

Санкт-Петербургский Государственный Политехнический Университет  
Институт компьютерных наук и технологий  
Высшая школа программной инженерии

**ОТЧЕТ ПО РАСЧЕТНОМУ ЗАДАНИЮ**  
по дисциплине «Теория вероятностей и математическая статистика»  
Вариант - 174

Выполнила  
студент гр. 3530202/90202



Потапова А.М.

Преподаватель

Зайцев И.В.

Санкт-Петербург  
2021

## Содержание

1. Таблица нумерованных денормализованных чисел.....	3
2. Построение нумерованного вариационного ряда.....	4
3. Оценка математического ожидания, дисперсии (смещенной, несмещенной), медианы.....	5
4. Группирование по интервалам, таблица разбиения на интервалы .....	6
5. Построение полигона, гистограммы, ступенчатой кривой .....	7
6. Точечная оценка характеристик распределения по сгруппированным данным .....	9
7. Оценка параметров распределения методом квантилей.....	11
8. Построение доверительных интервалов.....	12
9. Приближенный метод проверки гипотезы о нормальном распределении .	15
10. Критерий хи-квадрат для проверки статистических гипотез.....	16
11. Проверка гипотезы об однородности выборки с помощью критерия знаков и критерия Вилкоксона.....	18

## 1. Таблица нумерованных денормализованных чисел

1	2235	41	2209	81	2221	121	2201	161	2200
2	2204	42	2215	82	2198	122	2210	162	2224
3	2214	43	2220	83	2215	123	2206	163	2188
4	2202	44	2196	84	2222	124	2207	164	2221
5	2193	45	2204	85	2216	125	2210	165	2217
6	2208	46	2208	86	2213	126	2219	166	2212
7	2210	47	2212	87	2231	127	2206	167	2215
8	2219	48	2211	88	2229	128	2217	168	2216
9	2223	49	2207	89	2224	129	2233	169	2215
10	2212	50	2220	90	2215	130	2209	170	2225
11	2207	51	2216	91	2221	131	2210	171	2211
12	2230	52	2224	92	2215	132	2230	172	2209
13	2212	53	2204	93	2206	133	2209	173	2210
14	2211	54	2207	94	2222	134	2229	174	2215
15	2222	55	2202	95	2225	135	2214	175	2217
16	2217	56	2208	96	2224	136	2224	176	2205
17	2203	57	2212	97	2221	137	2221	177	2220
18	2214	58	2218	98	2214	138	2222	178	2212
19	2212	59	2193	99	2210	139	2213	179	2208
20	2227	60	2223	100	2208	140	2223	180	2191
21	2226	61	2221	101	2224	141	2216	181	2218
22	2214	62	2210	102	2208	142	2212	182	2200
23	2208	63	2212	103	2204	143	2218	183	2223
24	2217	64	2205	104	2209	144	2214	184	2219
25	2192	65	2214	105	2212	145	2214	185	2230
26	2213	66	2204	106	2210	146	2213	186	2214
27	2206	67	2214	107	2215	147	2207	187	2205
28	2208	68	2233	108	2218	148	2225	188	2228
29	2208	69	2214	109	2218	149	2221	189	2224
30	2217	70	2201	110	2212	150	2217	190	2216
31	2220	71	2213	111	2204	151	2216	191	2229
32	2220	72	2211	112	2221	152	2221	192	2223
33	2203	73	2215	113	2228	153	2200	193	2210
34	2213	74	2215	114	2222	154	2224	194	2216
35	2217	75	2205	115	2222	155	2219	195	2216
36	2201	76	2220	116	2213	156	2207	196	2212
37	2225	77	2205	117	2203	157	2212	197	2203
38	2210	78	2223	118	2225	158	2199	198	2217
39	2227	79	2232	119	2225	159	2226	199	2222
40	2204	80	2213	120	2203	160	2212	200	2203

Табл. 1. Таблица нумерованных денормализованных чисел

## 2. 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

1	2188	41	2207	81	2212	121	2216	161	2222
2	2191	42	2207	82	2212	122	2216	162	2222
3	2192	43	2207	83	2212	123	2216	163	2222
4	2193	44	2207	84	2212	124	2217	164	2223
5	2193	45	2208	85	2212	125	2217	165	2223
6	2196	46	2208	86	2212	126	2217	166	2223
7	2198	47	2208	87	2213	127	2217	167	2223
8	2199	48	2208	88	2213	128	2217	168	2223
9	2200	49	2208	89	2213	129	2217	169	2223
10	2200	50	2208	90	2213	130	2217	170	2224
11	2200	51	2208	91	2213	131	2217	171	2224
12	2201	52	2208	92	2213	132	2217	172	2224
13	2201	53	2208	93	2213	133	2218	173	2224
14	2201	54	2209	94	2213	134	2218	174	2224
15	2202	55	2209	95	2214	135	2218	175	2224
16	2202	56	2209	96	2214	136	2218	176	2224
17	2203	57	2209	97	2214	137	2218	177	2224
18	2203	58	2209	98	2214	138	2219	178	2225
19	2203	59	2210	99	2214	139	2219	179	2225
20	2203	60	2210	100	2214	140	2219	180	2225
21	2203	61	2210	101	2214	141	2219	181	2225
22	2203	62	2210	102	2214	142	2220	182	2225
23	2204	63	2210	103	2214	143	2220	183	2225
24	2204	64	2210	104	2214	144	2220	184	2226
25	2204	65	2210	105	2214	145	2220	185	2226
26	2204	66	2210	106	2215	146	2220	186	2227
27	2204	67	2210	107	2215	147	2220	187	2227
28	2204	68	2210	108	2215	148	2221	188	2228
29	2204	69	2211	109	2215	149	2221	189	2228
30	2205	70	2211	110	2215	150	2221	190	2229
31	2205	71	2211	111	2215	151	2221	191	2229
32	2205	72	2211	112	2215	152	2221	192	2229
33	2205	73	2212	113	2215	153	2221	193	2230
34	2205	74	2212	114	2215	154	2221	194	2230
35	2206	75	2212	115	2215	155	2221	195	2230
36	2206	76	2212	116	2216	156	2221	196	2231
37	2206	77	2212	117	2216	157	2222	197	2232
38	2206	78	2212	118	2216	158	2222	198	2233
39	2207	79	2212	119	2216	159	2222	199	2233
40	2207	80	2212	120	2216	160	2222	200	2235

Табл. 2. Нумерованный вариационный ряд

### 3. Оценка математического ожидания, дисперсии (смещенной, несмещенной), медианы

1) Оценка математического ожидания – среднее арифметическое – вычисляется по формуле:

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

$$\bar{x} = 2214,12$$

2) Смещенную оценку дисперсии по всей выборке вычисляем по сокращенной формуле:

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

$$S^2 = 77,5156$$

$$S = 8,8042$$

3) Несмещенную оценку дисперсии вычисляют по формуле:

$$s^{*2} = \frac{ns^2}{n-1}$$

$$s^{*2} = 77,9051$$

$$s^* = 8,8263$$

4) 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:

$$\tilde{M}_e = \frac{x_{100} + x_{101}}{2}$$

$$\tilde{M}_e = 2214,0$$

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

$$R = x_{\max} - x_{\min}$$

$$R = 47$$

#### 4. 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  $l = 10$ . Для удобства вычислений примем  $R = 50$ . Тогда ширина интервала (шаг разбиения) будет равна:

$$\Delta x = \frac{R}{L}$$

$$\Delta x = 5$$

№ интервала	Границы интервалов		Частота в интервале	Частность в интервале	Середина интервала
1	2185	2190	1	0,005	2187,5
2	2190	2195	4	0,02	2192,5
3	2195	2200	6	0,03	2197,5
4	2200	2205	23	0,115	2202,5
5	2205	2210	34	0,17	2207,5
6	2210	2215	47	0,235	2212,5
7	2215	2220	32	0,16	2217,5
8	2220	2225	36	0,18	2222,5
9	2225	2230	12	0,06	2227,5
10	2230	2235	5	0,025	2232,5

Табл. 3. Подсчет частот и частотностей по интервалам вариационного ряда

## 5. Построение полигона, гистограммы, ступенчатой кривой

### 1) Построение полигона распределения

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

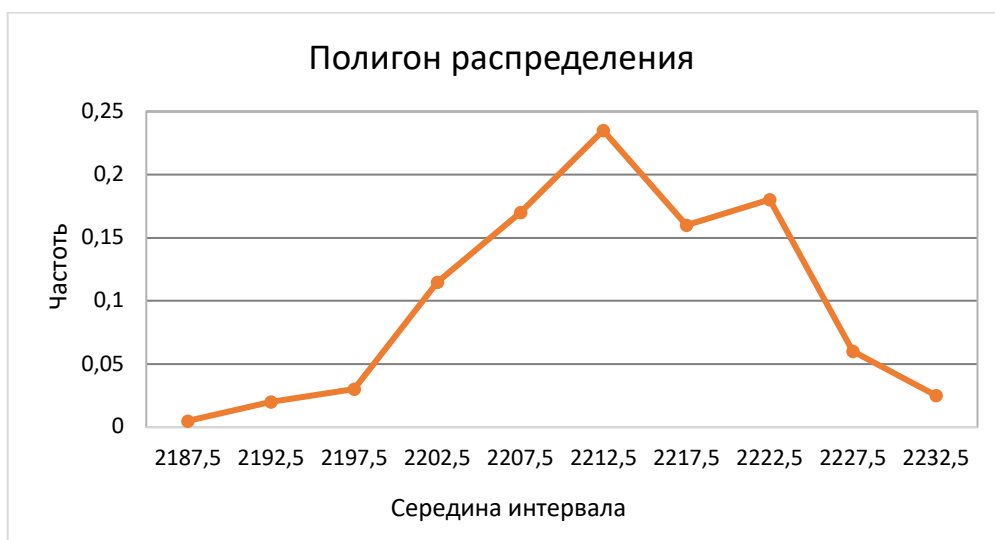


Рис. 1. Полигон распределения

### 2) Построение гистограммы распределения

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



Рис. 2. Гистограмма распределения

### 3) Построение ступенчатой кривой 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.

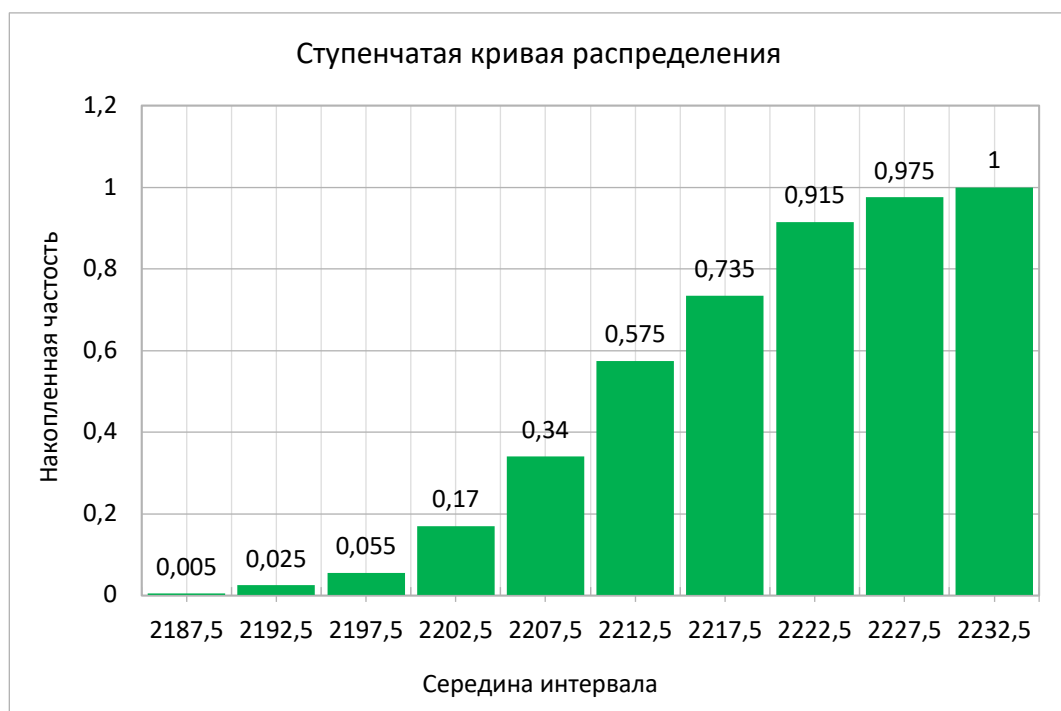


Рис. 3. Ступенчатая кривая распределения



## 6. Точечная оценка характеристик распределения по сгруппированным данным

1) Среднее арифметическое вычисляется по формуле:

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

$x_i$  — середина интервала;  $l$  — число интервалов;  $v_i$  — частота в интервале;  $n$  — число элементов в выборке

$$\bar{x} = 2213,525$$

2) Оценкой моды является середина самого многочисленного интервала:

$$M = 2212,5$$

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

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

$$s^2 = 78,0744$$

$$s = 8,8359$$

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

$$s^{*2} = s^2 - \frac{(\Delta x)^2}{12}$$

$$s^{*2} = 76$$

$$s^* = 8,7178$$

5) Коэффициент вариации считается по формуле:

$$v = \frac{s}{\bar{x}} \quad (\text{при } \bar{x} \neq 0)$$

$$v = 0,004$$

#### 6) Вычисление оценки асимметрии:

$$\tilde{s}_k = \frac{m_3}{s_3}$$

$m_3$  – центральный момент третьего порядка, который считается по формуле:

$$m_3 = \frac{\sum_{i=1}^l (x_i - \bar{x})^3 * v_i}{\sum_{i=1}^l v_i}$$

$$m_3 = -114,28$$

$$\tilde{s}_k = -0,1721$$

#### 7) Вычисление оценки эксцесса:

$$\bar{E}_k = \frac{m_4}{s_4} - 3$$

$m_4$  – центральный момент четвертого порядка, который считается по формуле:

$$m_4 = \frac{\sum_{i=1}^l (x_i - \bar{x})^4 * v_i}{\sum_{i=1}^l v_i}$$

$$m_4 = 16840,91$$

$$\bar{E}_k = -0,05$$

## 7. 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  $\sigma$ , используя метод квантилей. Для определения оценок двух неизвестных параметров  $m$  и  $\sigma$  составим, два уравнения, используя формулу:

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

Возьмем две точки  $x_4$  и  $x_8$  на кривой функции распределения со значениями  $F(x)$ , равными  $\text{асс}F_4/N$  и  $\text{асс}F_8/N$ . Составим систему из двух уравнений:

$$\begin{cases} \Phi\left(\frac{2202,5 - m}{\sigma}\right) = 0,17 \\ \Phi\left(\frac{2222,5 - m}{\sigma}\right) = 0,915 \end{cases}$$

Решение системы дает в результате:

$$\begin{cases} m = 2210 \\ \sigma = 8,5984 \end{cases}$$

## 8. Построение доверительных интервалов

### 1) Доверительный интервал для математического ожидания

Рассмотрим выборку малого объема из первых двадцати значений табл. 1. Построим таблицу, вариационный ряд (табл. 5 и 6). Найдем оценки математического ожидания и дисперсии.

№ пп.	$x_i$
1	2235
2	2204
3	2214
4	2202
5	2193
6	2208
7	2210
8	2219
9	2223
10	2212
11	2207
12	2230
13	2212
14	2211
15	2222
16	2217
17	2203
18	2214
19	2212
20	2227

Таблица 5

№ пп.	$x_i$	$x_i^2$
1	2188	4787344
2	2191	4800481
3	2192	4804864
4	2193	4809249
5	2193	4809249
6	2196	4822416
7	2198	4831204
8	2199	4835601
9	2200	4840000
10	2200	4840000
11	2200	4840000
12	2201	4844401
13	2201	4844401
14	2201	4844401
15	2202	4848804
16	2202	4848804
17	2203	4853209
18	2203	4853209
19	2203	4853209
20	2203	4853209

Таблица 6

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

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

$$\left\{ \begin{array}{l} \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \\ s = \sqrt{\left[ \frac{1}{n} \sum_{i=1}^n x_i^2 \right] - \bar{x}^2} \end{array} \right.$$

$$\bar{x} = 2198,5$$

$$s^2 = 20,3475$$

$$s = 4,5108$$

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 уровням значимости  $q$ . Уровень значимости  $q$  здесь и в дальнейшем предполагается заданным.

$$q = 5\%, t_{5,19} = 2,093$$

$$\bar{x} - t_{q,n-1} * \frac{s}{\sqrt{n-1}} = 2196,28$$

$$\bar{x} + t_{q,n-1} * \frac{s}{\sqrt{n-1}} = 2200,62$$

Доверительный интервал:

$$2196,28 < m_x < 2200,62$$

$$q = 5\%, t_{5,19} = 2,093$$

Доверительный интервал:

$$2195,49 < m_x < 2201,41$$

$$q = 1\%, t_{5,19} = 2,8615$$

Доверительный интервал:

$$2196,66 < m_x < 2200,24$$

$$q = 10\%, t_{5,19} = 1,7295$$

## 2) Доверительный интервал для дисперсии и среднего квадратического отклонения

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

$$\frac{\sqrt{n} * s}{x_{\frac{1-\gamma}{2}, n-1}} \leq \sigma \leq \frac{\sqrt{n} * s}{x_{\frac{1+\gamma}{2}, n-1}}$$

Принимаем надежность интервала за 0.9 и получаем такие доверительные интервалы:

$$13,5 \leq \sigma_x^2 \leq 40,2243$$

$$3,6742 \leq \sigma_x \leq 6,3422$$

## 9. 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  $\tilde{S}_k = 0$  и  $\tilde{E}_x = 0$ . 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:

Для  $n = 200$  получаем:

$$\begin{cases} \sigma \tilde{E}_k = \sqrt{\frac{6(n-2)}{(n+1)(n+3)}} = 0,171 \\ \sigma \tilde{E}_x = \sqrt{\frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}} = 0,334 \end{cases}$$

Зная эти величины, можно оценить, существенно ли отличаются оценки от оцениваемых асимметрии  $\tilde{S}_k$  и эксцесса  $\tilde{E}_x$ , т. е. от нуля. Критические области для асимметрии  $|\tilde{S}_k| > 3\sigma \tilde{S}_k$  и для эксцесса  $|\tilde{E}_x| > 5\sigma \tilde{E}_x$  получены на основании неравенства Чебышева.

$$1) \tilde{S}_k = \frac{m_3}{s^3} = -0,1721$$

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

$$|\tilde{S}_k| < 3\sigma \tilde{S}_k$$

$$2) \tilde{E}_k = \frac{m_4}{s^4} - 3$$

$$\tilde{E}_k = -0,05$$

$$5\sigma \tilde{E}_x = 1,67$$

$$|\tilde{E}_x| < 5\sigma \tilde{E}_x$$

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

## 10. 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.

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

Значение критерия  $\chi^2$  определяется по данным выборки. Если это значение попадет в область допустимых значений  $\chi^2 < \chi_{q2}$ , то следует признать, что данные выборки не противоречат гипотезе о нормальности распределения. Если же численное значение критерия  $\chi^2$  попадает в критическую область  $\chi^2 > \chi_{q2}$ , то гипотеза отвергается. Вычисления критерия удобно свести в табл. 7.

Сумма, называемая критерием  $\chi^2$ , 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  $n\tilde{p}_i > 5$  (более строгое требование  $n\tilde{p}_i > 10$ ). Поэтому в табл. 7 интервалы с номерами 1, 2 и 11, 12 объединены. Оценку вероятности  $\tilde{p}_i$  попадания в интервал находим по формуле:

$$\tilde{p}_i = p(\alpha < X < \beta) = \Phi\left(\frac{\beta - \bar{x}}{s}\right) - \Phi\left(\frac{\alpha - \bar{x}}{s}\right),$$

где  $\alpha$  и  $\beta$  – границы интервалов,  $\bar{x}$  и  $s$  вычислены в п. 6 по данной выборке, а значения функции Лапласа берутся из таблицы.

Так как по данным выборки мы оценили два параметра  $m_x$  и  $\sigma_x$  нормального закона (т.е.  $c = 2$ ), то в нашем случае число степеней свободы будет равно:  $k = l' - c - 1 = 7 - 2 - 1 = 4$ , где  $l' = 7$  – число интервалов, получившихся после объединения интервалов 1, 2, 3 и 9, 10.

По таблице распределения хи-квадрат найдем значения  $\chi_{q2}$  для числа степеней свободы  $k = 4$  и уровней значимости  $q$ , равным 1%, 5% и 10%:

1)  $q = 1\%$

$$\chi_{q2}^2 = 13,276$$

Поскольку  $9,4556 < 13,2767$ , то гипотеза о нормальности выборки не противоречит данным измерений.



$$2) q = 5\%$$

$$\chi^2_q = 9,48$$

Поскольку  $9,4556 < 9,4877$ , то гипотеза о нормальности выборки не противоречит данным измерений

$$3) q = 10\%$$

$$\chi^2_q = 7,7794$$

Поскольку  $9,4556 > 7,7794$ , то гипотеза о нормальности выборки противоречит данным измерений

№ интервала	Истинные границы интервала		Границы интервалов, приведенные в долях	$\Phi(Z_i)$ от верхней границы	Оценка вероятности	Оценка мат. ожидания		Частота в интервале		Уклонение	Взвешенные квадратные уравнения
1	2185	2190	-2,719	-0,4969	0,0045	0,8987	15,504	1	11	-4,504	1,3085
2	2190	2195	-2,153	-0,4849	0,0183	3,6663		4			
3	2195	2200	-1,586	-0,4452	0,0547	10,9392		6			
4	2200	2205	-1,020	-0,3493	0,1194	23,8787		23		-0,879	0,0323
5	2205	2210	-0,453	-0,1797	0,1907	38,1409		34		-4,141	0,4496
6	2210	2215	0,113	0,0397	0,2229	44,5839		47		2,416	0,1309
7	2215	2220	0,680	0,2474	0,1907	38,1409		32		-6,141	0,9887
8	2220	2225	1,246	0,3912	0,1194	23,8787		36		12,121	6,1530
9	2225	2230	1,813	0,4640	0,0547	10,9392	14,605	12	17	2,394	0,3926
10	2230	2235	2,379	0,4910	0,0183	3,6663		5			
				Суммы:	1	199		200			9,4556

Таблица 7

## 11. 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.

$x_i$	$y_i$	$z_i = x_i - y_i$
2235	2218	17
2204	2200	4
2214	2223	-9
2202	2219	-17
2193	2230	-37
2208	2214	-6
2210	2205	5
2219	2228	-9
2223	2224	-1
2212	2216	-4
2207	2229	-22
2230	2223	7
2212	2210	2
2211	2216	-5
2222	2216	6
2217	2212	5
2203	2203	0
2214	2217	-3
2212	2222	-10
2227	2203	24

Таблица 8

### 1) Критерий знаков

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 таблицы 8 имеем:  $k(+) = 9$ ,  $k(-) = 11$ . Из таблицы для  $n = 20$  и уровней значимости  $q$ , равным 1%, 5% и 10% находим:

1)  $q = 1\%$ ,  $\bar{m}_{20} = 4$

2)  $q = 5\%$ ,  $\bar{m}_{20} = 5$

3)  $q = 10\%$ ,  $\bar{m}_{20} = 5$

Так как меньшее из чисел знаков разностей  $k_{20}(-) = 11$  при всех рассмотренных уровнях значимости оказалось больше значения  $\bar{m}_{20}$ , то нулевая гипотеза  $H_0$  об однородности выборки не противоречит данным выборки.

## 2) Критерий Вилкоксона

Критерий Вилкоксона основан на числе инверсий. Суммарное число инверсий обозначим за  $u = 247$ .

Последовательность	Буква	Число инверсий
2193	х	
2202	х	
2200	у	2
2203	х	
2203	у	3
2203	у	3
2204	х	
2205	у	4
2207	х	
2208	х	
2210	х	
2210	у	7
2211	х	
2212	х	
2212	х	
2212	х	
2212	у	11
2214	х	
2214	х	
2214	у	13
2216	у	13
2216	у	13
2216	у	13
2217	х	

2217	y	14
2218	y	14
2219	x	
2219	y	15
2222	x	
2222	y	16
2223	x	
2223	y	17
2223	y	17
2224	y	17
2227	x	
2228	y	18
2229	y	18
2230	x	
2230	y	19
2235	x	
Сумма		247

Таблица 9

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:

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

Тогда:

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

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

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

Задаем уровень значимости  $q = 5\%$  и строим критическую область, используя соотношение:

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

где  $\Phi_0$  – функция Лапласа.

Находим критическую область для гипотезы  $H_0$ :

$$u \leq M[u] - t_{q,III} * \sigma[u]$$

$$u \geq M[u] + t_{q,III} * \sigma[u]$$

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

1)  $q = 1\%$

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

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

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

$$\begin{cases} u \leq 104,54 \\ u \geq 295,46 \end{cases}$$

2)  $q = 5\%$

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

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

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

$$\begin{cases} u \leq 127,5 \\ u \geq 272,5 \end{cases}$$

3)  $q = 10\%$

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

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

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

$$\begin{cases} u \leq 139,32 \\ u \geq 260,68 \end{cases}$$

При всех рассмотренных уровнях значимости число инверсий  $u$  не лежит в критической области, а потому нулевая гипотеза  $H_0$  об однородности выборки не противоречит данным выборки.