

Министерство образования Республики Беларусь

Учреждение образования
«Белорусский государственный университет
информатики и радиоэлектроники»

Факультет компьютерных систем и сетей
Кафедра программного обеспечения информационных технологий

Дисциплина «Теория вероятностей и математическая статистика»

ОТЧЁТ
к расчётной работе

СБОР И АНАЛИЗ ДИСКРЕТНЫХ И НЕПРЕРЫВНЫХ
ДАННЫХ

Студент группы
№351001
Ушаков А. Д.

Преподаватель
Петюкевич Н. С.

Минск 2024

СОДЕРЖАНИЕ

| | |
|---|----|
| Введение | 3 |
| 1 Теоретические сведения | 4 |
| 1.1 Оценка закона распределения..... | 4 |
| 1.2 Точечные оценки числовых характеристик | 5 |
| 1.3 Формулы и вычисления..... | 6 |
| 2 Анализ дискретных данных | 9 |
| 2.1 Вариационный ряд | 9 |
| 2.2 Общие характеристики собранных данных | 10 |
| 2.3 Дискретный ряд распределения и числовые характеристики | 10 |
| 2.4 Диаграммы для дискретных данных | 12 |
| 2.5 Проверка гипотезы: распределение Пуассона | 13 |
| 2.6 Вывод по второй главе | 16 |
| 3 Анализ непрерывных данных | 17 |
| 3.1 Общие характеристики выборки | 17 |
| 3.2 Оценки числовых характеристик | 17 |
| 3.3 Вычисление дисперсий..... | 19 |
| 3.4 Диаграммы для непрерывных данных | 20 |
| 3.5 Проверка гипотезы: нормальное распределение | 21 |
| 3.6 Вывод по третьей главе | 24 |
| Заключение | 25 |
| Список использованных источников | 26 |
| Приложение А (обязательное) Собранные данные | 26 |
| Приложение Б (обязательное) Вариационный и интервальный ряды распределения для непрерывных данных | 28 |
| Приложение В (справочное) Анализ дискретных данных в Excel | 29 |
| Приложение Г (справочное) Анализ непрерывных данных в Excel..... | 30 |

ВВЕДЕНИЕ

Математической статистикой называется наука, занимающаяся методами обработки опытных данных, полученных в результате наблюдений над случайными явлениями. Любой такой результат можно представить как совокупность значений, принятых в результате n опытов случайной одномерной или многомерной величиной.

Генеральной совокупностью опыта называется множество объектов, из которых производится выборка. Каждый из объектов задает фиксированное значение случайной величины X . Количество (N) входящих в генеральную совокупность объектов называют объемом генеральной совокупности. Она может состоять из бесчисленного множества объектов.

Выборка – множество случайно отобранных объектов (значений) из генеральной совокупности. Объемом выборки n называется число входящих в нее объектов. К выборке предъявляется требование, чтобы она адекватно представляла генеральную совокупность, т.е. была репрезентативной (представительной). В силу закона больших чисел можно утверждать, что выборка будет репрезентативной, если ее осуществлять случайно, т.е. каждый из объектов генеральной совокупности имеет одинаковую вероятность попасть в выборку. Очевидно, что можно осуществить в одинаковых условиях k выборок объема n и получить различные совокупности значений случайной величины X :

Вариационным рядом называется выборка, полученная в результате расположения значений исходной выборки в порядке возрастания. Ее значения называются вариантами.

Одной из главных задач математической статистики является определение закона распределения случайной величины X .

1 ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

1.1 Оценка закона распределения

Эмпирическая функция распределения [1] случайной величины X равна частоте того, что X примет значение меньшее, чем аргумент функции x , и определяется формулой (1).

$$F^*(x) = p^*(X < x) = \begin{cases} 0, & x \leq \hat{x}_1, \\ \vdots \\ \frac{i}{n}, & \hat{x}_i < x \leq \hat{x}_{i+1}, \\ \vdots \\ 1, & x > \hat{x}_n. \end{cases} \quad (1)$$

При $n \rightarrow \infty$ эмпирическая функция распределения $F^*(x)$ сходится по вероятности к теоретической функции распределения $F(x)$.

Основные свойства функции $F^*(x)$:

1 $0 \leq F^*(x) \leq 1$.

2 $F^*(x)$ – неубывающая ступенчатая функция.

3 $F^*(x) = 0$, для $x \leq \hat{x}_1$.

4 $F^*(x) = 1$, для $x > \hat{x}_n$.

Эмпирическая функция распределения является наилучшей оценкой закона распределения (несмещенной, состоятельной, эффективной). Недостаток функции $F^*(x)$ заключается в ее невысокой наглядности: визуально сложно определить закон распределения случайной величины X . Статистический ряд распределения вероятностей определяется по исходной выборке объемом n , если анализируемая случайная величина X является дискретной с известным множеством значений $\{x_1, \dots, x_m\}$.

Интервальный статистический ряд вероятностей строится по исходной выборке, если анализируемая случайная величина X является непрерывной.

При построении интервального статистического ряда вероятностей используют следующие методы разбиения диапазона значений на интервалы:

- равноинтервальный, т.е. все интервалы одинаковой длины;
- равновероятностный, т.е. границы интервалов выбирают так, чтобы в каждом интервале было одинаковое число выборочных значений.

Гистограмма – статистический аналог графика плотности вероятности $F^*(x)$ случайной величины, и она строится по интервальному статистическому ряду (см. рисунок 1.1.1). Гистограмма представляет собой совокупность прямоугольников, построенных, как на основаниях, на интервалах

статистического ряда с высотой равной статистической плотности вероятности в соответствующем интервале.

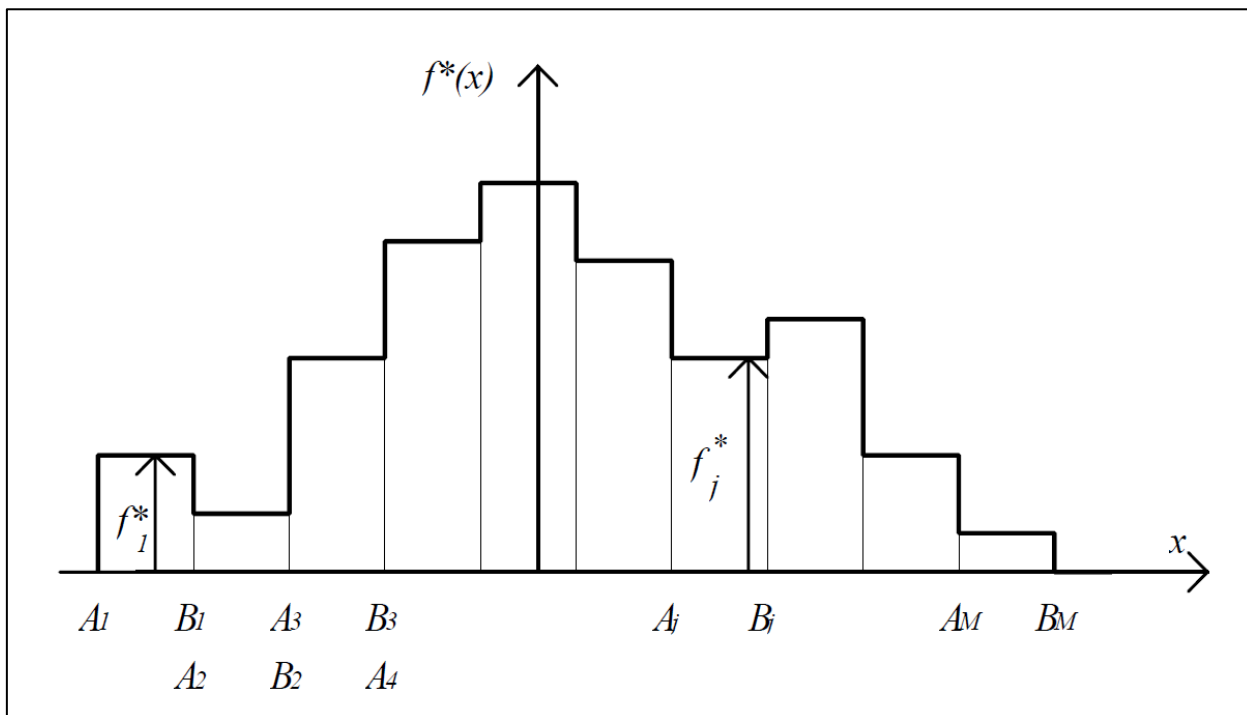


Рисунок 1.1.1 – Гистограмма

Для равноинтервального метода все прямоугольники гистограммы имеют одинаковую ширину, а для равновероятностного метода – одинаковую площадь. Сумма площадей всех прямоугольников гистограммы равна 1.

Достоинства гистограммы: простота построения, высокая наглядность.

1.2 Точечные оценки числовых характеристик

Статистической оценкой \hat{Q} параметра Q распределения называется приближенное значение параметра, вычисленное по результатам эксперимента (по выборке). Статистические оценки делятся на точечные и интервальные.

Оценка математического ожидания. На основании теоремы Чебышева в качестве состоятельной оценки математического ожидания может быть использовано среднее арифметическое значений выборки \bar{X} , называемое выборочным средним.

Оценка дисперсии. В качестве состоятельной оценки дисперсии может быть использовано среднее арифметическое квадратов отклонений значений выборки от выборочного среднего.

Оценка вероятности. На основании теоремы Бернулли несмещенная состоятельная и эффективная оценка вероятности случайного события А в схеме независимых опытов равна частоте этого события.

1.3 Формулы и вычисления

На рисунке 1.3.1 показано, каким образом строить интервальный ряд, гистограмму и вычислять ошибки.

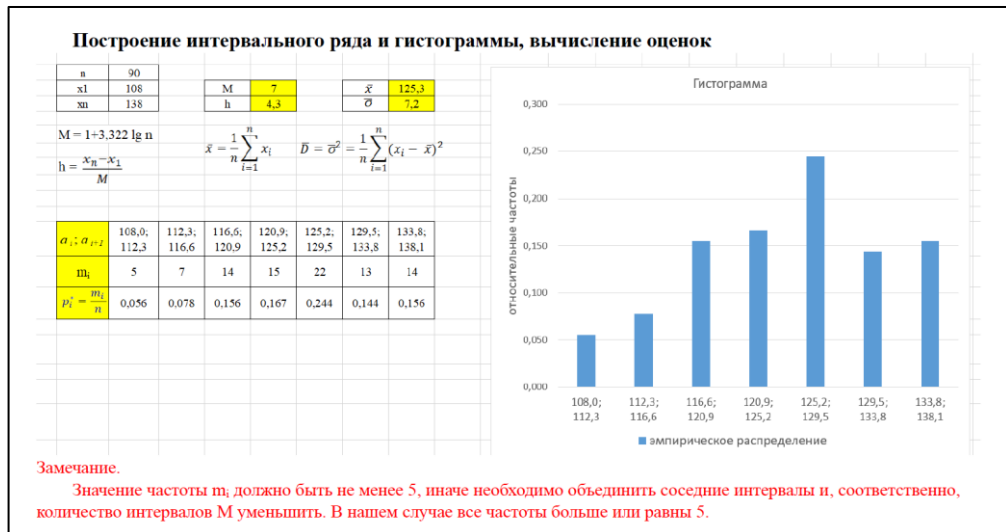


Рисунок 1.3.1 – Интервальный ряд, гистограмма, вычисление ошибок

На рисунке 1.3.2 показано, каким образом выдвигать, проверять и отвергать гипотезы.

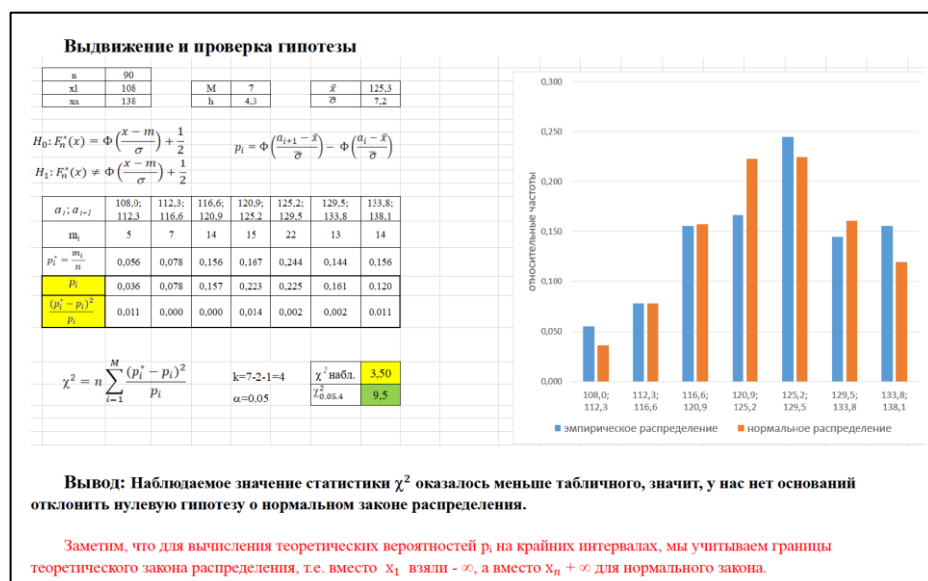


Рисунок 1.3.2 – Выдвижение и проверка гипотез

На рисунке 1.3.3 продемонстрированы формулы, используемые в расчётах при анализе данных.

X_i – случайная величина

$X_{\text{ср}}, \bar{X}$ – средняя выборочная

N_i – частота

N – объем выборки

$F(X)$ – функция распределения

$h = R / M$ – длина интервала

$M = \text{int}(1 + 3,322 * \lg(N))$ – количество интервалов при непрерывных данных

$D = \frac{\sum (X - X_{\text{ср}})^2 n}{\sum n}$ – дисперсия

$D_{\text{общ}} = \frac{\sum (X - \bar{X}_{\text{общ}})^2}{n}$ – общая дисперсия

$D_{\text{ср.гр.}} = \frac{\sum \sigma_i^2 n_i}{\sum n_i}$ – средняя из групповых дисперсий

$D_{\text{гр.}} = \frac{\sum_{j=1}^k (X_j - \bar{X}_i)^2 n_j}{n_i}$ – дисперсия группы

$D_{\text{межгр}} = \frac{\sum_{i=1}^k (\bar{X}_i - \bar{X})^2 * n_i}{n}$ – межгрупповая дисперсия

Теорема сложения дисперсий: $D_{\text{общ}} = D_{\text{ср гр}} + D_{\text{межгр}}$

$R = X_{\text{max}} - X_{\text{min}}$ – размах вариации

$l = \frac{\sum |X - \bar{X}| n}{\sum n}$ – среднее линейное отклонение

$\sigma = \sqrt{D}$ – среднее квадратическое отклонение

$V = \frac{\sigma}{\bar{X}} 100\%$ – коэффициент вариации

$E_x = \frac{\mu_4}{\sigma^4} - 3$ – эксцесс

$\nu_k = \frac{\sum x^k f}{\sum f}$ – Начальным моментом k -го порядка называется величина

Центральным моментом m -го порядка называется величина

$\mu_m = \frac{\sum (x - \bar{x})^m f}{\sum f}, \quad \nu_1 = \bar{x}, \quad \mu_2 = \sigma^2$

Рисунок 1.3.3 – Основные формулы для анализа данных

На рисунке 1.3.4 показаны основные положения теоремы о сложении дисперсий.

Для сгруппированных данных вводятся следующие понятия:

1) Групповая дисперсия

$$\sigma_i^2 = \frac{\sum_{j=1}^{n_i} (x_j - \bar{x}_i)^2}{n_i},$$

где n_i – кол-во вариант в i -ой группе (частота);
 \bar{x}_i – выборочное среднее i -ой группы;

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_j.$$

2) Средняя из групповых дисперсий

$$\bar{\sigma}^2 = \frac{\sum_{i=1}^k \sigma_i^2 * n_i}{n},$$

где k – количество групп;

$$n = \sum_{i=1}^k n_i \text{ – объем выборки;}$$

3) Межгрупповая дисперсия

$$\delta_{\text{межгр}}^2 = \frac{\sum_{i=1}^k (\bar{x}_i - \bar{x})^2 * n_i}{n},$$

где \bar{x} – средняя величина;

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j.$$

4) Общая дисперсия

$$\sigma_{\text{общ}}^2 = \frac{\sum_{j=1}^n (x_j - \bar{x})^2}{n}$$

Теорема о сложении дисперсий

$$\sigma_{\text{общ}}^2 = \bar{\sigma}^2 + \delta_{\text{межгр}}^2$$

Рисунок 1.3.4 – Теорема о сложении дисперсий

Проверить качество группировки можно с помощью эмпирического коэффициента детерминации (см. рисунок 1.3.5).

Эмпирический коэффициент детерминации

$$\eta^2 = \frac{\delta_{\text{межгр}}^2}{\sigma_{\text{общ}}^2}$$

характеризует качество группировки.

Рисунок 1.3.5 – Эмпирический коэффициент детерминации

2 АНАЛИЗ ДИСКРЕТНЫХ ДАННЫХ

Для просмотра собранных данных можно обратиться к приложению А. В непрерывные данные были внесены незначительные корректировки с целью грамотного использования методов статистического анализа на практике.

2.1 Вариационный ряд

На рисунке 2.1.1 представлена упорядоченная выборка собранных дискретных данных.

[illegible]

Рисунок 2.1.1 – Вариационный ряд

Вариационный ряд представляет собой упорядоченное множество значений количества приёмов пищи в день, полученных в результате наблюдений на протяжении 50 дней. В ряду можно выделить преобладающие значения: наименьшее количество приёмов пищи зафиксировано на уровне 2 раз в день, наиболее часто встречающееся значение — 3 и 4 приёма пищи в день, что указывает на стабильность в распорядке питания в исследуемый период. Более редкие значения, такие как 6 приёмов пищи в день, являются единичными наблюдениями и могут рассматриваться как исключения или особые случаи. Данный ряд позволяет наглядно оценить распределение данных и подготовить основу для дальнейших статистических расчетов, включая определение числовых характеристик и построение дискретного ряда распределения.

2.2 Общие характеристики собранных данных

На рисунке 2.2.1 показаны данные, которые в целом характеризуют выборку.

| Общие характеристики | | | | |
|----------------------|------------------------|-------|-------|--|
| Объём выборки | Сумма всех значений СВ | X_min | X_max | |
| 50 | 191 | 2 | 6 | |

Рисунок 2.2.1 – Общие характеристики выборки

2.3 Дискретный ряд распределения и числовые характеристики

На рисунке 2.3.1 представлен дискретный ряд распределения частот и частостей для анализируемой выборки, а рисунок 2.3.2 отражает точечные оценки числовых характеристик.

| Дискретный ряд распределения частот и частостей | | | | | | Итоговые суммы |
|---|--------|---------|--------|---------|--------|----------------|
| X _i | 2 | 3 | 4 | 5 | 6 | - |
| N _i | 1 | 19 | 20 | 8 | 2 | 50 |
| w _i | 0,02 | 0,38 | 0,4 | 0,16 | 0,04 | 1 |
| X _i - X̄ | 1,82 | 0,82 | 0,18 | 1,18 | 2,18 | 6,18 |
| (X _i - X̄) * N _i | 1,82 | 15,58 | 3,6 | 9,44 | 4,36 | 34,8 |
| (X _i - X̄) ² | 3,3124 | 0,6724 | 0,0324 | 1,3924 | 4,7524 | 10,162 |
| N _i * (X _i - X̄) ² | 3,3124 | 12,7756 | 0,648 | 11,1392 | 9,5048 | 37,38 |

Рисунок 2.3.1 – Дискретный ряд распределения

| Числовые характеристики | | | | | | | | | |
|------------------------------|-------------------|----|----|---------------|-------------|--------------|-------------|---------------|------------|
| Выборочная средняя \bar{X} | Размах вариации R | Mo | Me | Ср_лин_откл I | Дисперсия D | СКО σ | Кэфф_вар V | Эксцесс E_x | Асимметрия |
| 3,82 | 4 | 4 | 4 | 0,696 | 0,7476 | 0,864638653 | 0,226345197 | -0,011259235 | -0,18 |

Рисунок 2.3.2 – Точечные оценки числовых характеристик

Выборочная средняя (\bar{X}) представляет собой среднее арифметическое всех наблюдаемых значений случайной величины и в данном случае равна 3,82. Этот показатель характеризует центр распределения и позволяет определить усреднённое значение, вокруг которого группируются данные.

Размах вариации (R) равен 4 и показывает разницу между максимальным и минимальным значениями в выборке. Это значение отражает степень разброса данных и демонстрирует, насколько сильно отличаются крайние наблюдения.

Мода (Mo) принимает значение 4 и показывает наиболее часто встречающееся значение в выборке. Этот показатель позволяет определить, какое значение имеет наибольшую вероятность появления среди наблюдаемых данных.

Медиана (Me) также равна 4 и делит набор данных на две равные части. Половина значений в выборке меньше медианы, а другая половина — больше. Медиана является устойчивым показателем, который менее чувствителен к выбросам, чем выборочная средняя.

Среднее линейное отклонение (Ср_лин_откл I) имеет значение 0,696. Этот показатель показывает, насколько в среднем каждое значение выборки отклоняется от выборочной средней. Он является более наглядной мерой разброса данных по сравнению с дисперсией, поскольку представляет отклонения в тех же единицах измерения, что и исходные данные.

Дисперсия (D) равна 0,7476 и характеризует средний квадрат отклонений значений от выборочной средней. Дисперсия демонстрирует разброс данных относительно среднего значения: чем выше дисперсия, тем больше разброс.

Среднеквадратическое отклонение (СКО, σ) составляет 0,8646. Этот показатель является корнем из дисперсии и показывает, насколько в среднем значения отклоняются от среднего арифметического. Он также измеряется в тех же единицах, что и исходные данные, что облегчает его интерпретацию.

Коэффициент вариации (V) равен 0,2263 (22,63%) и показывает относительный разброс данных относительно выборочной средней. Этот показатель особенно полезен для сравнения разброса между наборами данных с разными средними значениями.

Эксцесс (E_x) принимает значение -0,0113 и характеризует остроту пика распределения. Положительный эксцесс указывает на остроконечное

распределение, тогда как отрицательный — на более плоский пик по сравнению с нормальным распределением.

Асимметрия составляет $-0,18$ и показывает степень смещения распределения. Отрицательное значение говорит о том, что распределение имеет небольшое смещение влево, что означает более длинный левый хвост.

Формулы для оценки числовых характеристик приведены в теоретической части расчётной работы.

2.4 Диаграммы для дискретных данных

На рисунке 2.4.1 показан график эмпирической функции распределения.

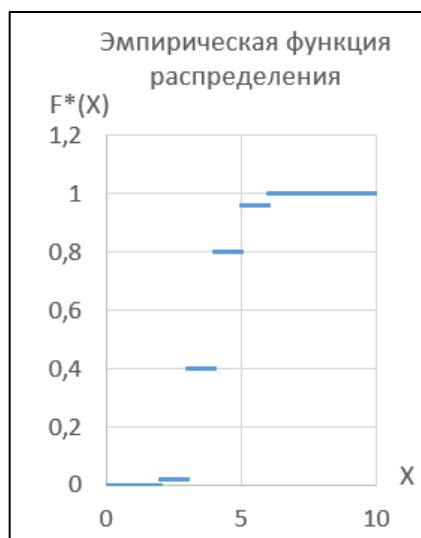


Рисунок 2.4.1 – Эмпирическая функция распределения

Получившаяся эмпирическая функция распределения задаётся формулой (2.1):

$$F^*(X) = \begin{cases} 0,00, & X \leq 2; \\ 0,02, & 2 < X \leq 3; \\ 0,40, & 3 < X \leq 4; \\ 0,80, & 4 < X \leq 5; \\ 0,96, & 5 < X \leq 6; \\ 1,00, & X > 6. \end{cases} \quad (2.1)$$

На рисунке 2.4.2 (следующая страница) изображён полигон [2], задающий зависимость относительных частот от случайной величины.

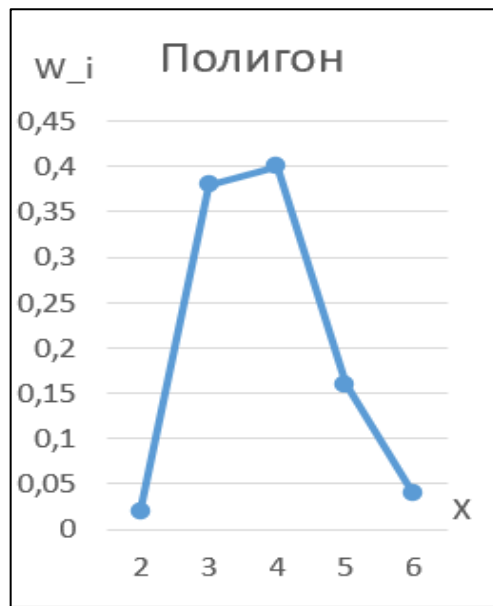


Рисунок 2.4.2 – Полигон

2.5 Проверка гипотезы: распределение Пуассона

Для начала имеет место привести диаграммы сравнения эмпирических данных и теоретических (рис. 2.5.1).

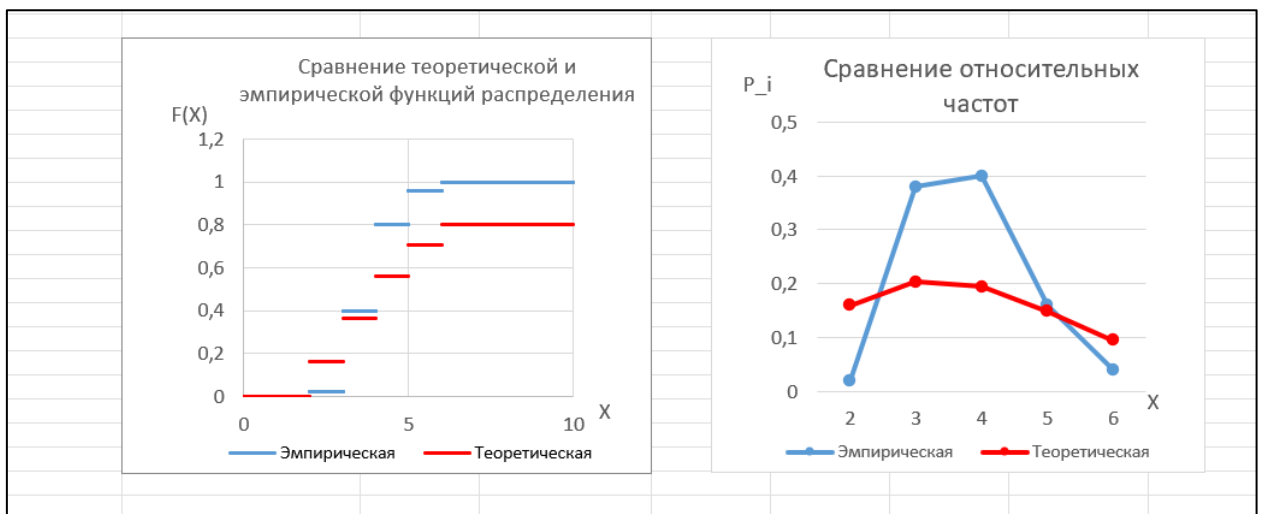


Рисунок 2.5.1 – Диаграммы сравнения теоретических и эмпирических данных

На первом графике представлено сравнение эмпирической и теоретической функций распределения. Эмпирическая функция (синяя линия) отображается в виде ступенчатой кривой, которая показывает накопленные относительные частоты значений случайной величины. Теоретическая функция распределения (красная линия) демонстрирует ожидаемое поведение

распределения. Однако заметны существенные расхождения между этими функциями на отдельных участках, что говорит о значительных отклонениях эмпирических данных от теоретической модели.

Получившаяся теоретическая функция распределения задаётся формулой (2.2):

$$F(X) = \begin{cases} 0.00, & X \leq 2; \\ 0.159989621, & 2 < X \leq 3; \\ 0.363709738, & 3 < X \leq 4; \\ 0.55826245, & 4 < X \leq 5; \\ 0.706900722, & 5 < X \leq 6; \\ 0.801533755, & X > 6. \end{cases} \quad (2.2)$$

На втором графике проведено сравнение относительных частот для эмпирических (синяя линия) и теоретических (красная линия) данных. Эмпирическая кривая демонстрирует выраженные колебания, в то время как теоретическая кривая выглядит более сглаженной. Максимальные значения относительных частот для эмпирических данных не всегда совпадают с ожидаемыми значениями, а некоторые интервалы демонстрируют заметные расхождения.

Оба графика указывают на то, что эмпирические данные не в полной мере соответствуют теоретическому распределению. Наблюдаемые отклонения не могут быть объяснены только случайными флуктуациями, что подтверждается результатами проверки гипотезы критерием Пирсона.

На рисунке 2.5.2 приведена таблица, которая содержит достаточное количество данных, чтобы судить о виде распределения, опираясь на критерий Пирсона.

| Проверка гипотезы H_0 – распределение Пуассона | | | | | | | | |
|--|-------------|-------------|-------------|-------------|-------------|----------------|------------------------|-----------------------|
| Дискретный ряд распределения частот и частостей | | | | | | Итоговые суммы | χ^2 наблюдаемо | χ^2 табличное |
| X_i | 2 | 3 | 4 | 5 | 6 | - | | |
| N_i^* | 1 | 19 | 20 | 8 | 2 | 50 | 26,21931894 | 7,8 |
| W_i | 0,02 | 0,38 | 0,4 | 0,16 | 0,04 | 1 | | |
| P_i | 0,159989621 | 0,203720117 | 0,194552712 | 0,148638272 | 0,094633033 | 0,801533755 | Уровень | Кол-во ст. |
| Теор_частота N_i | 7,999481044 | 10,18600586 | 9,727635599 | 7,431913598 | 4,731651657 | 40,07668776 | значимости | свободы k |
| $(N_i^* - N_i)^2 / N_i$ | 6,124489153 | 7,626786563 | 10,84759696 | 0,043423831 | 1,577022426 | 26,21931894 | 0,05 | 3 |
| $F(X)$ | 0,159989621 | 0,363709738 | 0,55826245 | 0,706900722 | 0,801533755 | 2,590396286 | | |

Рисунок 2.5.2 – Данные для проверки по критерию Пирсона

В таблице представлена проверка статистической гипотезы H_0 о соответствии наблюдаемых данных распределению Пуассона с использованием критерия согласия Пирсона (χ^2).

Наблюдаемое значение статистики критерия χ^2 равно 26,2193, тогда как табличное критическое значение при уровне значимости 0,05 и числе степеней

свободы 3 составляет 7,8. Поскольку наблюдаемое значение критерия значительно превышает табличное, это говорит о существенных различиях между эмпирическим и теоретическим распределениями.

Такое превышение критического значения означает, что вероятность того, что расхождения между наблюдаемыми и ожидаемыми частотами вызваны случайностью, крайне мала. Следовательно, нулевая гипотеза H_0 о соответствии данных распределению Пуассона должна быть отвергнута.

В итоге можно сделать вывод, что наблюдаемые данные не подчиняются закону распределения Пуассона на заданном уровне значимости.

Также стоит отметить, что контрольное соотношение не выполняется, что было проверено средствами Excel: сумма p_i приблизительно равна 0,8 (рис. 2.5.3).

| | |
|--|------|
| $\left 1 - \sum_{j=1}^M p_i \right \leq 0,01.$ | ложь |
|--|------|

Рисунок 2.5.3 – Контрольное соотношение

2.6 Вывод по второй главе

Во второй главе был проведён детальный статистический анализ дискретных данных, полученных в ходе сбора в течение 50 дней. На основе исходной выборки был построен вариационный ряд, который позволил упорядочить данные и наглядно отразить их распределение. Далее был составлен дискретный ряд распределения, с вычислением всех необходимых точечных оценок числовых характеристик, таких как математическое ожидание, дисперсия и среднее квадратическое отклонение. Эти характеристики позволили получить обобщённое представление о структуре данных и основных тенденциях их распределения.

Для наглядного представления результатов были построены диаграммы эмпирической функции распределения и полигон эмпирических относительных частот. Эмпирическая функция распределения отразила накопленные частоты значений дискретной случайной величины, а полигон относительных частот позволил оценить распределение вероятностей в отдельных дискретных точках.

В результате анализа было выявлено, что эмпирические данные демонстрируют заметные отклонения от теоретической функции распределения. Графики показали, что пики и характерные участки эмпирических данных не полностью совпадают с ожидаемыми теоретическими значениями. Проверка гипотезы с использованием критерия Пирсона окончательно подтвердила, что дискретные эмпирические данные не соответствуют распределению Пуассона.

Таким образом, во второй главе был проведён комплексный анализ дискретных статистических данных, включающий вычисление числовых характеристик, построение графических представлений и статистическую проверку гипотезы. Полученные результаты демонстрируют, что распределение выборки не подчиняется распределению Пуассона, что требует дальнейшего анализа и поиска более подходящей модели для описания дискретных данных.

В приложении В можно увидеть все этапы анализа для дискретных данных средствами Excel.

3 АНАЛИЗ НЕПРЕРЫВНЫХ ДАННЫХ

Для просмотра вариационного ряда непрерывных данных, интервального ряда распределения и вычислений для его числовых характеристик следует обратиться к приложению Б.

3.1 Общие характеристики выборки

На рисунке 3.1.1 показаны данные, которые в целом характеризуют выборку.

| Общие характеристики | |
|----------------------|--------|
| Объём выборки | 50 |
| Кол-во интервалов M | 6 |
| Шаг h | 75 |
| X_min | 3 |
| X_max | 451 |
| \bar{X} | 278.18 |

Рисунок 3.1.1 – Общие характеристики собранных данных

3.2 Оценки числовых характеристик

Размах вариации (R) равен 448, что указывает на широкий разброс значений в выборке и помогает понять диапазон изменения величины. Мода (M_o) составляет 409,9615 и представляет наиболее часто встречающееся значение в выборке. Этот показатель позволяет определить значение, которое имеет наибольшую вероятность появления среди наблюдаемых данных.

Медиана (M_e) равна 312 и делит набор данных на две равные части. Половина значений меньше медианы, а другая половина — больше. Существенное отличие медианы от моды и выборочной средней может указывать на наличие асимметрии в распределении. Среднее линейное отклонение ($Ср_лин_откл\ I$) имеет значение 115,6712 и показывает среднее абсолютное отклонение значений от среднего арифметического. Этот показатель помогает оценить степень разброса данных.

Дисперсия (D) равна 18349,9076 и характеризует средний квадрат отклонений значений от выборочной средней. Высокое значение дисперсии указывает на значительную изменчивость данных относительно среднего значения. Среднеквадратическое отклонение (СКО, σ) составляет 135,4618 и представляет собой корень из дисперсии. Оно показывает, насколько в

среднем значения отклоняются от среднего арифметического и выражается в тех же единицах, что и исходные данные.

Коэффициент вариации (V) равен 0,4869 (48,69%) и демонстрирует относительный разброс данных относительно среднего значения. Высокое значение коэффициента вариации указывает на значительную изменчивость данных. Эксцесс (E_x) принимает значение -0,9045 и характеризует остроту пика распределения. Отрицательное значение указывает на более сглаженный пик распределения по сравнению с нормальным распределением.

Асимметрия составляет -131,7815 и показывает значительное смещение распределения влево. Это означает, что большая часть значений сосредоточена в правой части, а левый хвост распределения вытянут.

Таким образом, числовые характеристики позволяют оценить центр распределения, разброс, асимметрию и остроту пика. Полученные данные демонстрируют значительный разброс значений, смещение распределения влево и сглаженность пика.

Рисунок 3.2.1 отражает точечные оценки числовых характеристик интервального ряда распределения. Формулы для вычисления оценок приведены в теоретической части расчётной работы.

| Числовые характеристики | |
|-------------------------|--------------|
| Размах вариации R | 448 |
| M_o | 409,9615385 |
| M_e | 312 |
| Ср_лин_откл I | 115,6712 |
| Дисперсия D | 18349,9076 |
| СКО σ | 135,4618308 |
| Коэфф_вар V | 0,486957476 |
| Эксцесс E_x | -0,904457966 |
| Асимметрия | -131,7815385 |

Рисунок 3.2.1 – Точечные оценки числовых характеристик

3.3 Вычисление дисперсий

На рисунке 3.3.1 показаны значения дисперсий для интервального ряда распределения. Формулы для вычисления приведены в теоретической части расчётной работы. На рисунках 3.3.2 и 3.3.3 можно увидеть значение эмпирического коэффициента детерминации и проверку средствами Excel истинности теоремы о сложении дисперсий.

| Дисперсии | |
|------------------------------|-------------|
| $(\sigma_{\text{ср_гр}})^2$ | 391,8286765 |
| $(\delta_{\text{межгр}})^2$ | 17958,07892 |
| $(\sigma_{\text{общ}})^2$ | 18349,9076 |

Рисунок 3.3.1 – Оценки дисперсий

| Эмпирический коэффициент детерминации | |
|---------------------------------------|-------------|
| η^2 | 0,978646831 |

Рисунок 3.3.2 – Эмпирический коэффициент детерминации

| Теорема о сложении дисперсий |
|--|
| $(\sigma_{\text{общ}})^2 = (\sigma_{\text{ср_гр}})^2 + (\delta_{\text{межгр}})^2$ |
| ИСТИНА |

Рисунок 3.3.3 – Проверка теоремы о сложении дисперсий

3.4 Диаграммы для непрерывных данных

На рисунке 3.4.1 показан график эмпирической функции распределения, а на рисунке 3.4.2 – гистограмма для относительных частот.

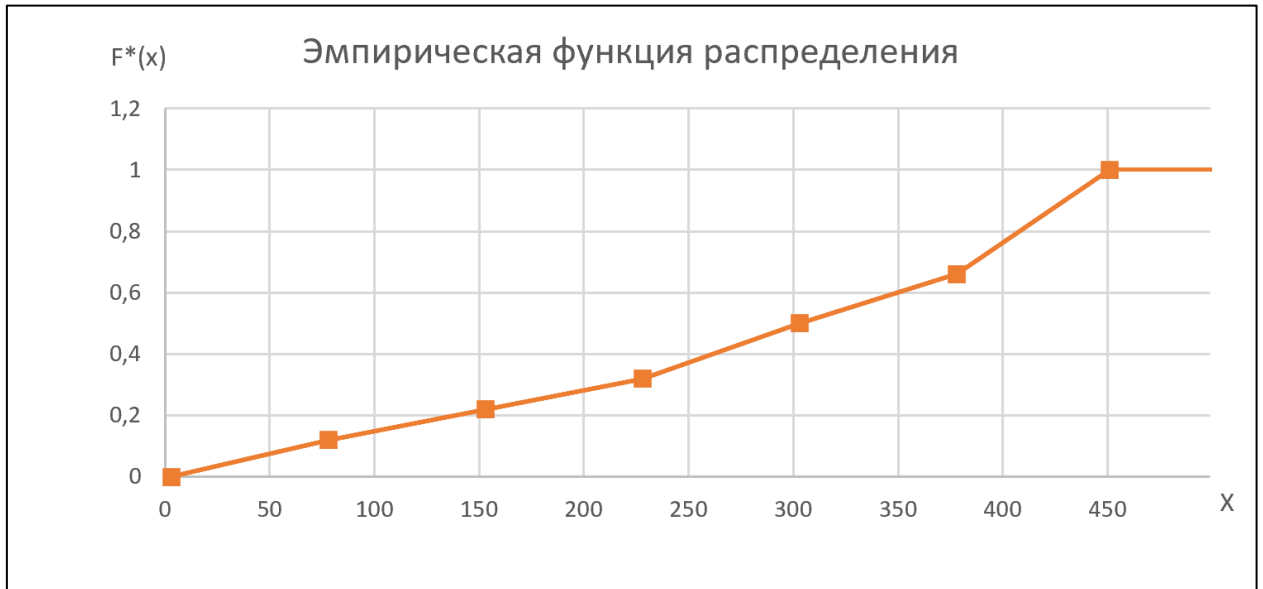


Рисунок 3.4.1 – Эмпирическая функция распределения



Рисунок 3.4.2 – Гистограмма относительных частот

3.5 Проверка гипотезы: нормальное распределение

Будет уместным сначала привести диаграммы сравнения теоретических и эмпирических данных (см. рис 3.5.1, 3.5.2) и проанализировать их.

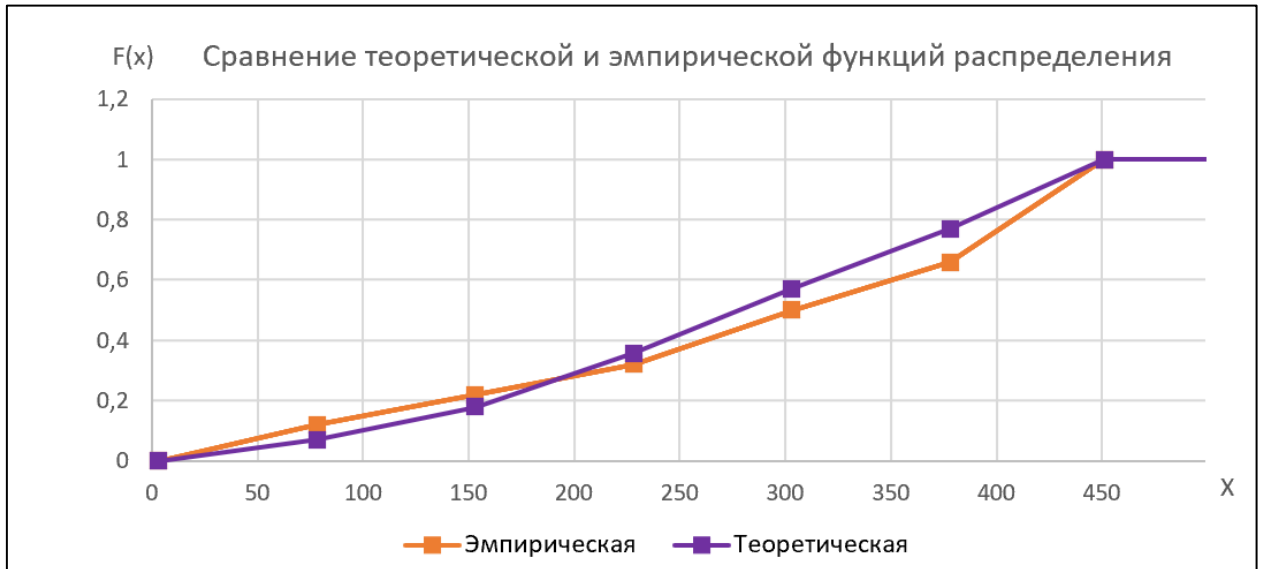


Рисунок 3.5.1 – Сравнение эмпирических функций распределения

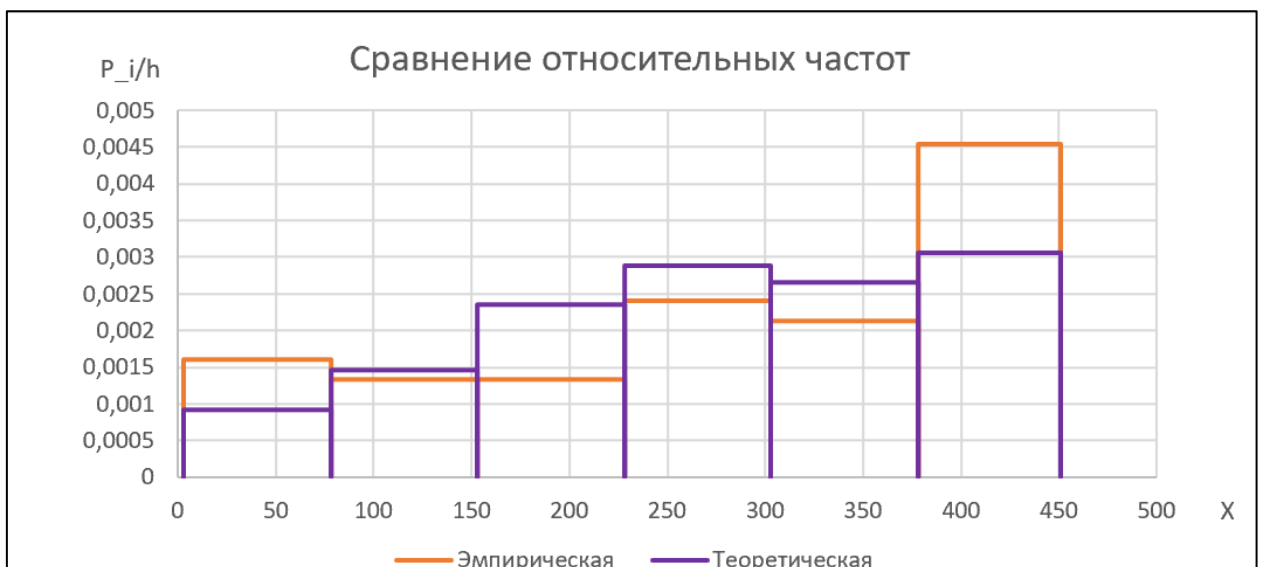


Рисунок 3.5.2 – Гистограммы теоретических и эмпирических относительных частот

Анализ представленных графиков для непрерывных данных позволяет сделать предварительные выводы о характере распределения и степени соответствия теоретическим ожиданиям.

На первой диаграмме представлено сравнение эмпирической и теоретической функций распределения. Видно, что обе кривые в целом следуют схожей тенденции: они монотонно возрастают, что является характерным признаком нормального распределения. Однако на отдельных участках, особенно в области значений около 300–400, наблюдаются заметные отклонения эмпирической функции от теоретической. Тем не менее, общее направление и форма функций позволяют предположить, что распределение данных приближается к нормальному, хотя и с небольшими отклонениями.

На второй диаграмме представлены гистограммы относительных частот для эмпирических и теоретических данных. Гистограммы демонстрируют схожее распределение, однако присутствуют видимые различия в высоте столбцов на отдельных интервалах. В частности, в области 250–350 эмпирические частоты оказываются выше теоретических, а в области 400–450 наблюдается обратная ситуация. Тем не менее, общая форма гистограммы напоминает колоколообразную кривую, что соответствует признакам нормального распределения.

В целом, на основании анализа графиков можно предположить, что распределение данных приближается к нормальному, хотя и содержит некоторые отклонения. Для более точного подтверждения гипотезы о нормальном распределении данных потребуется статистическая проверка с использованием критерия Пирсона.

На рисунке 3.5.3 представлены необходимые данные для проверки гипотезы по критерию Пирсона, а 3.5.2 – результаты.

| Проверка гипотезы H0 - нормальное распределение | | | | | | | |
|--|--------------|--------------|--------------|--------------|-------------|------------|-------|
| Интервал | 3 | 78 | 153 | 228 | 303 | 378 | |
| | 78 | 153 | 228 | 303 | 378 | 451 | |
| N _i | 6 | 5 | 5 | 9 | 8 | 17 | 50 |
| (X _{гран} - \bar{X})/ σ | -∞ | -1,477759446 | -0,924097949 | -0,370436452 | 0,183225045 | 0,73688654 | |
| | -1,477759446 | -0,924097949 | -0,370436452 | 0,183225045 | 0,736886542 | ∞ | |
| Φ(X) | -0,5 | -0,4306 | -0,3212 | -0,1443 | 0,0714 | 0,2703 | |
| | -0,4306 | -0,3212 | -0,1443 | 0,0714 | 0,2703 | 0,5 | |
| P _i | 0,0694 | 0,1094 | 0,1769 | 0,2157 | 0,1989 | 0,2297 | 1 |
| P _i * | 0,12 | 0,1 | 0,1 | 0,18 | 0,16 | 0,34 | |
| (P _i - P _i *) ² /P _i | 0,036892795 | 0,000807678 | 0,033429112 | 0,005908623 | 0,007607893 | 0,05296513 | |
| P _i /h | 0,000925333 | 0,001458667 | 0,002358667 | 0,002876 | 0,002652 | 0,00306267 | |
| | | | | | | | Сумма |

Рисунок 3.5.1 – Данные для проверки

| χ^2 наблюдаемое | χ^2 табличное | Уровень значимости α | Кол-во ст. свободы k |
|----------------------|--------------------|-----------------------------|-------------------------|
| 6,880561553 | 7,8 | 0,05 | 3 |

Рисунок 3.5.2 – Критерий Пирсона

На основе результатов проверки гипотезы о нормальном распределении данных с использованием критерия χ^2 Пирсона можно сделать следующие выводы.

Наблюдаемое значение критерия χ^2 составляет 6,880561553, а табличное значение при уровне значимости $\alpha = 0,05$ и 3 степенях свободы равно 7,8. Согласно правилу проверки гипотезы, если наблюдаемое значение χ^2 меньше табличного, то оснований для отклонения нулевой гипотезы нет.

В данном случае, поскольку $6,88 < 7,8$, нулевая гипотеза о нормальном распределении данных не отвергается. Это означает, что наблюдаемые данные не противоречат предположению о нормальности распределения и могут рассматриваться как соответствующие нормальному закону распределения.

Таким образом, на основе критерия χ^2 Пирсона можно заключить, что гипотеза о нормальном распределении данных подтверждается с уровнем значимости $\alpha = 0,05$.

Также стоит отметить, что выполняется контрольное соотношение, что было проверено средствами Excel: сумма p_i равна 1 (рис. 3.5.3).

| | |
|--|--------|
| $\left 1 - \sum_{j=1}^M p_i \right \leq 0,01.$ | ИСТИНА |
|--|--------|

Рисунок 3.5.3 – Контрольное соотношение

3.6 Вывод по третьей главе

В третьей главе был проведён всесторонний анализ непрерывных данных, полученных в результате исследования. В начале работы был построен интервальный ряд распределения, что позволило наглядно представить структуру и закономерности в данных. На основе ряда были рассчитаны точечные оценки числовых характеристик, таких как математическое ожидание, дисперсия и стандартное отклонение, что обеспечило количественное описание выборки.

Дополнительно была проверена теорема сложения дисперсий, что позволило подтвердить корректность распределения данных и их соответствие основным теоретическим предпосылкам. Важным этапом анализа стало построение вариационного ряда, который обеспечил детальное представление о распределении данных по уровням и дал возможность выявить закономерности в их изменениях.

На заключительном этапе была проверена гипотеза о нормальном распределении данных с использованием критерия χ^2 Пирсона. Результаты проверки показали, что наблюдаемое значение критерия оказалось меньше табличного, что позволило не отвергнуть нулевую гипотезу. Это указывает на то, что анализируемые данные могут рассматриваться как подчиняющиеся нормальному распределению с заданным уровнем значимости.

Таким образом, в ходе анализа непрерывных данных были подтверждены основные теоретические предположения, а также получены все необходимые числовые и графические характеристики выборки, что позволило сделать обоснованные выводы о её распределении и свойствах.

В приложении Г можно увидеть все этапы анализа для непрерывных данных средствами Excel.

ЗАКЛЮЧЕНИЕ

В результате проделанной работы был проведён детальный статистический анализ данных, представленных в дискретной и непрерывной формах. Во второй главе основное внимание было уделено анализу дискретных данных: были построены вариационный и дискретный ряды распределения, рассчитаны точечные оценки числовых характеристик, а также построены диаграммы эмпирической функции распределения и полигон относительных частот. Проверка гипотезы о соответствии данных распределению Пуассона с использованием критерия χ^2 Пирсона показала расхождение между эмпирическими и теоретическими значениями, что привело к отклонению нулевой гипотезы.

В третьей главе был проведён анализ непрерывных данных, который включал построение интервального ряда распределения и вариационного ряда. Для проверки корректности расчетов была использована теорема сложения дисперсий, которая подтвердила их достоверность. Построенные графики эмпирической и теоретической функций распределения, а также гистограммы относительных частот, позволили визуально оценить характер распределения данных и предположить их соответствие нормальному распределению. Проверка гипотезы с использованием критерия χ^2 Пирсона показала, что наблюдаемые данные не противоречат нормальному закону распределения.

Подводя итог, можно отметить, что в ходе исследования были успешно применены методы математической статистики для анализа данных, включая расчёт числовых характеристик, построение графических моделей и проверку гипотез. Результаты работы подтвердили эффективность выбранных подходов и позволили сделать обоснованные выводы о закономерностях в исследуемых выборках. Полученные данные могут служить основой для дальнейших статистических исследований и практического применения.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

[1] Волковец, А. И. Теория вероятностей и математическая статистика : для студентов всех специальностей и форм обучения БГУИР / А. И. Волковец, А. Б. Гуринович. – Мн. : БГУИР, 2003. – 84с. : ил.

[2] Гмурман, В. Е. Теория вероятностей и математическая статистика : учебное пособие для вузов / В. Е. Гмурман. – 9 изд., стер. – М. : Высш. шк., 2003. – 479с. : ил.

ПРИЛОЖЕНИЕ А **(обязательное)** **Собранные данные**

На рисунке А.1 приведены данные, которые накапливались в течение пятидесяти дней для дальнейшего анализа.

| | | | | | |
|--|----|------------------|-----|---------------------|---|
| Примечание. В требованиях сказано: "Обе выборки могут быть получены в результате собственных наблюдений. Я выбрал именно этот вариант" | | | | | |
| Выполнил: Ушаков Александр Дмитриевич, 351001 | | | | | |
| | | Непрерывные | | Дискретные | |
| Дни (в две колонки) | | Время сна (мин.) | | Кол-во приёмов пищи | |
| 1 | 26 | 413 | 407 | 3 | 3 |
| 2 | 27 | 388 | 302 | 4 | 5 |
| 3 | 28 | 239 | 434 | 5 | 4 |
| 4 | 29 | 342 | 263 | 3 | 3 |
| 5 | 30 | 401 | 408 | 3 | 3 |
| 6 | 31 | 421 | 403 | 4 | 6 |
| 7 | 32 | 368 | 213 | 4 | 3 |
| 8 | 33 | 433 | 432 | 5 | 4 |
| 9 | 34 | 412 | 338 | 4 | 4 |
| 10 | 35 | 392 | 387 | 3 | 3 |
| 11 | 36 | 297 | 312 | 6 | 5 |
| 12 | 37 | 428 | 248 | 4 | 3 |
| 13 | 38 | 253 | 23 | 3 | 3 |
| 14 | 39 | 404 | 412 | 3 | 4 |
| 15 | 40 | 361 | 391 | 4 | 3 |
| 16 | 41 | 448 | 428 | 5 | 4 |
| 17 | 42 | 407 | 361 | 4 | 5 |
| 18 | 43 | 294 | 292 | 3 | 4 |
| 19 | 44 | 3 | 419 | 2 | 4 |
| 20 | 45 | 398 | 202 | 4 | 3 |
| 21 | 46 | 443 | 408 | 5 | 3 |
| 22 | 47 | 328 | 389 | 4 | 5 |
| 23 | 48 | 373 | 27 | 4 | 4 |
| 24 | 49 | 429 | 451 | 3 | 3 |
| 25 | 50 | 384 | 278 | 4 | 4 |

Рисунок А.1 – Собранные данные

ПРИЛОЖЕНИЕ Б (обязательное)

Вариационный и интервальный ряды распределения для непрерывных данных

На рисунке Б.2 приведены вариационный ряд непрерывных данных, интервальный ряд распределения и вычисленные параметры для его числовых характеристик

| Вариационный ряд для непрерывных данных | Интервал | N _i | W _i | W _i /h | X _{i_ср_гр} | (X _i - X _{i_ср_гр}) ² | (σ _i) ² - Групповая дисперсия | (X _i - \bar{X}) ² | N _i *(X _{i_ср_гр} - \bar{X}) ² | X _i - \bar{X} |
|---|------------|----------------|----------------|-------------------|----------------------|---|--|--|--|----------------------------|
| 3 | [3; 78] | 6 | 0,12 | 0,0016 | 34,5 | 992,25 | 436,5833333 | 75724,0324 | 356279,6544 | 275,18 |
| 23 | | | | | | 132,25 | | 65116,8324 | | 255,18 |
| 27 | | | | | | 56,25 | | 63091,3924 | | 251,18 |
| 33 | | | | | | 2,25 | | 60113,2324 | | 245,18 |
| 54 | | | | | | 380,25 | | 50256,6724 | | 224,18 |
| 67 | | | | | | 1056,25 | | 44596,9924 | | 211,18 |
| 81 | [78; 153] | 5 | 0,1 | 0,001333333 | 93,8 | 163,84 | 231,76 | 38879,9524 | 169979,922 | 197,18 |
| 84 | | | | | | 96,04 | | 37705,8724 | | 194,18 |
| 87 | | | | | | 46,24 | | 36549,7924 | | 191,18 |
| 94 | | | | | | 0,04 | | 33922,2724 | | 184,18 |
| 123 | | | | | | 852,64 | | 24080,8324 | | 155,18 |
| 186 | | | | | | 158,76 | | 8497,1524 | | 92,18 |
| 194 | [153; 228] | 5 | 0,1 | 0,001333333 | 198,6 | 21,16 | 79,84 | 7086,2724 | 31664,882 | 84,18 |
| 198 | | | | | | 0,36 | | 6428,8324 | | 80,18 |
| 202 | | | | | | 11,56 | | 5803,3924 | | 76,18 |
| 213 | | | | | | 207,36 | | 4248,4324 | | 65,18 |
| 239 | | | | | | 1225 | | 1535,0724 | | 39,18 |
| 248 | | | | | | 676 | | 910,8324 | | 30,18 |
| 253 | [228; 303] | 9 | 0,18 | 0,0024 | 274 | 441 | 501,7777778 | 634,0324 | 157,2516 | 25,18 |
| 263 | | | | | | 121 | | 230,4324 | | 15,18 |
| 278 | | | | | | 16 | | 0,0324 | | 0,18 |
| 292 | | | | | | 324 | | 190,9924 | | 13,82 |
| 294 | | | | | | 400 | | 250,2724 | | 15,82 |
| 297 | | | | | | 529 | | 354,1924 | | 18,82 |
| 302 | | | | | | 784 | | 567,3924 | | 23,82 |
| 312 | | | | | | 1287,015625 | | 1143,7924 | | 33,82 |
| 328 | | | | | | 395,015625 | | 2482,0324 | | 49,82 |
| 338 | | | | | | 97,515625 | | 3578,4324 | | 59,82 |
| 342 | [303; 378] | 8 | 0,16 | 0,002133333 | 347,875 | 34,515625 | 399,359375 | 4072,9924 | 38859,1442 | 63,82 |
| 361 | | | | | | 172,265625 | | 6859,1524 | | 82,82 |
| 361 | | | | | | 172,265625 | | 6859,1524 | | 82,82 |
| 368 | | | | | | 405,015625 | | 8067,6324 | | 89,82 |
| 373 | | | | | | 631,265625 | | 8990,8324 | | 94,82 |
| 384 | | | | | | 741,7612457 | | 11197,8724 | | 105,82 |
| 387 | [378; 451] | 17 | 0,34 | 0,004533333 | 411,2352941 | 587,349481 | 453,1211073 | 11841,7924 | 300963,092 | 108,82 |
| 388 | | | | | | 539,8788927 | | 12060,4324 | | 109,82 |
| 389 | | | | | | 494,4083045 | | 12281,0724 | | 110,82 |
| 392 | | | | | | 369,9965398 | | 12954,9924 | | 113,82 |
| 398 | | | | | | 175,1730104 | | 14356,8324 | | 119,82 |
| 401 | | | | | | 104,7612457 | | 15084,7524 | | 122,82 |
| 404 | | | | | | 52,34948097 | | 15830,6724 | | 125,82 |
| 407 | | | | | | 17,93771626 | | 16594,5924 | | 128,82 |
| 412 | | | | | | 0,584775087 | | 17907,7924 | | 133,82 |
| 413 | | | | | | 3,114186851 | | 18176,4324 | | 134,82 |
| 419 | | | | | | 60,29065744 | | 19830,2724 | | 140,82 |
| 421 | | | | | | 95,34948097 | | 20397,5524 | | 142,82 |
| 434 | | | | | | 518,2318339 | | 24279,8724 | | 155,82 |
| 443 | | | | | | 1008,99654 | | 27165,6324 | | 164,82 |
| 448 | | | | | | 1351,643599 | | 28838,8324 | | 169,82 |
| 451 | | | | | | 1581,231834 | | 29866,7524 | | 172,82 |
| Сумма | | 50 | 1 | | | | | | | |

Рисунок Б.2 – Ряды распределения и параметры для вычисления оценок числовых характеристик

ПРИЛОЖЕНИЕ В **(справочное)** **Анализ дискретных данных в Excel**

На рисунке В.1 можно увидеть все этапы анализа для дискретных данных средствами Excel.



Рисунок В.1 – Дискретные данные в Excel

Анализ непрерывных данных в Excel

На рисунке Г.1 можно увидеть все этапы анализа для дискретных данных средствами Excel.



Рисунок Г.1 – Непрерывные данные в Excel