Министерство образования Республики Беларусь

Учреждение образования

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ИНФОРМАТИКИ ИРАДИОЭЛЕКТРОНИКИ

Факультет компьютерных систем и сетей

Кафедра программного обеспечения информационных технологий

ОТЧЕТ

по расчётной работе

по МАТЕМАТИЧЕСКОЙ

СТАТИСТИКЕ

Выполнил Студент гр. 351004

Проверил

Бражалович А.И.

Н.С. Петюкевич

Минск 2024

# **1.Обозначения и используемые формулы**

## **Обозначения**

|  |  |
| --- | --- |
| **Обозначения** | **Пояснение** |
|  | Случайная величина |
|  | Частота |
|  | Относительная частота (частотность) |
| N | Объем выборки |
| F(x) | Функция распределения |
|  | Средняя величина |
|  | Средняя величина для интервала |
| h | Длина интервала |
|  | Число интервалов |
|  | Общая дисперсия |
|  | Дисперсия группы |
|  | Средняя из групповых дисперсий |
|  | Межгрупповая дисперсия |
|  | Среднее квадратичное отклонение |
|  | Мода |
|  | Медиана |
|  | Коэффициент эксцесса |
|  | Коэффициент асимметрии |
|  | Теоретическая частотность |
| V | Коэффициент вариации |
| R | Размах вариации |

## **Формулы**

|  |  |
| --- | --- |
| **Формула** | **Пояснение** |
|  | Размах вариации |
|  | Размер интервала |
|  | Относительная частота (частость) |
|  | Cреднее линейное отклонение |
|  | Оценка дисперсии |
|  | Cреднее квадратическое отклонение |
|  | Коэффициент вариации |
|  | Дисперсия группы |
|  | Общая дисперсия |
|  | Средняя из групповых дисперсий |
|  | Межгрупповая дисперсия |
|  | Коэффициент асимметрии |
|  | Коэффициент эксцесса |
|  | Количество интервалов |

Теорема сложения дисперсий: = +

# **2. Непрерывные данные**

## **2.1 Описание набора данных**

В качестве непрерывных данных в течении 100 дней собиралось кол-во шагов в день. Данные записывались в таблицу Exсel для удобности расчётов в будущем.

## **2.2 Вариационный ряд**

360, 416, 516, 1102, 1223, 3859, 3927, 4095, 4379, 4393, 4413, 4463, 5191, 5398, 5844, 5881, 5943, 6503, 6840, 6980, 7094, 7176, 7210, 7382, 7459, 7465, 7493, 7540, 7791, 7838, 7916, 8009, 8296, 8302, 8422, 8551, 8592, 8596, 8822, 8832, 8903, 8935, 9020, 9043, 9380, 9488, 9493, 9503, 9524, 9554, 9751, 10027, 10193, 10223, 10304, 10350, 10377, 10464, 10486, 10577, 10590, 10865, 10950, 11023, 11118, 11191, 11357, 11364, 11484, 11600, 11660, 11716, 11776, 12519, 12805, 13015, 13372, 13515, 13570, 13836, 13978, 14025, 14133, 14244, 14247, 14385, 14786, 14930, 15091, 15584, 15747, 15904, 16166, 17152, 17233, 17904, 18403, 18659, 20118, 21416

**2.3 Интервальный ряд**

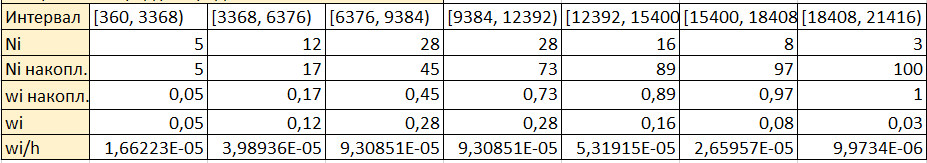
Для дальнейших вычислений был построен интервальный ряд распределения частот и частостей. Чтобы узнать количество интервалов, на которые следует разбивать данные, была использована формула Стерджеса.

В итоге по формуле получилось, что количество интервалов равно 7, а вычисленный позже шаг равен 3008.

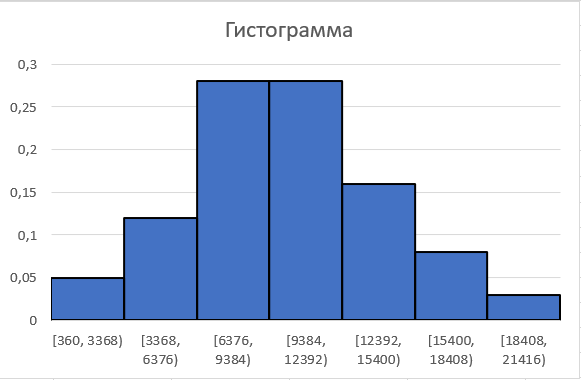
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| шаг | | | | 3008 | |  | | |
| **кол-во интервалов М** | | | | 7 | |
|  | |  | | |
|  | интервал | | Ni | | | | wi |
|  | 360 | | 5 | | | | 0,05 |
|  | 3368 | |  | | | |  |
|  | 3368 | | 12 | | | | 0,12 |
|  | 6376 | |  | | | |  |
|  | 6376 | | 28 | | | | 0,28 |
|  | 9384 | |  | | | |  |
|  | 9384 | | 28 | | | | 0,28 |
|  | 12392 | |  | | | |  |
|  | 12392 | | 16 | | | | 0,16 |
|  | 15400 | |  | | | |  |
|  | 15400 | | 8 | | | | 0,08 |
|  | 18408 | |  | | | |  |
|  | 18408 | | 3 | | | | 0,03 |
|  | 21416 | |  | | | |  |
|  | сумма | | 100 | | | | 1 |

**2.4 Гистограмма**

Для построения гистограммы была рассчитана плотность относительных частот, которая и послужила входными данными для гистограммы.



Гистограмма была построена с помощью встроенных возможностей Excel.



**2.5 Эмпирическая функция распределения**

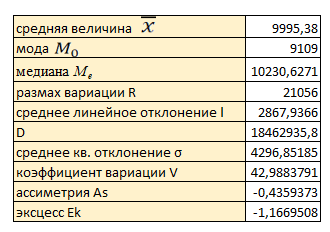
Для построения эмпирической функции распределения, следует вычислить вероятность того, что случайная величина Xi будет меньше, чем x, за x были взяты границы интервалов.



По вычисленным данным, с помощью встроенных возможностей Excel была построена эмпирическая функция распределения.

**2.6 Числовые характеристики**

Используя вышеописанные формулы были вычислены все основные числовые характеристики.



Анализируя данные характеристики, можно сделать несколько выводов о нашей выборке.

Как видно, средняя величина, мода и медиана практически совпадают, что означает, что выборка относительно симметрична относительно среднего значения. Небольшое отклонение медианы от среднего значения говорит о легкой асимметрии в сторону меньших значений.

Коэффициент вариации показывает нам степень рассеянности выборки, и 40% означает довольно значительную рассеянность.

Асимметрия получилась довольно маленькой и можно сказать, что выборка практически симметричная, но всё-таки она не равна нулю и указывает на незначительную асимметрию в левую сторону.

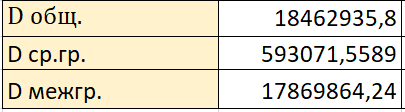
Эксцесс немного ниже нуля, что указывает на чуть более плосковершинное распределение относительно нормального распределения.

**2.7 Теорема сложения дисперсий**

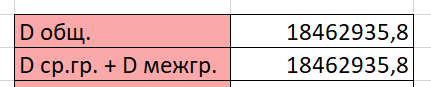
Для проверки теоремы сложения дисперсий были вычислены средние значения для каждого интервала, чтобы позже вычислить групповые дисперсии, а также некоторые промежуточные значения.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| интервал | Ni | wi | Xi ср. | D гр | D гр\*ni | (Xi ср. - X ср.)^2\*Ni |
| 360 | 5 | 0,05 | 723,4 | 132501,44 | 662507,2 | 429848065,6 |
| 3368 |  |  |  |  |  |  |
| 3368 | 12 | 0,12 | 4815,5 | 568617,5833 | 6823411 | 321973881,8 |
| 6376 |  |  |  |  |  |  |
| 6376 | 28 | 0,28 | 8013,929 | 596728,7806 | 16708406 | 109932193,4 |
| 9384 |  |  |  |  |  |  |
| 9384 | 28 | 0,28 | 10607,43 | 542289,8163 | 15184115 | 10488896,71 |
| 12392 |  |  |  |  |  |  |
| 12392 | 16 | 0,16 | 13903,19 | 515224,9023 | 8243598,4 | 244335351,3 |
| 15400 |  |  |  |  |  |  |
| 15400 | 8 | 0,08 | 16761,63 | 985034,2344 | 7880273,9 | 366256571,2 |
| 18408 |  |  |  |  |  |  |
| 18408 | 3 | 0,03 | 20064,33 | 1268281,556 | 3804844,7 | 304151463,7 |
| 21416 |  |  |  |  |  |  |

Далее вычислим межгрупповую и среднюю групповую дисперсии для проверки теоремы. А также возьмем уже вычисленную общую дисперсию



Подставив получившиеся данные в теорему, получим полное сходство.



Посчитав коэффициент Детерминации по формуле:

|  |  |
| --- | --- |
| коэф. Детерм. | 96,78777219 |

Это означает, что данные достаточно качественно сгруппированы.

**2.8 Проверка теорем**

Посмотрев на гистограмму плотностей относительных частот, можно выдвинуть гипотезу о нормальном распределении данной случайной величины Х.

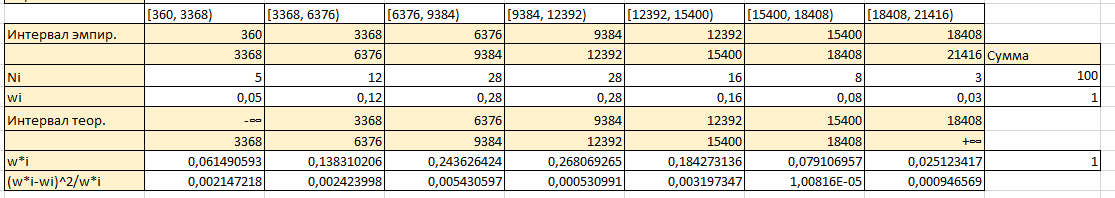
H0: Величина Х распределена по нормальному закону

H1: Величина Х не распределена по нормальному закону

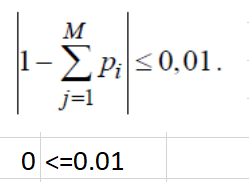
Для начала вычислим степень свободы:

|  |  |
| --- | --- |
| уровень значимости | 0.05 |
| Степени свободы | 4 |

Построим таблицу с интервалами для эмпирической функции и теоретической, а также вычислим теоретическую вероятности попадания случайной величины в интервал.



Проверим контрольную сумму теоретических вероятностей.



Критерий согласия Пирсона:

|  |  |
| --- | --- |
| Х^2набл. | 1,468680274 |
| Х^2табл | 9,487729037 |
|  |  |
| нет оснований отклонить гипотезу |  |

По графику также видно, что случайная величина близка к нормальному распределению.

Критерий согласия Романовского:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| R | 0,894956672 | | | < | 3 |
|  |  | | |  |  |
| нет оснований отклонить гипотезу |  |  |

Критерий согласия Ястремского:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| R | 0,625064853 | < | | 3 |
|  |  |  | |  |
| нет оснований отклонить гипотезу |  | |  | | |

Вывод. Ни по одному из критериев мы не имеем оснований отклонить гипотезу, следовательно у нас есть основания полагать что наша случайная величина Х распределена по нормальному закону.

# **3. Дискретные данные**

## **2.1 Описание набора данных**

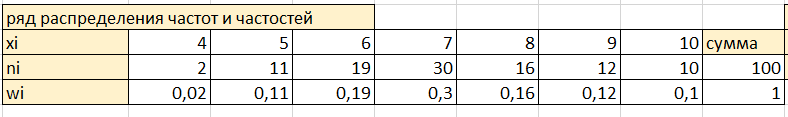
В качестве дискретных данных в течении 100 дней собиралось кол-во часов сна в день с математическим округлением. Данные записывались в таблицу Exсel для удобности расчётов в будущем.

## **2.2 Вариационный ряд**

4, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 7, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10

**2.3 Дискретный ряд**

Для дальнейших вычислений был построен дискретный ряд распределения частот и частостей. Для этого просто подсчитаем количество повторений каждого варианта.



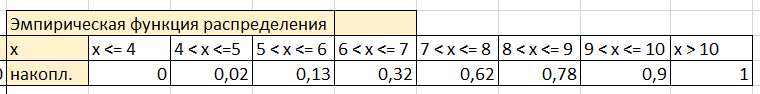
**2.4 Полигон**

Для построения полигона нам понадобится частость каждого варианта.

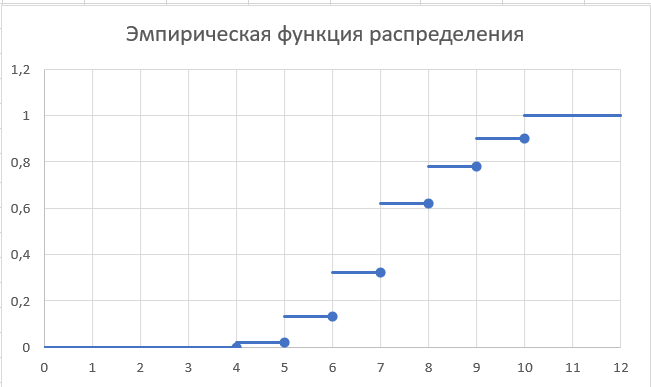


**2.5 Эмпирическая функция распределения**

Для построения эмпирической функции распределения, следует вычислить вероятность того, что случайная величина Xi будет меньше, чем x.

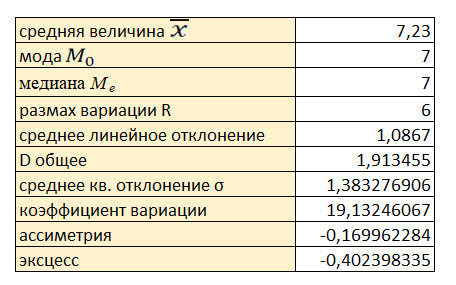


По вычисленным данным, с помощью встроенных возможностей Excel была построена эмпирическая функция распределения.



**2.6 Числовые характеристики**

Используя вышеописанные формулы были вычислены все основные числовые характеристики.



Анализируя данные характеристики, можно сделать несколько выводов о нашей выборке.

Как видно, средняя величина, мода и медиана практически совпадают, что означает, что выборка относительно симметрична относительно среднего значения.

Коэффициент вариации показывает степень рассеянности выборки. Значение около 15% указывает на умеренную рассеянность данных относительно среднего значения.

Асимметрия получилась довольно маленькой и можно сказать, что выборка практически симметричная, но всё-таки значение асимметрии -0.12 указывает на легкую асимметрию в левую сторону, в сторону меньших значений.

Эксцесс немного выше нуля (0.598410155), что указывает на чуть более остроконечное распределение относительно нормального распределения.

**2.7 Проверка теорем**

Проверим гипотезу о Пуассоновском распределении.

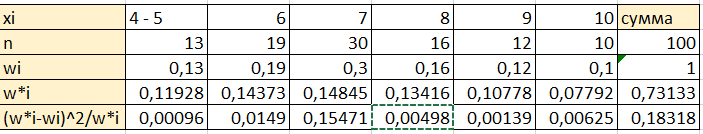
H0: Величина Х соответствует распределению Пуассона

H1: Величина Х не соответствует распределению Пуассона

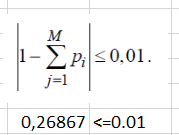
Для начала вычислим степень свободы:



Построим таблицу с интервалами для эмпирической функции и теоретической, а также вычислим теоретическую вероятности попадания случайной величины в интервал.



Проверим контрольную сумму теоретических вероятностей.



Критерий согласия Пирсона:



|  |  |
| --- | --- |
| гипотеза отвергается |  |

Критерий согласия Романовского:

|  |  |  |  |
| --- | --- | --- | --- |
| Критерий Романовского |  |  |  |
|  |  |  |  |
| R | 8,15901 | < | 3 |
|  |  |  |  |
| гипотеза отвергается |  |  |  |

Критерий согласия Ястремского:

|  |  |  |  |
| --- | --- | --- | --- |
| Критерий Ястремского |  |  |  |
|  |  |  |  |
| R | 4,634006 | < | 3 |
|  |  |  |  |
| гипотеза отвергается |  |  |  |

Вывод. Каждый критерий отверг гипотезу, следовательно у нас есть основания полагать что наша случайная величина Х не соответствует распределению Пуассона.