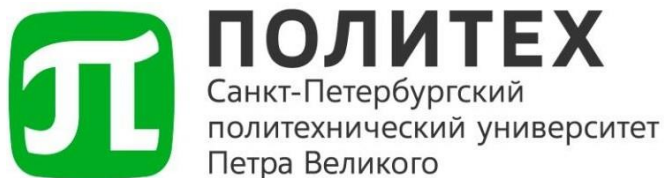


ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ПЕТРА ВЕЛИКОГО»
Высшая школа программной инженерии



ОТЧЕТ ПО РАССЧЕТНОМУ ЗАДАНИЮ
ВАРИАНТ №178
по дисциплине «Теория вероятностей и математическая статистика»

Студенты

группы № 3530202/00201

Козлова Е. А.

Руководитель

Зайцев И. В.

Санкт-Петербург

2022 г.

ОГЛАВЛЕНИЕ

ТАБЛИЦА НУМЕРОВАННЫХ ДЕНОРМАЛИЗОВАННЫХ ЧИСЕЛ	3
ПОСТРОЕНИЕ НУМЕРОВАННОГО ВАРИАЦИОННОГО РЯДА	4
ГРУППИРОВАНИЕ ПО ИНТЕРВАЛАМ, ТАБЛИЦА РАЗБИЕНИЯ НА ИНТЕРВАЛЫ.....	6
ПОСТРОЕНИЕ ПОЛИГОНА, ГИСТОГРАММЫ, СТУПЕНЧАТОЙ КРИВОЙ.....	7
ТОЧЕЧНАЯ ОЦЕНКА ХАРАКТЕРИСТИК РАСПРЕДЕЛЕНИЯ ПО СГРУППИРОВАННЫМ ДАННЫМ	9
ОЦЕНКА ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ МЕТОДОМ КВАНТИЛЕЙ.....	11
ПОСТРОЕНИЕ ДОВЕРИТЕЛЬНЫХ ИНТЕРВАЛОВ	12
ПРИБЛИЖЕННЫЙ МЕТОД ПРОВЕРКИ ГИПОТЕЗЫ О НОРМАЛЬНОМ РАСПРЕДЕЛЕНИИ	15
ПРОВЕРКА ГИПОТЕЗЫ ОБ ОДНОРОДНОСТИ ВЫБОРКИ С ПОМОЩЬЮ КРИТЕРИЯ ЗНАКОВ И КРИТЕРИЯ ВИЛКОКСОНА	20

ТАБЛИЦА НУМЕРОВАННЫХ ДЕНОРМАЛИЗОВАННЫХ ЧИСЕЛ

№ изм.	xi	№ изм.	xi	№ изм.	xi	№ изм.	xi	№ изм.	xi
1	441,0	41	453,0	81	442,0	121	425,0	161	460,0
2	423,0	42	428,0	82	437,0	122	446,0	162	425,0
3	456,0	43	450,0	83	457,0	123	456,0	163	426,0
4	421,0	44	428,0	84	432,0	124	427,0	164	432,0
5	451,0	45	440,0	85	430,0	125	419,0	165	436,0
6	435,0	46	418,0	86	433,0	126	414,0	166	446,0
7	453,0	47	428,0	87	422,0	127	452,0	167	409,0
8	438,0	48	439,0	88	440,0	128	454,0	168	460,0
9	428,0	49	445,0	89	433,0	129	435,0	169	431,0
10	422,0	50	468,0	90	462,0	130	455,0	170	426,0
11	445,0	51	441,0	91	450,0	131	437,0	171	458,0
12	426,0	52	442,0	92	437,0	132	450,0	172	418,0
13	457,0	53	435,0	93	444,0	133	457,0	173	430,0
14	426,0	54	427,0	94	435,0	134	438,0	174	442,0
15	433,0	55	437,0	95	441,0	135	467,0	175	441,0
16	440,0	56	447,0	96	439,0	136	432,0	176	438,0
17	436,0	57	423,0	97	434,0	137	451,0	177	448,0
18	431,0	58	439,0	98	447,0	138	436,0	178	433,0
19	443,0	59	445,0	99	410,0	139	430,0	179	452,0
20	414,0	60	446,0	100	441,0	140	447,0	180	434,0
21	422,0	61	427,0	101	438,0	141	447,0	181	444,0
22	411,0	62	428,0	102	443,0	142	443,0	182	449,0
23	436,0	63	455,0	103	435,0	143	434,0	183	427,0
24	449,0	64	413,0	104	439,0	144	429,0	184	422,0
25	457,0	65	422,0	105	429,0	145	444,0	185	420,0
26	433,0	66	451,0	106	439,0	146	430,0	186	432,0
27	441,0	67	453,0	107	433,0	147	431,0	187	425,0
28	442,0	68	446,0	108	437,0	148	428,0	188	426,0
29	430,0	69	443,0	109	444,0	149	443,0	189	428,0
30	443,0	70	434,0	110	422,0	150	418,0	190	441,0
31	445,0	71	446,0	111	455,0	151	448,0	191	440,0
32	435,0	72	437,0	112	450,0	152	443,0	192	405,0
33	433,0	73	467,0	113	424,0	153	427,0	193	434,0
34	428,0	74	410,0	114	434,0	154	440,0	194	444,0
35	456,0	75	442,0	115	450,0	155	425,0	195	431,0
36	451,0	76	443,0	116	428,0	156	428,0	196	435,0
37	443,0	77	452,0	117	431,0	157	426,0	197	431,0
38	441,0	78	450,0	118	439,0	158	445,0	198	433,0
39	437,0	79	436,0	119	445,0	159	451,0	199	418,0
40	430,0	80	456,0	120	445,0	160	423,0	200	439,0

Таблица 1. Таблица нумерованных денормализованных чисел

ПОСТРОЕНИЕ НУМЕРОВАННОГО ВАРИАЦИОННОГО РЯДА

Совокупность значений признака, записанных в порядке их возрастания, называют вариационным рядом (упорядоченной выборкой), а сам признак - вариантой (случайной величиной). Вариационный ряд, построенный по данным табл. 1, приведен в табл. 2

№ изм.	X_i	№ изм.	X_i	№ изм.	X_i	№ изм.	X_i	№ изм.	X_i
1	405,0	41	428,0	81	434,0	121	441,0	161	448,0
2	409,0	42	428,0	82	434,0	122	441,0	162	448,0
3	410,0	43	428,0	83	435,0	123	441,0	163	449,0
4	410,0	44	428,0	84	435,0	124	441,0	164	449,0
5	411,0	45	428,0	85	435,0	125	441,0	165	450,0
6	413,0	46	428,0	86	435,0	126	442,0	166	450,0
7	414,0	47	428,0	87	435,0	127	442,0	167	450,0
8	414,0	48	428,0	88	435,0	128	442,0	168	450,0
9	418,0	49	428,0	89	435,0	129	442,0	169	450,0
10	418,0	50	428,0	90	436,0	130	442,0	170	450,0
11	418,0	51	429,0	91	436,0	131	443,0	171	451,0
12	418,0	52	429,0	92	436,0	132	443,0	172	451,0
13	419,0	53	430,0	93	436,0	133	443,0	173	451,0
14	420,0	54	430,0	94	436,0	134	443,0	174	451,0
15	421,0	55	430,0	95	437,0	135	443,0	175	451,0
16	422,0	56	430,0	96	437,0	136	443,0	176	452,0
17	422,0	57	430,0	97	437,0	137	443,0	177	452,0
18	422,0	58	430,0	98	437,0	138	443,0	178	452,0
19	422,0	59	431,0	99	437,0	139	443,0	179	453,0
20	422,0	60	431,0	100	437,0	140	444,0	180	453,0
21	422,0	61	431,0	101	437,0	141	444,0	181	453,0
22	423,0	62	431,0	102	438,0	142	444,0	182	454,0
23	423,0	63	431,0	103	438,0	143	444,0	183	455,0
24	423,0	64	431,0	104	438,0	144	444,0	184	455,0
25	424,0	65	432,0	105	438,0	145	445,0	185	455,0
26	425,0	66	432,0	106	439,0	146	445,0	186	456,0
27	425,0	67	432,0	107	439,0	147	445,0	187	456,0
28	425,0	68	432,0	108	439,0	148	445,0	188	456,0
29	425,0	69	433,0	109	439,0	149	445,0	189	456,0
30	426,0	70	433,0	110	439,0	150	445,0	190	457,0
31	426,0	71	433,0	111	439,0	151	445,0	191	457,0
32	426,0	72	433,0	112	439,0	152	446,0	192	457,0
33	426,0	73	433,0	113	440,0	153	446,0	193	457,0
34	426,0	74	433,0	114	440,0	154	446,0	194	458,0
35	426,0	75	433,0	115	440,0	155	446,0	195	460,0
36	427,0	76	433,0	116	440,0	156	446,0	196	460,0
37	427,0	77	434,0	117	440,0	157	447,0	197	462,0
38	427,0	78	434,0	118	441,0	158	447,0	198	467,0
39	427,0	79	434,0	119	441,0	159	447,0	199	467,0
40	427,0	80	434,0	120	441,0	160	447,0	200	468,0

Таблица 2. Нумерованный вариационный ряд

ОЦЕНКА МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ, ДИСПЕРСИИ (СМЕЩЕННОЙ, НЕСМЕЩЕННОЙ), МЕДИАНЫ

Оценка математического ожидания – среднее арифметическое – вычисляется по формуле:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$
$$\bar{x} = 437,49$$

Смещенную оценку дисперсии по всей выборке вычисляем по формуле:

$$s^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2$$
$$s^2 = 463,26$$
$$s = 21,523$$

Несмещенную оценку дисперсии вычисляют по формуле:

$$s^{*2} = \frac{n * s^2}{n - 1}$$
$$s^{*2} = 465,588$$
$$s^* = 21,577$$

Оценка медианы – значение варианты, которое делит вариационный ряд на две равные по числу членов части. При четном числе членов ($n = 2k$) в качестве медианы принимают:

$$\tilde{Me} = \frac{x_k + x_{k+1}}{2} = 437,00$$

Размахом варьирования (широтой распределения) называют разность между наибольшим и наименьшим значениями варианты:

$$R = x_{max} - x_{min} = 63,00$$

ГРУППИРОВАНИЕ ПО ИНТЕРВАЛАМ, ТАБЛИЦА РАЗБИЕНИЯ НА ИНТЕРВАЛЫ

При большом объеме выборки для удобства вычислений прибегают к группированию данных в интервалы. Число таких интервалов при объеме выборки, превышающем 100 – 300 элементов, рекомендуется брать в пределах от 10 до 20. Возьмем число интервалов $l = 11$. Тогда ширина интервала (шаг разбиения) будет равна:

$$\Delta x = \frac{R}{l} = 5,3$$

Таблицу подсчета частот и частотностей по интервалам вариационного ряда удобно представить в виде табл. 3.

№ интервала	Границы интервалов	Частота в интервале	Частость в интервале	Середина интервала
1	405-411	5	0,025	408
2	411-417	3	0,015	414
3	417-423	16	0,08	420
4	423-429	28	0,14	426
5	429-435	37	0,185	432
6	435-441	36	0,18	438
7	441-447	35	0,175	444
8	447-453	21	0,105	450
9	453-459	13	0,065	456
10	459-465	3	0,015	462
11	465-471	2	0,015	468

Таблица 3. Таблица подсчета

ПОСТРОЕНИЕ ПОЛИГОНА, ГИСТОГРАММЫ, СТУПЕНЧАТОЙ КРИВОЙ

На оси абсцисс откладываются интервалы значений величины x , в серединах интервалов строятся ординаты, пропорциональные частотам, и концы ординат соединяются отрезками прямых линий. На рис. 1 показан полигон распределения, построенный по данным табл. 3.



Рисунок 1. Полигон распределения

Над каждым отрезком оси абсцисс, изображающим интервал значений x , строится прямоугольник, площадь которого пропорциональна частоте в данном интервале. На рис.2 показана гистограмма распределения, построенная по данным табл. 3.

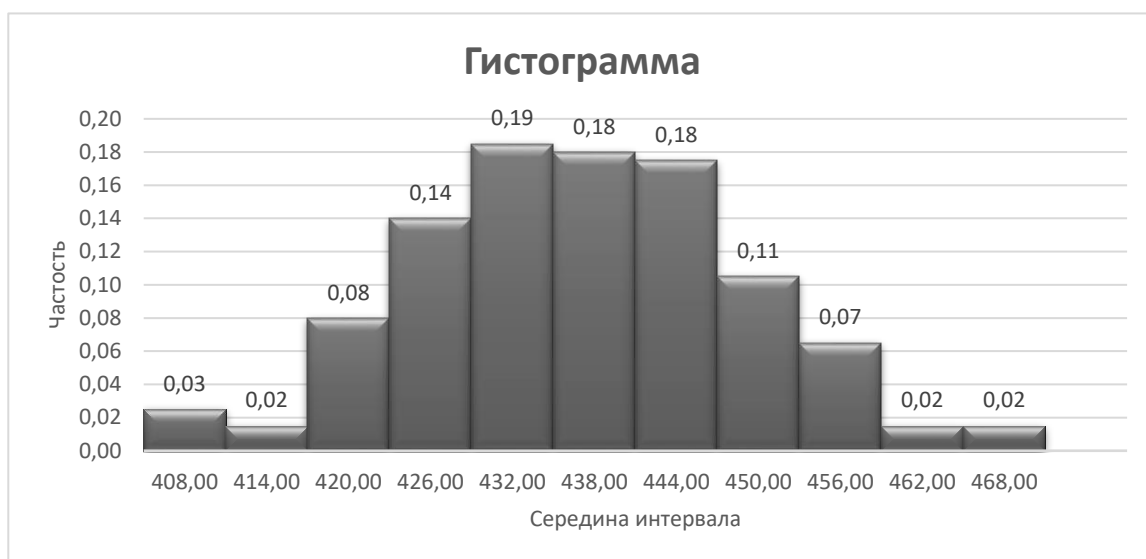


Рисунок 2. Гистограмма распределения

Над каждым отрезком оси абсцисс, изображающим расстояние между серединами интервалов значений x , проводится отрезок горизонтальной прямой на высоте, пропорциональной накопленной частоте (или накопленной частоте) в данном интервале. Концы отрезков соединяются. Накопленной частотой в данном интервале называется сумма всех частот, начиная с первого интервала до данного интервала включительно. На рис. 3 показана ступенчатая кривая распределения, построенная по данным табл. 3.

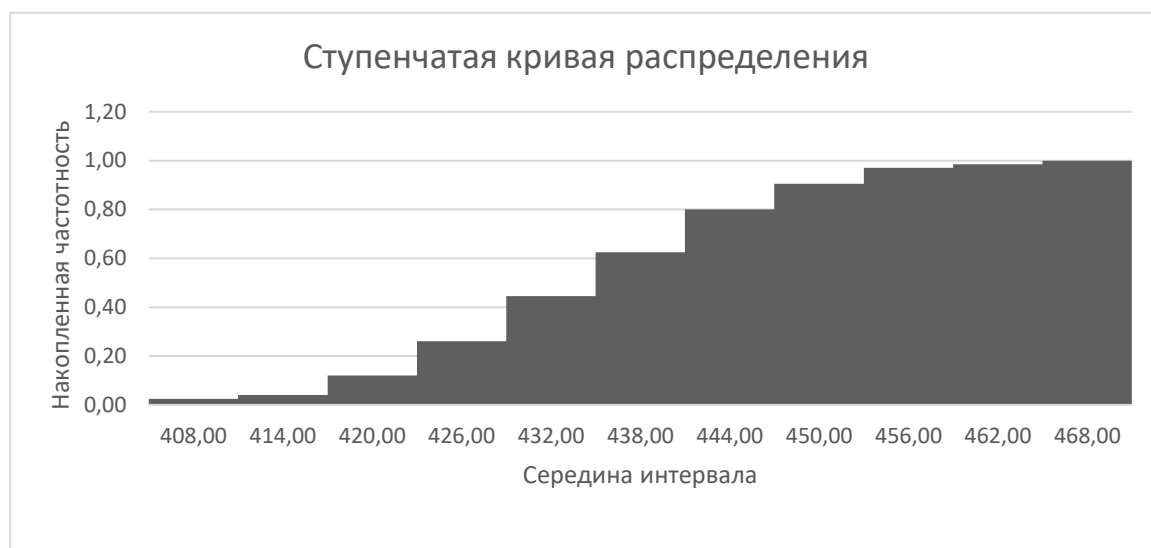


Рисунок 3. Ступенчатая кривая распределения

ТОЧЕЧНАЯ ОЦЕНКА ХАРАКТЕРИСТИК РАСПРЕДЕЛЕНИЯ ПО СГРУППИРОВАННЫМ ДАННЫМ

Для нахождения среднего арифметического значения выборки воспользуемся формулой:

$$\bar{x} = \frac{\sum_{i=1}^l x_i v_i}{n}$$
$$\bar{x} = 436,95$$

Оценкой моды является середина самого многочисленного интервала.

$$M = 432$$

Эмпирическая дисперсия s^2 и среднее квадратическое отклонение s вычисляются по формулам:

$$S^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 v_i - \bar{x}^2$$
$$s^2 = 180571,808$$
$$S = 424,937$$

Оценка дисперсии, полученная по группированным данным, оказывается смещенной (несколько увеличенной). Исправляют это смещение введением поправки Шеппарда:

$$s_*^2 = s^2 - \frac{(\Delta x)^2}{11}$$
$$s_*^2 = 29,331$$

Коэффициент вариации считается по формуле:

$$v = \frac{s}{\bar{x}}$$
$$v = 0,973$$

Далее вычислим оценки асимметрии и эксцесса по формулам:

1. Относительные середины интервалов

$$y_i = \frac{x_i - c}{\Delta x}$$

2. Начальные моменты

$$\bar{h}_i = \frac{1}{n} \sum_{j=1}^l \nu_j y_j^i$$

3. Среднее арифметическое \bar{x} и центральные эмпирические моменты m_2, m_3 и

$$\begin{aligned} \bar{x} &= \Delta x \bar{h}_1 + c &= 441,9 \\ m_2 &= s^2 = (\Delta x)^2 (\bar{h}_2 - \bar{h}_1^2) \\ m_3 &= (\Delta x)^3 (\bar{h}_3 - 3\bar{h}_2 \bar{h}_1 + 2\bar{h}_1^3) \\ m_4 &= (\Delta x)^4 (\bar{h}_4 - 4\bar{h}_3 \bar{h}_1 + 6\bar{h}_1^2 \bar{h}_2 - 3\bar{h}_1^4) \end{aligned} \left. \vphantom{\begin{aligned} \bar{x} \\ m_2 \\ m_3 \\ m_4 \end{aligned}} \right\} \begin{aligned} & \\ &= 180571,80836 \\ &= -3559,3020 \\ &= 157810,3317 \end{aligned}$$

4. Далее вычислим оценки асимметрии и эксцесса по формулам:

$$\tilde{S}_k = \frac{m_3}{s^3} = -0,000046$$

$$\tilde{E}_x = \frac{m_4}{s^4} - 3 = -2,999$$

ОЦЕНКА ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ МЕТОДОМ КВАНТИЛЕЙ

Предположим, что выборка, данная в табл. 1, подчиняется нормальному закону распределения. Определим оценки параметров этого закона математического ожидания m и среднего квадратического отклонения σ , используя метод квантилей. Для определения оценок двух неизвестных параметров m и σ составим, два уравнения, используя формулу:

$$\phi\left(\frac{x_n - m}{\sigma}\right) + 0,5 = F(x_n)$$

Для определения оценок двух неизвестных параметров m и σ составим два уравнения, используя формулу. Для этого из вариационного ряда или из ступенчатой кривой накопленной частоты (см. рис. 3) возьмем два любых значения с соответствующими им вероятностями.

$$\phi\left(\frac{426 - m}{\sigma}\right) + 0,5 = 0,26$$

$$\phi\left(\frac{456 - m}{\sigma}\right) + 0,5 = 0,97$$

Решение системы дает в результате:

$$m = 441$$

$$\sigma = 60$$

ПОСТРОЕНИЕ ДОВЕРИТЕЛЬНЫХ ИНТЕРВАЛОВ

Рассмотрим выборку малого объема из первых двадцати значений табл. 1. Построим таблицу, вариационный ряд (табл. 4 и 5). Найдем оценки математического ожидания и дисперсии.

№ изм.	xi
1	441,0
2	423,0
3	456,0
4	421,0
5	451,0
6	435,0
7	453,0
8	438,0
9	428,0
10	422,0
11	445,0
12	426,0
13	457,0
14	426,0
15	433,0
16	440,0
17	436,0
18	431,0
19	443,0
20	414,0

Таблица 4. Нумерованный вариационный ряд

№ изм.	xi	xi^2
1	441,0	194481
2	414,0	171396
3	421,0	177241
4	422,0	178084
5	423,0	178929
6	426,0	181476
7	426,0	181476
8	428,0	183184
9	431,0	185761
10	433,0	187489
11	435,0	189225
12	436,0	190096
13	438,0	191844
14	440,0	193600
15	443,0	196249
16	445,0	198025
17	451,0	203401
18	453,0	205209
19	456,0	207936
20	457,0	208849

Таблица 5. Нумерованный вариационный ряд

В предположении нормального распределения отклонения случайной величины от среднего построим доверительный интервал для математического ожидания при неизвестном σ по значениям данной выборки по формуле:

$$\bar{x} - t_{q,n-1} \frac{s}{\sqrt{n-1}} < m_x \leq \bar{x} + t_{q,n-1} \frac{s}{\sqrt{n-1}}$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}_2}$$

По формулам получаем:

$$\bar{x} = 458,89$$

$$s = 12,048 \text{ (E432 в xls)}$$

Определим доверительные интервалы для математического ожидания m_x , задаваясь различными уровнями значимости q . Значения квантилей $t_{q,n-1}$ берутся из таблиц распределения Стьюдента по двум входам: по числу степеней свободы $(n-1)$ и уровнями значимости q . Уровень значимости q здесь и в дальнейшем предполагается заданным.

Для $q=5\%$:

$$q = 5\%, t_{5,19} = 2,093 \text{ (по таблице распределения Стьюдента)}$$

$$430,165 < m_x \leq 441,73$$

Для $q=10\%$:

$$q = 10\%, t_{10,19} = 1,7295$$

$$431,169 < m_x \leq 440,73$$

Для $q=1\%$:

$$q = 1\%, t_{1,19} = 2,860935$$

$$428,04 < m_x \leq 443,86$$

Доверительный интервал для параметров σ^2 и σ строятся также по первым двадцати значениям выборки по формуле:

$$\left. \begin{aligned} \frac{ns^2}{\chi_2^2} < \sigma_x^2 < \frac{ns^2}{\chi_1^2} \\ \frac{\sqrt{ns}}{\chi_2} < \sigma_x < \frac{\sqrt{ns}}{\chi_1} \end{aligned} \right\}$$

При уровне значимости q получаем такие доверительные интервалы:

Для $q=1\%$:

$$72,58 < \sigma^2 < 390,5$$

$$8,52 < \sigma < 19,76$$

Для $q=10\%$:

$$92,42 < \sigma^2 < 267,53$$

$$9,61 < \sigma < 16,36$$

Для $q=5\%$:

$$84,96 < \sigma^2 < 302,68$$

$$9,22 < \sigma < 17,4$$

Таким образом, определены доверительные интервалы для параметров σ^2 и σ .

ПРИБЛИЖЕННЫЙ МЕТОД ПРОВЕРКИ ГИПОТЕЗЫ О НОРМАЛЬНОМ РАСПРЕДЕЛЕНИИ

Примем гипотезу о том, что выборка, данная в табл. 1, подчиняется нормальному закону распределения. Для приближенной проверки этой гипотезы могут быть использованы эмпирические асимметрия \tilde{S}_k и эксцесс \tilde{E}_x .

Для нормального распределения, как известно $S\tilde{k} = 0$ и $E\tilde{x} = 0$. Поэтому показатели асимметрии и эксцесса, отличные от нуля, указывают на отклонение рассматриваемого распределения от нормального.

Выборочные асимметрия и эксцесс, как и все оценки, являются случайными величинами и могут не совпадать с теоретическими. Среднеквадратические отклонения этих характеристик при заданном объеме выборки n вычисляются по формулам:

$$\left. \begin{aligned} \sigma_{\tilde{S}_k} &= \sqrt{\frac{6(n-2)}{(n+1)(n+3)}} \\ \sigma_{\tilde{E}_x} &= \sqrt{\frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}} \end{aligned} \right\}$$

$$n = 200$$

$$\sigma_{\tilde{S}_k} = 0,171$$

$$\sigma_{\tilde{E}_x} = 0,334$$

Зная эти величины, можно оценить, существенно ли отличаются оценки от оцениваемых асимметрии $S\tilde{k}$ и эксцесса $E\tilde{x}$, т.е. от нуля. Критические области для асимметрии $|S\tilde{k}| > 3\sigma_{\tilde{S}_k}$ и для эксцесса $|E\tilde{x}| > 3\sigma_{\tilde{E}_x}$ получены на основании неравенства Чебышева.

$$s_k = \frac{m_3}{s^3} = -0,000046$$

$$3 * \sigma_{\tilde{S}_k} = 0,513$$

$$\tilde{s}_k < 3\sigma_{\tilde{S}_k}$$

$$\bar{E}_x = -2,999$$

$$5 * \sigma_{E_x} = 1,67$$

$$\bar{E}_x < 5 * \sigma_{E_x}$$

Таким образом, $\tilde{E}x$ и $\tilde{S}k$ не попадают в критическую область, значит гипотеза нормальности не противоречит данным табл.1

КРИТЕРИЙ ХИ-КВАДРАТ ДЛЯ ПРОВЕРКИ СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Примем гипотезу о том, что выборка, данная в табл. 1, подчиняется нормальному закону распределения. Для проверки этой гипотезы воспользуемся критерием.

$$\chi^2 = \sum_{i=1}^l \frac{(v_i - \tilde{n}p)^2}{\tilde{n}p}$$

Значение критерия χ^2 определяется по данным выборки. Если это значение попадет в область допустимых значений $\chi^2 < \chi^2_q$, то следует признать, что данные выборки не противоречат гипотезе о нормальности распределения. Если же численное значение критерия 2 попадает в критическую область $\chi^2 > \chi^2_q$, то гипотеза отвергается. Вычисления критерия удобно свести в табл. 6.

Сумма, называемая критерием χ^2 , асимптотически распределена как хи-квадрат. При практических расчетах для нахождения критического значения этой суммы можно пользоваться таблицами распределения хи-квадрат только в том случае, если для всех интервалов $\tilde{n}p > 5$. Поэтому в табл. 6 интервалы с номерами 1, 2 и 9, 10, 11 объединены. Оценку вероятности \tilde{p} попадания в интервал находим по формуле:

$$P_i = P(\alpha < X < \beta) = \Phi\left(\frac{\beta - \bar{x}}{s}\right) - \Phi\left(\frac{\alpha - \bar{x}}{s}\right)$$

где α и β – границы интервалов, \bar{x} и s вычислены по данной выборке, а значения функции Лапласа берутся из таблицы.

Так как по данным выборки мы оценили два параметра m_x и σ_x нормального закона (т.е. $c = 2$), то в нашем случае число степеней свободы будет равно: $k = l' - c - 1 = 8 - 2 - 1 = 5$, где $l' = 8$ – число интервалов, получившихся после объединения интервалов.

№ интервала	Истинные границы интервалов		Границы интервалов, приведенные в долях	$\Phi(Z_i)$ от верхней границы	Оценка вероятности	Оценка мат. ожидания		Частота в интервале		Уклонение	Взвешенные квадраты уклонения
1	405	411	-0,09	-0,0359	0,46410	92,82	94,42	5	8	-86,42	79,1
2	411	417	-0,07	-0,0279	0,00800	1,6		3			
3	417	423	-0,06	-0,0239	0,00400	0,8		16		4,2	22,05
4	423	429	-0,04	-0,016	0,00790	1,58		28		5,42	18,59
5	429	435	-0,03	-0,012	0,00400	0,8		37		13,2	217,8
6	435	441	-0,02	-0,008	0,00400	0,8		36		13,2	217,8
7	441	447	0,00	0	0,00800	1,6		35		10,4	67,6
8	447	453	0,01	0,004	0,00400	0,8		21		15,2	288,8
9	453	459	0,03	0,012	0,00800	1,6	3,18	13	11	7,82	19,23
10	459	465	0,04	0,016	0,00400	0,8		3			
11	465	471	0,05	0,0199	0,00390	0,78		3			
				Суммы:	1,00	200		200			930,97

Таблица 6. Вычисления критерия хи-квадрат

По таблице распределения хи-квадрат найдем значения χ^2_q для числа степеней свободы $k = 5$ и уровней значимости q , равным 1%, 5% и 10%. Для $q = 1\%$:

$$\chi^2_q = 20,0902$$

Поскольку $3,71257 < 20,0902$, то гипотеза о нормальности выборки не противоречит данным измерений.

Для $q = 5\%$:

$$\chi^2_q = 15,5073$$

Поскольку $3,71257 < 15,5073$, то гипотеза о нормальности выборки не противоречит данным измерений.

Для $q = 10\%$:

$$\chi^2_q = 13,3615$$

Поскольку $3,71257 < 13,3615$, то гипотеза о нормальности выборки не противоречит данным измерений.

ПРОВЕРКА ГИПОТЕЗЫ ОБ ОДНОРОДНОСТИ ВЫБОРКИ С ПОМОЩЬЮ КРИТЕРИЯ ЗНАКОВ И КРИТЕРИЯ ВИЛКОКСОНА

Введем нулевую гипотезу H_0 о том, что выборка, данная в табл. 1, является однородной. Для проверки этой гипотезы возьмем две выборки из двадцати первых и двадцати последних значений табл. 1 и представим данные в табл. 8.

Воспользуемся непараметрическими (независимыми от формы распределения) критериями: критерием знаков и критерием Вилкоксона.

N_i	x_i	y_i	$z_i = x_i - y_i$
1	441	444	-3
2	423	449	-26
3	456	427	29
4	421	422	-1
5	451	420	31
6	435	432	3
7	453	425	28
8	438	426	12
9	428	428	0
10	422	441	-19
11	445	440	5
12	426	405	21
13	457	434	23
14	426	444	-18
15	433	431	2
16	440	435	5
17	436	431	5
18	431	433	-2
19	443	418	25
20	414	439	-25

Таблица 6. Вычисления критерия хи-квадрат

Составим разность $z_i = x_i - y_i$, где $i = 1, 2, \dots, 20$ – порядковые номера первых x_i и последних y_i двадцати значений выборки. Подсчитаем число положительных $k_n (+)$ и отрицательных $k_n (-)$ знаков разностей z_i ($n = 20$). Затем, выбрав уровень значимости q , находим по q и n в соответствующей

таблице критическое значение mn меньшего из чисел положительных и отрицательных знаков z_i . Если теперь меньшее из чисел знаков разностей окажется меньше mn , то гипотеза об однородности выборки отвергается, а если меньшее из чисел знаков разностей окажется больше mn , то следует признать, что гипотеза не противоречит данным выборки.

Для данных табл. 1 имеем: $k_{20}(+) = 13$ и $k_{20}(-) = 7$.

Из таблицы для $n=20$ и уровней значимости q , равным 1%, 5% и 10% находим:

- $q = 1\%$; $\bar{m}_{20}=3$
- $q = 5\%$; $\bar{m}_{20}=5$
- $q = 10\%$; $\bar{m}_{20}=5$

Т.к. меньшее из чисел знаков разностей $k_{20}(-) = 7$ при всех рассмотренных уровнях значимости оказалось больше значения \bar{m}_{20} , то нулевая гипотеза H_0 об однородности выборки не противоречит данным выборки.

Критерий Вилкоксона основан на числе инверсий.

Суммарное число инверсий обозначим за u .

№	Значение	x_i / y_i	Число инверсий
1	405	y	
2	414	x	1
3	418	y	
4	420	y	
5	421	x	3
6	422	x	3
7	422	y	
8	423	x	4
9	425	y	
10	426	x	5
11	426	x	5
12	426	y	
13	427	y	
14	428	x	7
15	428	y	
16	431	x	8
17	431	y	
18	431	y	
19	432	y	
20	433	x	11
21	433	y	
22	434	y	
23	435	x	13
24	435	y	

25	436	x	14
26	438	x	14
27	439	y	
28	440	x	15
29	440	y	
30	441	x	16
31	441	y	
32	443	x	17
33	444	y	
34	444	y	
35	445	x	19
36	449	y	
37	451	x	20
38	453	x	20
39	456	x	20
40	457	x	20
Суммарное значение			235

Далее введем нулевую гипотезу H_0 о том, что выборка, данная в табл. 1, является однородной. Эта гипотеза отвергается, если число инверсий u превосходит выбранную в соответствии с уровнем значимости границу, определяемую из того, что при объемах $n > 10$ и $m > 10$ выборочное число инверсий u распределено приблизительно нормально с центром:

$$M[u] = \frac{mn}{2}$$

дисперсией:

$$D[u] = \frac{mn}{12}(m + n + 1)$$

и средним квадратическим отклонением:

$$\sigma[u] = \sqrt{D[u]}$$

В нашем случае получим:

$$M[u] = \frac{20 \cdot 20}{2} = 200$$

$$D[u] = \frac{20 \cdot 20}{12}(20 + 20 + 1) = 1367$$

$$\sigma[u] = \sqrt{1367} = 37,0$$

Задавшись уровнем значимости $q = 5\%$, построим критическую область, используя соотношение:

$$q = 1 - 2\Phi_0(t_{q,III}) \text{ где } \Phi_0(t_{q,III}) \text{ — функция Лапласа.}$$

Используя эту формулу и таблицу нормального распределения, находим значение

$$t_{q,III} = 1,96$$

Критическая область для гипотезы H_0 :

$$\left. \begin{aligned} u &\leq M[u] - t_{q,III} * \sigma[u] \\ u &\geq M[u] + t_{q,III} * \sigma[u] \end{aligned} \right\}$$

Найдем критические области для выдвинутой нулевой гипотезы, задавшись уровнем значимости **q**, равным **1%**, **5%** и **10%**:

$$q = \mathbf{1\%}$$

$$0,01=q = 1 - 2\Phi_0(t_{q,III})$$

$$\Phi_0(t_{q,III}) = 0,495$$

$$t_{q,III} = 2,58$$

$$\left. \begin{aligned} u &\leq 104,54 \\ u &\geq 295,46 \end{aligned} \right\}$$

$$q = \mathbf{5\%}$$

$$0,05=q = 1 - 2\Phi_0(t_{q,III})$$

$$\Phi_0(t_{q,III}) = 0,475$$

$$t_{q,III} = 1,96$$

$$\left. \begin{aligned} u &\leq 127,5 \\ u &\geq 272,5 \end{aligned} \right\}$$

$$q = \mathbf{10\%}$$

$$0,1=q = 1 - 2\Phi_0(t_{q,III})$$

$$\Phi_0(t_{q,III}) = 0,45$$

$$t_{q,III} = 1,64$$

$$\left. \begin{aligned} u &\leq 139,32 \\ u &\geq 260,68 \end{aligned} \right\}$$

При всех рассмотренных уровнях значимости число инверсий **u** не лежит в критической области, а потому нулевая гипотеза **H₀** об однородности выборки не противоречит данным выборки.

