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 (ТВиМС)

ОТЧЕТ
по расчётной работе по статистике

Тема работы: Анализ дискретных и непрерывных статистических данных

Выполнил
студент: гр. 051003 Феськов С.В.

Проверил: Петюкевич Н.С.

Минск 2021

СОДЕРЖАНИЕ

1 ОБОЗНАЧЕНИЯ И ИСПОЛЬЗУЕМЫЕ ФОРМУЛЫ

$\boldsymbol{x_i}$ – варианта

$\boldsymbol{n_i}$ – частота варианты или интервала

$\boldsymbol{\omega_i}$ – частотость варианты или интервала

$F(x)$ – функция распределения

k – количество вариант

$\bar{x} = \frac{\sum_{i=1}^{k} x_i * n_i}{\sum_{i=1}^{k} n_i}$ – выборочное среднее

$M_0 = x_{\text{н}} + h \frac{n_{\text{мод}} - n_{\text{пред}}}{(n_{\text{мод}} - n_{\text{пред}}) + (n_{\text{мод}} - n_{\text{пост}})}$ – мода (формула только для интервальных рядов)

n_1, n_2, n_3 – частоты предмодального, модального и постмодального интервалов

$x_{\text{н}}$ – значение левой границы модального интервала

$M_e = x_{\text{н}} + h \frac{\frac{\sum_{i=1}^{k} n_i}{2} - \sum_{i=1}^{k_{M_e}} n_i}{n_{M_e}}$ – медиана (формула только для интервальных рядов)

$x_{\text{н}}$ – значение левой границы медианного интервала

$R = x_{max} - x_{min}$ – размах вариации

$l = \frac{\sum_{i=1}^{k} |x_i - \bar{x}| * n_i}{\sum_{i=1}^{k} n_i}$ – среднее линейное отклонение

$\bar{D} = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^2 * n_i}{\sum_{i=1}^{k} n_i}$ – выборочная дисперсия

$\bar{\sigma} = \sqrt{\bar{D}}$ – выборочное среднеквадратичное отклонение

$V = \frac{\bar{\sigma}}{\bar{x}}$ – коэффицент вариации

$\mu_m = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^m * n_i}{\sum_{i=1}^{k} n_i}$ – центральный момент m – го порядка

$S_x = \dfrac{\mu_3}{\sigma^3} -$ асимметрия

$\varepsilon_x = \dfrac{\mu_4}{\sigma^4} - 3 -$ эксцесс

$\sigma_j^2 = \dfrac{\sum_{i=1}^{k_{\text{в группе}}} (x_i - \bar{x}_j)^2 * n_i}{\sum_{i=1}^{k_{\text{в группе}}} n_i} -$ групповая дисперсия

$\bar{\sigma}^2 = \dfrac{\sum_{i=1}^{k_{\text{групп}}} \sigma_i^2 n_i}{\sum_{i=1}^{k_{\text{групп}}} n_i} -$ средняя групповая дисперсия

$\delta^2 = \dfrac{\sum_{i=1}^{k_{\text{групп}}} (\bar{x}_i - \bar{x})^2 * n_i}{\sum_{i=1}^{k_{\text{групп}}} n_i} -$ межгрупповая дисперсия

$p(x) = P(X = x) -$ функция вероятности

$p_i -$ вероятность попасть в интервал i

$\chi^2 = \sum_{i=1}^{k_{\text{интервалов}}} \dfrac{(n_i - np_i)^2}{np_i} -$ критерий согласия Пирсона

$R = \dfrac{|\chi^2 - k|}{\sqrt{2k}} -$ критерий Романовского ($k -$ число степене свободы)

$J = \dfrac{|\chi^2 - k|}{\sqrt{2n + 2.4}} -$ критерий Ястремского ($n -$ число интервалов)

Все расчёты были проведены в Python с использованием библиотек math, pandas и numpy. Графики построены в Python с помощью библиотеки matplotlib. Код исходников вычисления - code.docx, построения графиков - graphic.docx.

2 АНАЛИЗ ДИСКРЕТНЫХ ДАННЫХ

2.1 Описание набора данных

Каждый элемент из набора представляет собой количество потраченных мобильных данных на приложение «VK» в течение дня и округлённое до 50 Мб. Набор данных состоит из 101-го элемента, каждый из которых соответствует одному дню с 1 сентября по 10 декабря. С 24 по 27 сентября данные отсутствуют (фактически данные равны нулю, но они не учитываются, так как в этом промежутке мобильный интернет у меня в принципе отсутствовал), поэтому «очищенный» набор состоит из 97 элементов.

2.2 Статистические законы распределения

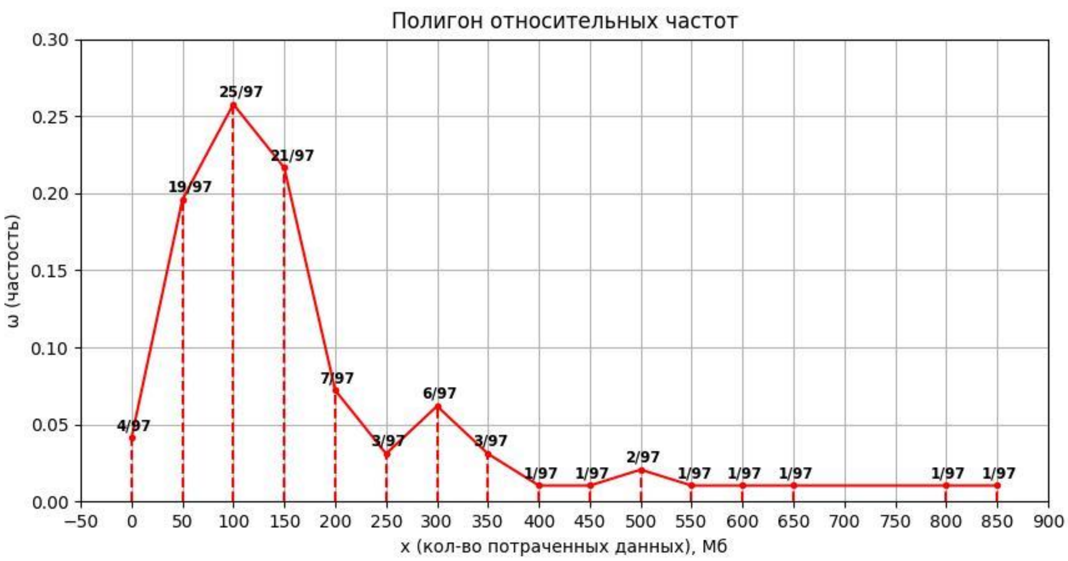
2.2.1 Вариационный ряд:

{0, 0, 0, 0, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 150, 150, 150, 150, 150, 150, 150, 150, 150, 150, 150, 150, 150, 150, 150, 150, 150, 150, 150, 150, 150, 200, 200, 200, 200, 200, 200, 200, 250, 250, 250, 300, 300, 300, 300, 300, 300, 350, 350, 350, 400, 450, 500, 500, 550, 600, 650, 800, 850}

2.2.2 Дискретный ряд распределения частот и частостей

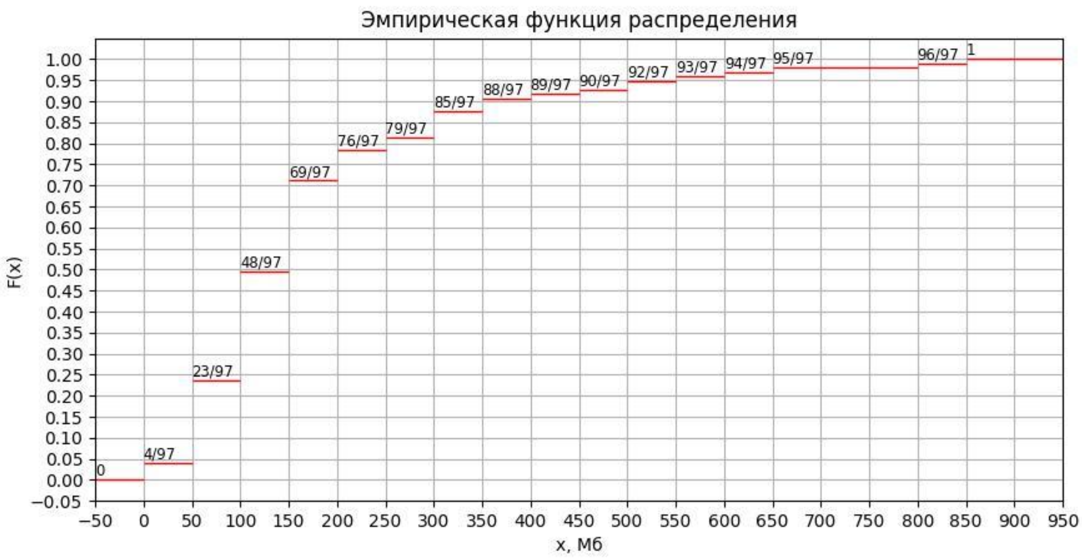
	x1	**x2**	**x3**	**x4**	**x5**	**x6**	**x7**	**x8**	**x9**	**x10**	**x11**	**x12**	**x13**	**x14**	**x15**	**x16**
x_i	0	50	100	150	200	250	300	350	400	450	500	550	600	650	800	850
n_i	4	19	25	21	7	3	6	3	1	1	2	1	1	1	1	1
ω_i	4/97	19/97	25/97	21/97	7/97	3/97	6/97	3/97	1/97	1/97	2/97	1/97	1/97	1/97	1/97	1/97

2.2.3 Полигон относительных частот (частостей)



2.2.4 Эмпирическая функция распределения

$$F(x) = \begin{cases} 0, & x \leq 0 \\ 4/97, & 0 < x \leq 50 \\ 23/97, & 50 < x \leq 100 \\ 48/97, & 100 < x < 150 \\ 69/97, & 150 < x < 200 \\ 76/97, & 200 < x < 250 \\ 79/97, & 250 < x < 300 \\ 85/79, & 300 \leq x < 350 \\ 88/97, & 350 \leq x < 400 \\ 89/97, & 400 \leq x < 450 \\ 90/97, & 450 \leq x < 500 \\ 92/97, & 500 \leq x < 550 \\ 93/97, & 550 \leq x < 600 \\ 94/97, & 600 \leq x < 650 \\ 95/97, & 650 \leq x < 800 \\ 96/97, & 800 \leq x < 850 \\ 1, & x > 850 \end{cases}$$



2.3 Числовые характеристики статистических данных

2.3.1 Выборочное среднее (оценка мат. ожидания)

$$\bar{x} = 174.22680412371133$$

2.3.2 Мода

$$M_0 = 100$$
$$\omega_{M_0} = \frac{25}{97}$$

2.3.3 Медиана

$$M_e = 125$$
$$F(M_e) = \frac{48}{97} \approx \frac{1}{2}$$

2.3.4 Размах вариации

$$R = 850$$

2.3.5 Среднее линейное отклонение

$$l = 111.78658731002233$$

2.3.6 Выборочные дисперсия и среднеквадратичное отклонение

$$\overline{D} = 25882.13412689978$$
$$\bar{\sigma} = 160.87925325193356$$

2.3.7 Асимметрия

$$S_x = 2.116317340851388$$

2.3.8 Эксцесс

$$\varepsilon_x = -1.5603414166897873$$

2.3.9 Коэффициент вариации

$$V = 0.923389796771453$$

2.4 Анализ статистических законов распределения и числовых характеристик

2.4.1 Близость оценок математического ожидания, моды и медианы

Мода и медиана достаточно близки друг к другу, а учитывая то, что медиана в распределении ДСВ рассчитывается неточно (так как функция распределения прерывна и она может не принимать значения ½), можно считать их равными. Однако значение выборочного среднего достаточно далеко от значений моды и медианы, что говорит о наличии некоторой асимметрии в этом распределении, которая к тому же видна на полигоне этих данных.

Проверим соотношение $|M_0 - \bar{x}| \cong 3|M_e - \bar{x}|$:

$$|M_0 - \bar{x}| = 49.226$$

$$3|M_e - \bar{x}| = 74.226$$

Соотношение не выполнено, поэтому данный статистический ряд нельзя назвать умеренно асимметричным.

2.4.2 Близость оценки дисперсии к нулю

Выборочная дисперсия значительно удалена от нуля, что связано с в принципе большими значениями данных в выборке и наличием слишком больших единичных значений данных, сильно влияющих на дисперсию, хоть их частота и минимальна.

2.4.3 Анализ значения коэффициента вариации

Коэффициент вариации V практически равен единице, что говорит о достаточно большом разбросе данных и высокой степени их случайности.

2.4.4 Анализ асимметрии

Асимметрия чуть больше 2, что о говорит о достаточно высокой степени положительной асимметрии, вызванной наличием единичных значений данных справа от мат. ожидания и отсутствием таких единичных данных слева от него.

2.4.5 Анализ эксцесса

Эксцесс приблизительно равен -1.5, что указывает на плосковерхие распределения относительно нормального распределения.

2.4.6 Соотношение между оценкой математического ожидания и дисперсией

Отношение выборочной дисперсии к выборочному мат. ожиданию приблизительно равно 150, что даёт нам основание усомниться в рассмотрении таких теоретических распределений ДСВ как распределение Пуассона (в котором $m = D$) и биномиальное распределение (в котором $D = m * q$, где $0 \leq q \leq 1$).

2.4.7 Общий вывод

Положительной асимметрии, такого большого значения дисперсии и коэффициента вариации и отличия выборочного среднего от моды и медианы можно было избежать, если бы мы убрали из рассмотрения данные со слишком большими значениями, однако тогда бы выборка стала намного меньше.

3 ПРОВЕРКА ГИПОТЕЗЫ О ВИДЕ РАСПРЕДЕЛЕНИЯ ДИСКРЕТНЫХ ДАННЫХ

3.1 Выдвижение гипотезы

Для этих данных, исходя из их анализа, хорошо не подходит ни одно из дискретных теоретических распределений, так как:

1. Отношение дисперсии к мат. ожиданию, равное приблизительно 150, не соответствует теоретическим соотношениям дисперсии и мат. ожидания в распределении Пуассона (D/m = 1) и в биномиальном распределении (D/m = q <= 1);
2. Вид полигона частостей и вид эмпирической функции распределения не соответствует теоретическому виду для геометрического распределения (в котором кривая распределения сразу убывает с 0 на полигоне).

И хоть для геометрического распределения соотношение выборочных дисперсии и мат. ожидания приемлемы, всё-таки графически оно будет меньше сходиться с выборкой чем распределение Пуассона, поэтому я выдвигаю гипотезу о том, что данные имеют распределение Пуассона.

3.2 Расчёт теоретических законов распределения и графический анализ

Так как собранные данные содержатся в диапазоне от 0 до 850 Мб с шагом 50 Мб, а случайная величина в распределение Пуассона – от 0 и далее с шагом 1, то для расчёта законов распределения Пуассона необходимо нормализовать значения случайной величины: поделить на 50. Т. о. мы будем рассчитывать значения законов распределения Пуассона в диапазоне от 0 до 17 с шагом 1 (850/50 = 17)

Вместо исходного выборочного среднего (равного ≈ 174) как оценку мат. ожидания рассмотрим нормализованное выборочное среднее ($\frac{\bar{x}}{50} \approx 3.48$) и нормализованную медиану ($\frac{M_e}{50} = 2.5$).

3.2.1 Функция вероятности

$$p(k) = \frac{e^{-\lambda} * \lambda^k}{k!},$$

где $\lambda = m -$ параметр распределения Пуассона.

3.2.2 Теоретические частоты

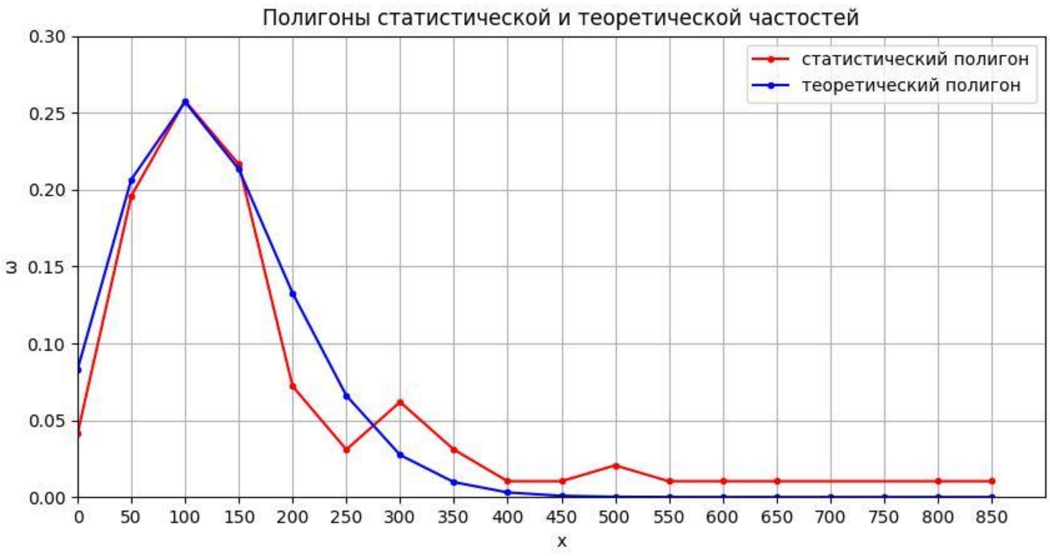
(при m = 2.5)

	x	k	p
x1(k1)	0	0	8.30e-02
x2(k2)	50	1	2.07e-01
x3(k3)	100	2	2.57e-01
x4(k4)	150	3	2.13e-01
x5(k5)	200	4	1.33e-01
x6(k6)	250	5	6.61e-02
x7(k7)	300	6	2.74e-02
x8(k8)	350	7	9.74e-03
x9(k9)	400	8	3.03e-03
x10(k10)	450	9	8.38e-04
x11(k11)	500	10	2.09e-04
x12(k12)	550	11	4.72e-05
x13(k13)	600	12	9.79e-06
x14(k14)	650	13	1.87e-06
x15(k15)	700	14	3.33e-07
x16(k16)	750	15	5.53e-08
x17(k17)	800	16	8.60e-09
x18(k18)	850	17	1.26e-09

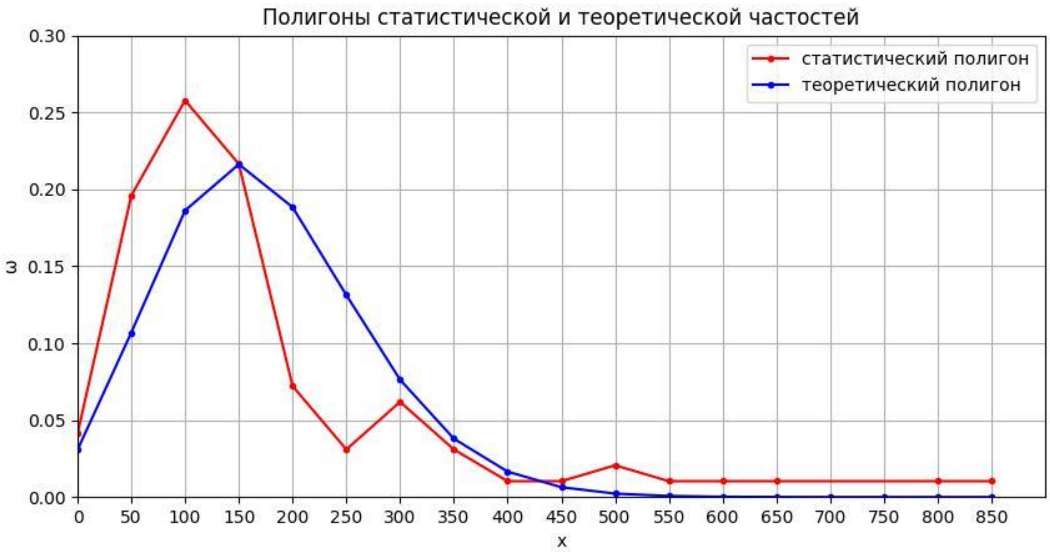
(при m = 3.48)

	x	k	p
x1(k1)	0	0	3.07e-02
x2(k2)	50	1	1.07e-01
x3(k3)	100	2	1.86e-01
x4(k4)	150	3	2.16e-01
x5(k5)	200	4	1.88e-01
x6(k6)	250	5	1.31e-01
x7(k7)	300	6	7.62e-02
x8(k8)	350	7	3.80e-02
x9(k9)	400	8	1.65e-02
x10(k10)	450	9	6.40e-03
x11(k11)	500	10	2.23e-03
x12(k12)	550	11	7.07e-04
x13(k13)	600	12	2.05e-04
x14(k14)	650	13	5.50e-05
x15(k15)	700	14	1.37e-05
x16(k16)	750	15	3.18e-06
x17(k17)	800	16	6.92e-07
x18(k18)	850	17	1.42e-07

3.2.3 Полигоны статистической и теоретической частостей

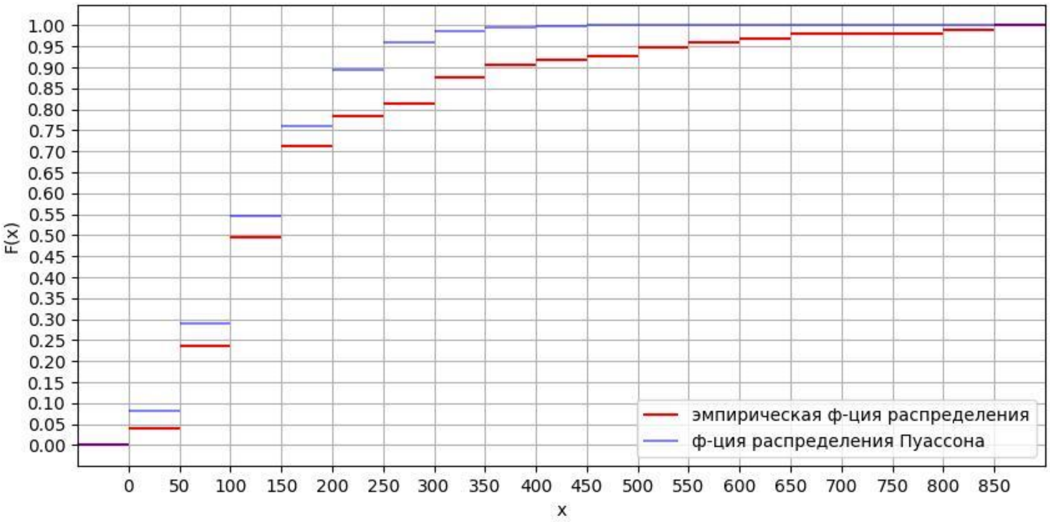


(при m = 2.5)

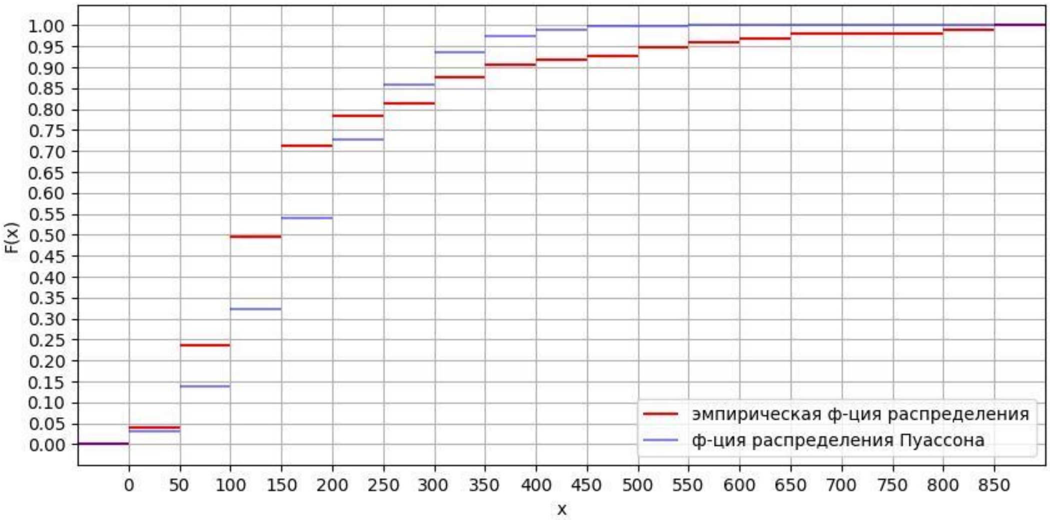


(при m = 3.48)

3.2.4 Графики функций распределения



(при m = 2.5)



(при m = 3.48)

3.2.5 Анализ графиков

Как мы видим по полигонам, и при m = 3.48, и при m = 2.5, распределение Пуассона соответствует статистике, однако при m = 2.5 соответствие практически полное. На графиках функций распределения расхождения есть, но общая статистическая тенденция совпадает с теоретическим распределением.

3.2.6 Проверка $\left|1-\sum_{j=1}^{M} p_i\right| \le 0,01$.

$\left|1-\sum_{j=1}^{M} p_i\right|$ = 2.0e-10 << 0,01

3.3 Проверка гипотезы о виде распределения

При расчёте критериев вариационный ряд был разбит на интервалы так, чтобы в каждом интервале частота была не меньше 5:

	i1	i2	i3	i4	i5	i6	i7
$\boldsymbol{i_i}$	0-50	100	150	200	250-300	350-450	500-850
$\boldsymbol{n_i}$	23	25	21	7	9	5	7
$\boldsymbol{p_i}$ **(m=2.5)**	2.90e-01	2.57e-01	2.13e-01	1.33e-01	9.35e-02	1.36e-02	2.68e-04
$\boldsymbol{p_i}$ **(m=3.48)**	1.38e-01	1.86e-01	2.16e-01	1.88e-01	2.08e-01	6.09e-02	3.21e-03

(здесь $\boldsymbol{i_i}$ – интервал с номером i)

3.3.1 По критерию согласия Пирсона

При $m = 2.5$:

$$\chi^2 = 1885.096$$

При $m = 3.48$:

$$\chi^2 = 166.359$$

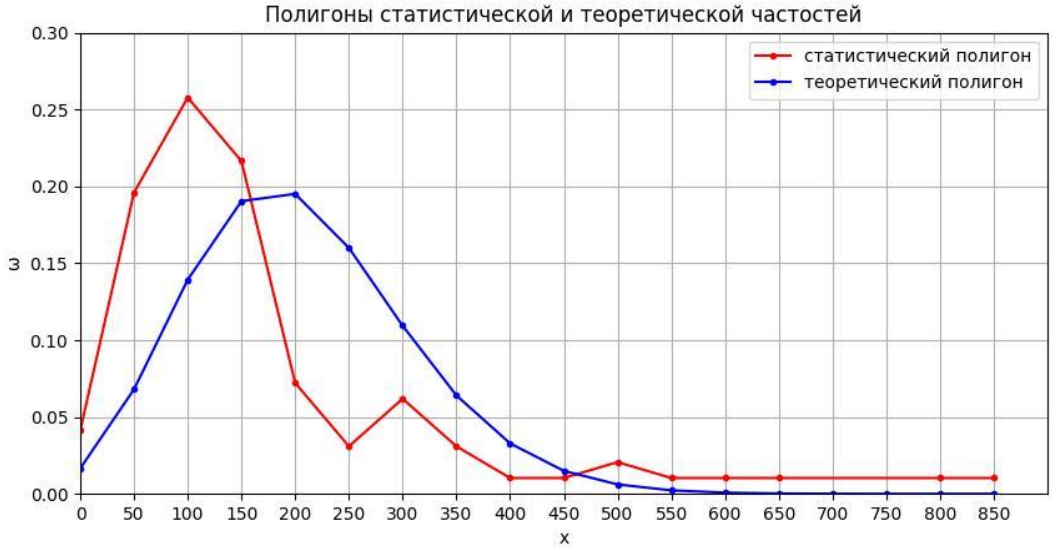
Как мы видим, более подходящее графически распределение Пуассона с m = 2.5, намного меньше подходит с точки зрения критерия Пирсона, что связано, по моему мнению, с недостаточно большой выборкой (так как в нашей выборке значения, вероятность которых теоретически должна быть близка к нулю, в выборке имеют достаточно большой вес).

Однако и при m = 3.48 критерий плох, так как его критическое значение при $\alpha = 0.05$ и $k = 7 - 1 - 1 = 5$ равно $\chi^2_{0.05,5} = 11.07$, что намного меньше, чем рассчитанные значения.

Но видна закономерность, что при увеличении m, критерий Пирсона уменьшается, поэтому можно попробовать подобрать наиболее подходящий m:

При m = 4 $\chi^2 = 100.221$;
При m = 4.1 $\chi^2 = 98.609$;
При m = 4.2 $\chi^2 = 99.338$.

Как мы видим, лучшее значение критерия получается при m = 4.1, однако график при таком m выглядит следующим образом:



Т. о. при всех возможных m данную гипотезу следует отвергнуть.

3.3.2 По критерию согласия Романовского

При $m = 2.5$:

$$R = 594.538$$

При $m = 3.48$:

$$R = 51.026$$

При $m = 4.1$:

$$R = 29.601$$

Как мы видим, ни при каком m полученный критерий R не меньше 3, значит во всех случаях данную гипотезу следует отвергнуть.

3.3.3 По критерию согласия Ястремского

При $m = 2.5$:

$$J = 464.256$$

При $m = 3.48$:

$$J = 39.844$$

При $m = 4.1$:

$$J = 23.115$$

Как мы видим, ни при каком m полученный критерий R не меньше 3, значит во всех случаях данную гипотезу следует отвергнуть.

4 АНАЛИЗ НЕПРЕРЫВНЫХ ДАННЫХ

4.1 Описание набора данных

Каждый элемент из набора представляет собой курс биткойна к доллару в определённый день. Набор данных состоит из 101-го элемента, каждый из которых соответствует одному дню с 1 сентября по 10 декабря. Данные были взяты с сайта https://www.rbc.ru/crypto/currency/btcusd.

4.1 Статистические законы распределения

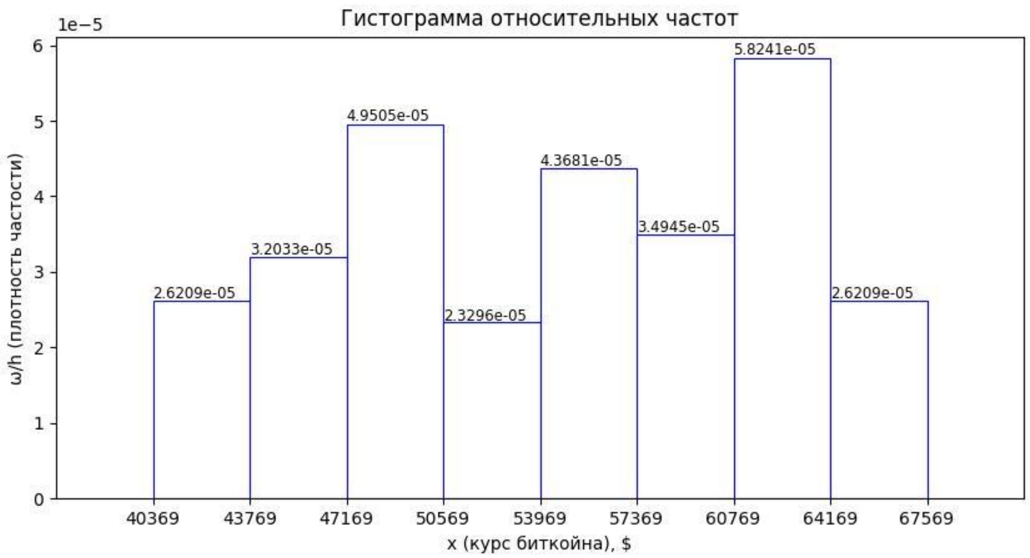
4.1.1 Вариационный ряд:

{40369.0, 40950.0, 41124.4, 42161.4, 42560.0, 42721.85, 42750.75, 43150.4, 43621.55, 43873.15, 44770.25, 44864.15, 44959.0, 45103.7, 46042.55, 46067.75, 46358.5, 46549.85, 47005.15, 47079.85, 47249.15, 47300.0, 47550.0, 47733.9, 47815.75, 48076.45, 48138.9, 48145.75, 48241.95, 48900.0, 49096.6, 49108.25, 49350.0, 49354.75, 49879.1, 49909.07, 50554.1, 50589.6, 50600.0, 51615.75, 51746.35, 52728.9, 53560.0, 53790.9, 53810.05, 53991.0, 54400.0, 54669.8, 55034.3, 55498.6, 56120.8, 56321.2, 56838.55, 56922.75, 57049.25, 57101.5, 57216.75, 57225.65, 57232.15, 57285.0, 57543.25, 57689.95, 57800.0, 58116.05, 58538.55, 58600.0, 58965.75, 59849.0, 60250.0, 60383.55, 60389.5, 60565.55, 60829.35, 60860.5, 60872.45, 61088.35, 61100.75, 61255.85, 61296.3, 61376.8, 61516.1, 61592.7, 61700.0, 61820.0, 62021.1, 62090.45, 62255.75, 62979.35, 63000.0, 63215.25, 63250.0, 63500.0, 64250.0, 64269.15, 64489.75, 64828.85, 64940.05, 65434.8, 66016.9, 66897.25, 67530.7}

4.1.2 Интервальный ряд распределения частот и частостей

	a1-a2	**a2-a3**	**a3-a4**	**a4-a5**	**a5-a6**	**a6-a7**	**a7-a8**	**a8-a9**
$a_i - a_{i+1}$	40369-43769	43769-47169	47169-50569	50569-53969	53969-57369	57369-60769	60769-64169	64169-67569
n_i	9	11	17	8	15	12	20	9
ω_i	9/101	11/101	17/101	8/101	15/101	12/101	20/101	9/101

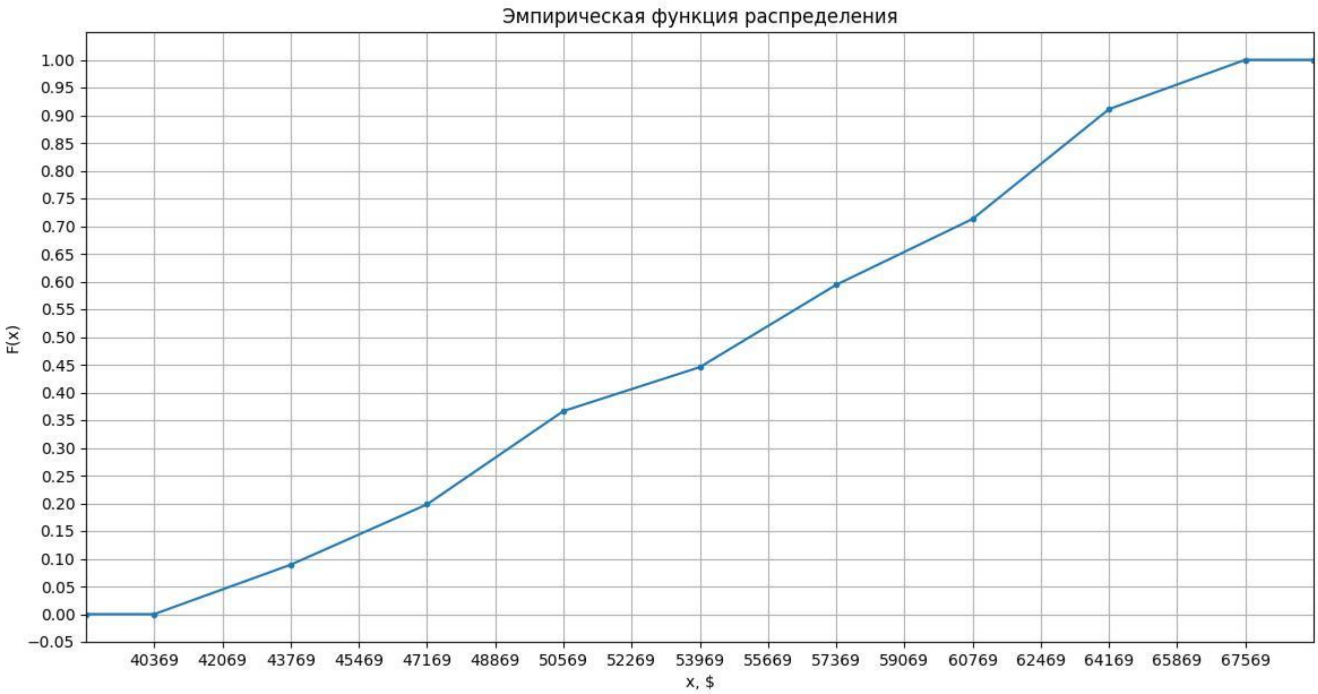
4.1.3 Гистограмма относительных частот (частостей)



4.1.4 Эмпирическая функция распределения

$$F(x)=\begin{cases}0, & x\le 40369\\ \sum_{i=1}^{k-1}\omega_i, & 40369<x\le 65869\\ 1, & x>65869\end{cases}$$

$$(k - \text{номер интервала, в котором находится } x)$$



4.2 Числовые характеристики статистических данных

4.2.1 Выборочное среднее (оценка мат. ожидания)

$$\bar{x} = 54503.02445544554$$

4.2.2 Мода

$$i = 7 \text{ (номер модального интервала)}$$
$$M_0 = 62200.57894736842$$
$$\omega_{M_0} = \frac{20}{101}$$

4.2.3 Медиана

$$i = 5 \text{ (номер медианного интервала)}$$
$$M_e = 55215.67$$

4.2.4 Размах вариации

$$R = 27161.7$$

4.2.5 Среднее линейное отклонение

$$l = 6461.834245662185$$

4.2.6 Выборочные дисперсия и среднеквадратичное отклонение

$$\overline{D} = 53591626.78816333$$
$$\bar{\sigma} = 7320.630217963705$$

4.2.7 Асимметрия

$$S_x = -0.17976445456751847$$

4.2.8 Эксцесс

$$\varepsilon_x = -1.2039011647433924$$

4.2.9 Коэффициент вариации

$$V = 0.13431603642377835$$

4.3 Анализ статистических законов распределения и числовых характеристик

4.3.1 Близость оценок математического ожидания, моды и медианы

Выборочное среднее и медиана практически равны и находятся в центре диапазона значений вариационного ряда, но они сильно отличаются от моды. Это говорит о том, что ряд не обладает абсолютной симметрией, однако ряд вполне может быть приближённо симметричным. Данные соотношения между выборочным средним, модой и медианой наводят на мысль о том, что этому вариационному ряду может соответствовать равномерное распределение и скорее всего не соответствует нормальное и экспоненциальное распределения.

4.3.2 Близость оценки дисперсии к нулю

Выборочная дисперсия значительно удалена от нуля, что связано с большими значениями данных в выборке.

4.3.3 Анализ значения коэффициента вариации

Коэффициент вариации V практически равен 0.13, что говорит о возможном маленьком разбросе случайной величины относительно мат. ожидания.

4.3.4 Анализ асимметрии

Асимметрия примерно равна -0.2, значит ряд достаточно симметричен (обладает небольшой отрицательной асимметрией).

4.3.5 Анализ эксцесса

Эксцесс приблизительно равен -1.2, что указывает на плосковерхие распределения относительно нормального распределения.

4.4 Анализ использования правила сложения дисперсий

4.4.1 Результаты расчётов

Интервалы	Групповые средние	Групповые дисперсии
40369-43769	42156,59444	1076203,054
43769-47169	45697,62727	983718,5774
47169-50569	48611,98353	923331,6152
50569-53969	52305,19375	1608070,971
53969-57369	56193,82	1274982,618
57369-60769	59057,59583	1247765,802
60769-64169	61881,0525	728965,6959
64169-67569	65406,38333	1236743,677

$$\bar{\sigma}^2 = 1077978.9402740872$$
$$\delta^2 = 52513647.84788923$$
$$\eta^2 = 0.97988$$

4.4.2 Анализ результатов

$$\bar{\sigma}^2 + \delta^2 = 53591626.78816$$
$$\overline{D} = 53591626.78816$$

Соотношение $\bar{\sigma}^2 + \delta^2 = \overline{D}$ выполняется, а значит теорема сложения дисперсий верна.

Эмпирический коэффициент детерминации практически равен 1, значит доля значимости группировки максимальна, то есть выполненная группировка является хорошей.

5 ПРОВЕРКА ГИПОТЕЗЫ О ВИДЕ РАСПРЕДЕЛЕНИЯ НЕПРЕРЫВНЫХ ДАННЫХ

5.1 Выдвижение гипотезы

Учитывая то, что интервальный ряд имеет несколько максимумов (что видно из графика гистограммы), а медиана и выборочное среднее совпадают и находятся в центре диапазона значений данных (но при этом мода находится в отдалении от медианы и выборочного среднего), можно достаточно точно отбросить из рассмотрения нормальное и показательное распределение и выдвинуть гипотезу о том, что данные имеют равномерное распределение.

5.2 Расчёт теоретических законов распределения и графический анализ

Параметры равномерного распределения a и b рассчитаем как границы диапазона значений данных, то есть:

$$a = 40369$$
$$b = 67530.7$$

5.2.1 Плотность распределения

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b \\ 0, & x < a \cup x > b \end{cases}$$

5.2.2 Функция распределения

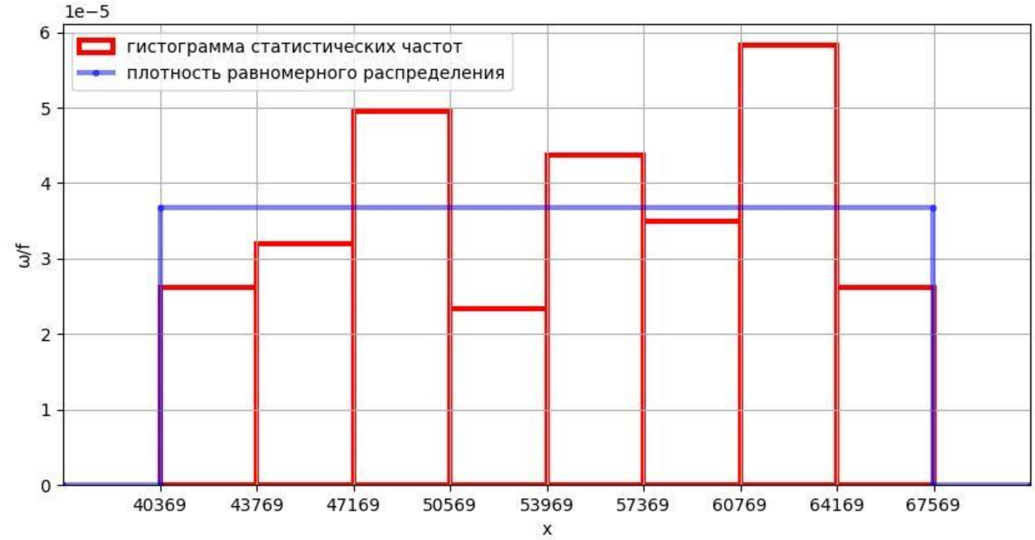
$$F(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \le x \le b \\ 1, & x > b \end{cases}$$

5.2.3 Теоретические частоты

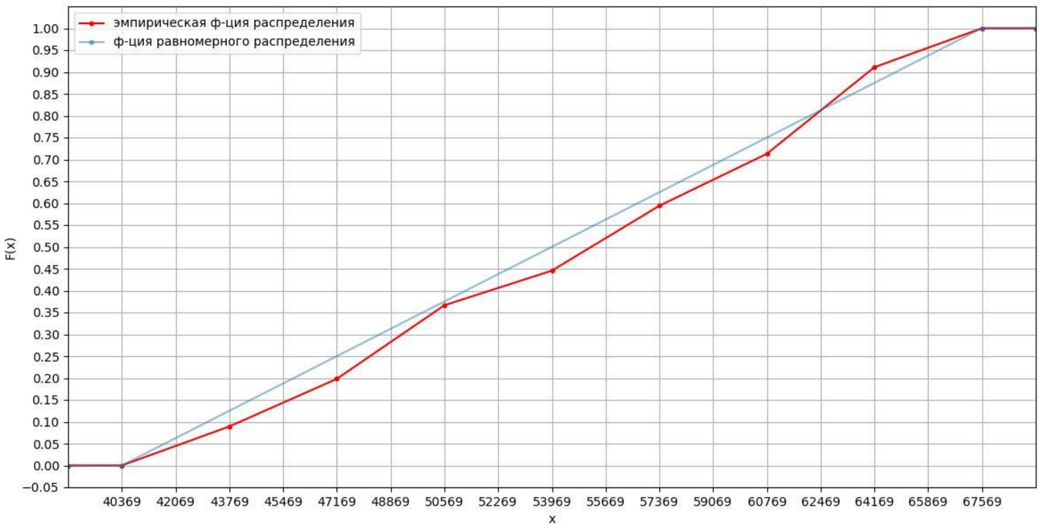
Вероятность попасть в интервал $a_i - a_{i+1}$будем рассчитывать по формуле $p_i = F(a_{i+1}) - F(a_i)$.

	$a_i - a_{i+1}$	n_i	ω_i	p_i
a0-a1	40369.0-43769.0	9	9/101	1.25e-01
a1-a2	43769.0-47169.0	11	11/101	1.25e-01
a2-a3	47169.0-50569.0	17	17/101	1.25e-01
a3-a4	50569.0-53969.0	8	8/101	1.25e-01
a4-a5	53969.0-57369.0	15	15/101	1.25e-01
a5-a6	57369.0-60769.0	12	12/101	1.25e-01
a6-a7	60769.0-64169.0	20	20/101	1.25e-01
a7-a8	64169.0-67569.0	9	9/101	1.24e-01

5.2.4 Гистограмма статистических частостей и плотность равномерного распределения



5.2.5 Графики функций распределения



5.2.6 Анализ графиков

Гистограмма и функция распределения достаточно сильно отличаются, хоть функция распределения и описывает общую тенденцию распределения как у гистограммы. А вот функции распределения достаточно похожи

5.2.7 Проверка $\left|1-\sum_{j=1}^{M} p_i\right| \le 0{,}01$.

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

5.3 Проверка гипотезы о виде распределения

При расчёте χ^2 будем использовать существующие интервалы, так как они все удовлетворяют необходимым условиям (частота для каждого интервала больше 4).

$$\chi^2 = 10.203$$

5.3.1 По критерию согласия Пирсона

$$\alpha = 0.05$$
$$k = 8 - 1 - 2 = 5$$

Для данных k и a критическое значение критерия согласия Пирсона равно $\chi^2_{0.05,5} = 11.07$.

Как видно, $\chi^2 < \chi^2_{0.05,5}$, значит гипотеза о равномерном виде распределения принимается.

5.3.2 По критерию согласия Романовского

$$R = 1.645$$

R < 3, значит гипотеза принимается.

5.3.3 По критерию согласия Ястремского

$$J = 1.213$$

J < 3, значит гипотеза принимается.