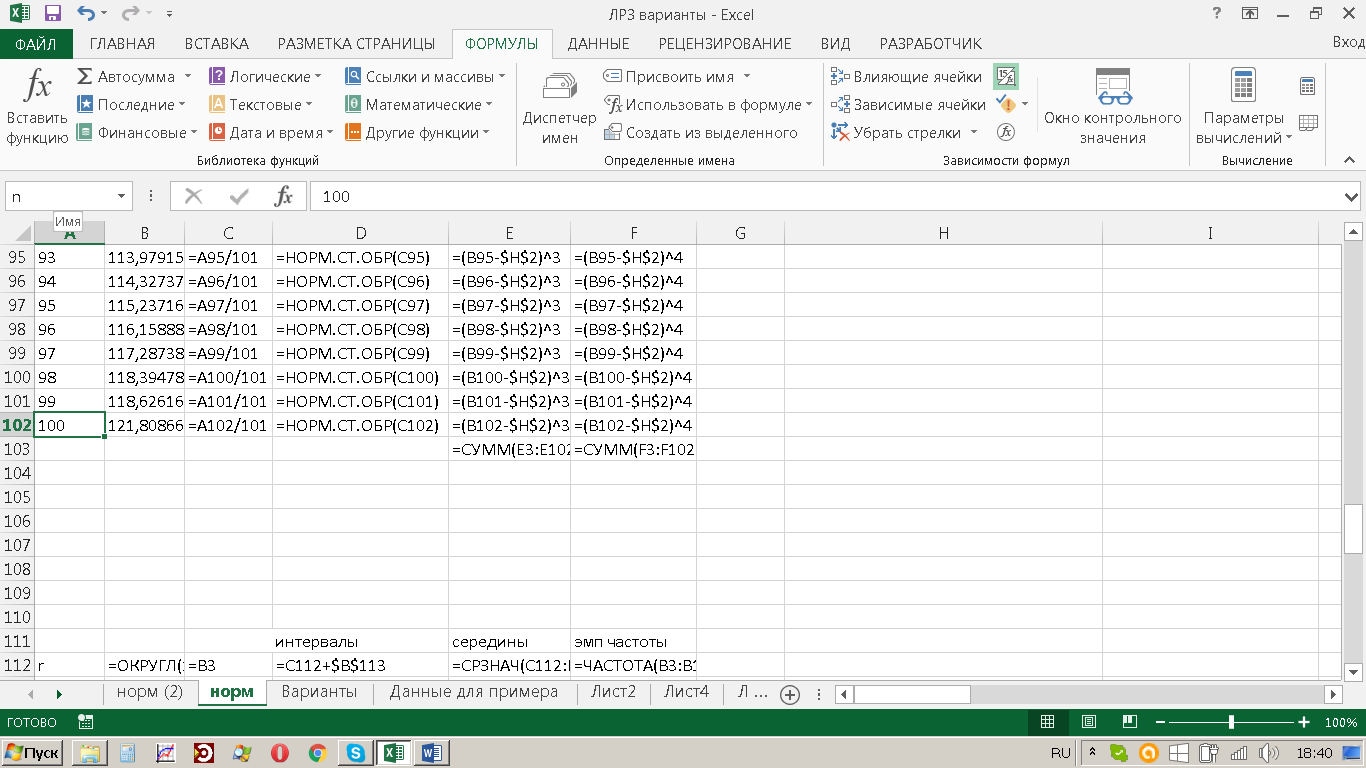


Рисунок 21 – Задание имени ячейке A102

Расчет критерия представлен рисунками 22-23.

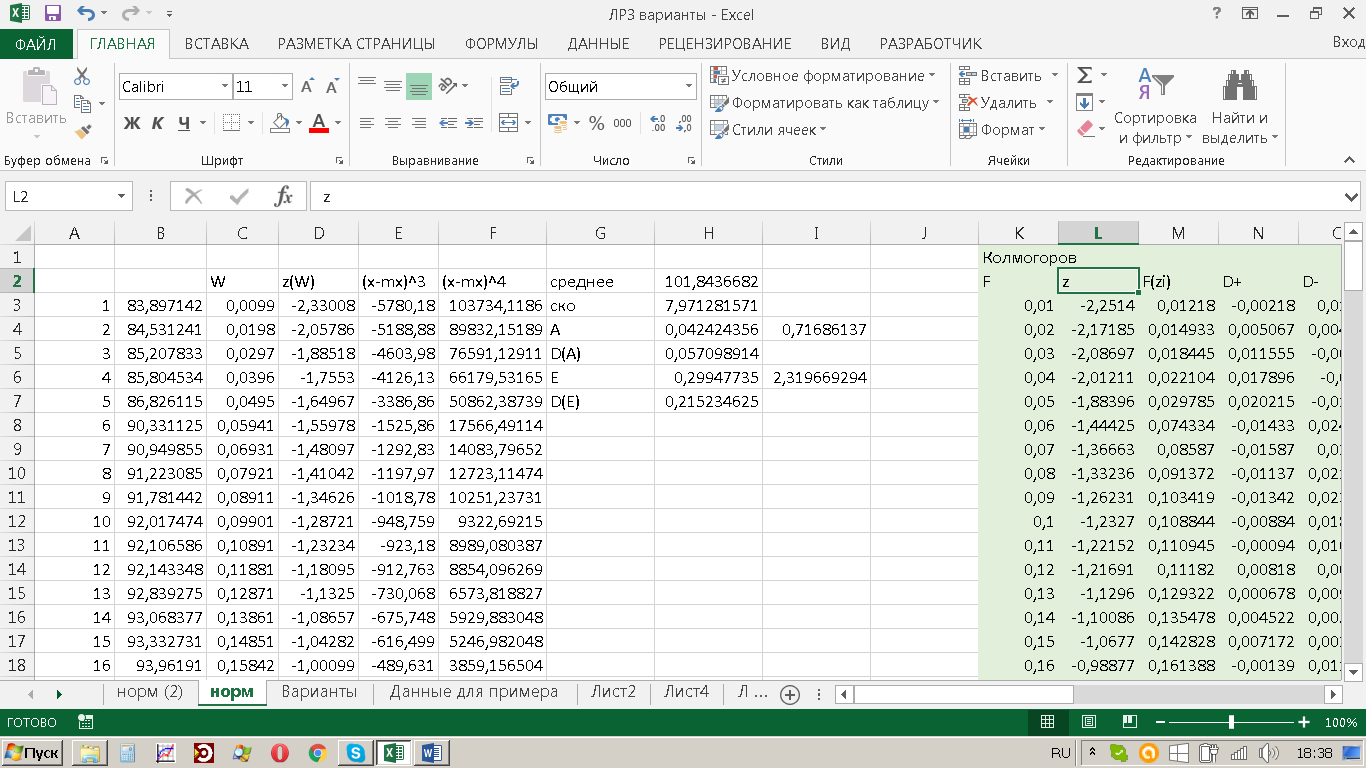


Рисунок 22 – Расчет критерия асимметрии и эксцесса

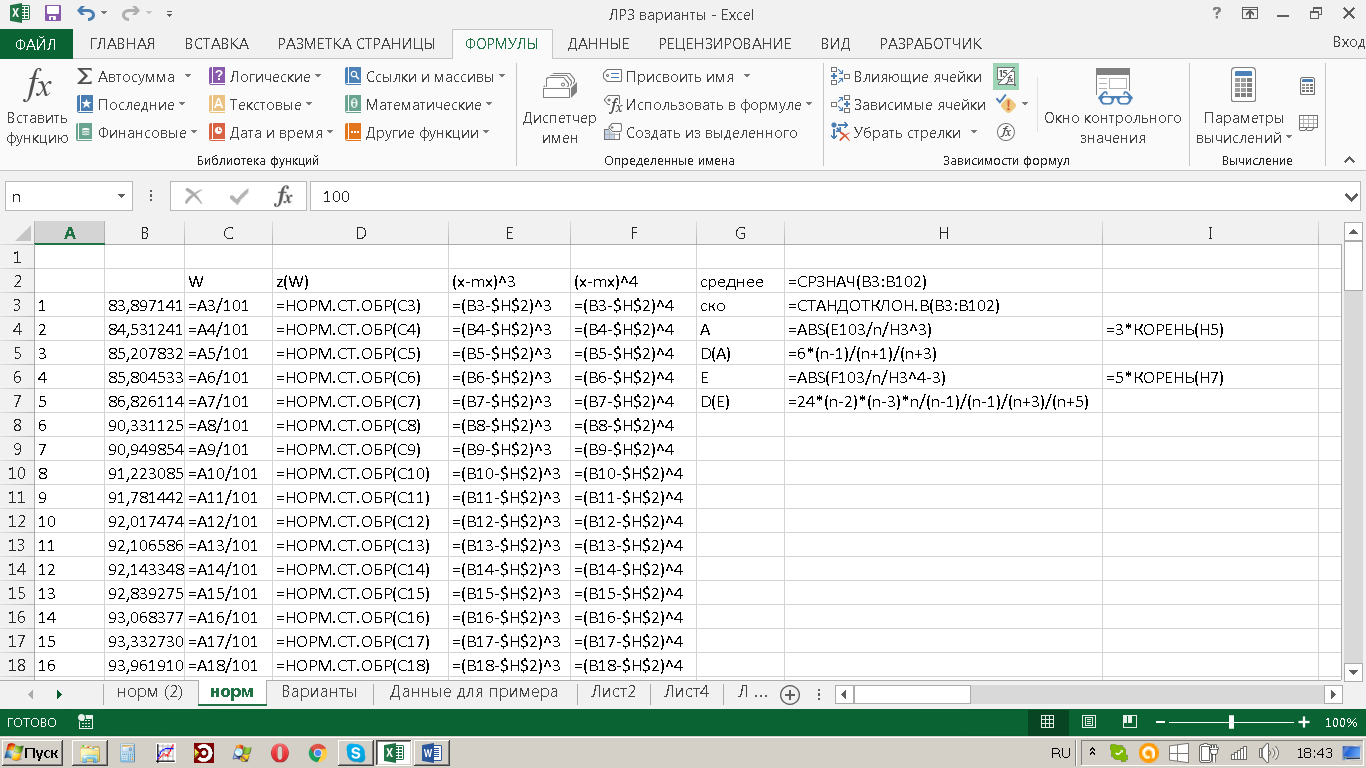


Рисунок 23 – Расчет критерия асимметрии и эксцесса в режиме отображения формул

Поскольку , то результаты испытаний считают распределёнными нормально.