ГУАП

КАФЕДРА № 42

ОТЧЕТ
ЗАЩИЩЕН С ОЦЕНКОЙ ____________________

ПРЕПОДАВАТЕЛЬ

Доцент		В.Н. Ассаул
должность, уч. степень, звание	подпись, дата	инициалы, фамилия

ОТЧЕТ О РАСЧЁТНО-ГРАФИЧЕСКОЙ РАБОТЕ

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Вариант 2

по курсу: ТЕОРИЯ ВЕРОЯТНОСТИ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

РАБОТУ ВЫПОЛНИЛ

СТУДЕНТ ГР. №	4128		В.А Воробьев
		подпись, дата	инициалы, фамилия

Санкт-Петербург 2022

СОДЕРЖАНИЕ

1. РАНЖИРОВАНИЕ И РАЗМАХ ВЫБОРКИ 3

2. ПРЕОБРАЗОВАНИЕ В ИНТЕРВАЛЬНЫЙ РЯД 5

3. ПОЛИГОН И ГИСТОГРАММА 6

4. ВЫБОРОЧНЫЕ МОДА И МЕДИАНА 6

4.1. Выборочная мода 6

4.1. Выборочная медиана 7

5. ВЫБОРОЧНОЕ СРЕДНЕЕ, ДИСПЕРСИЯ И СКО 9

5.1. Выборочное среднее 9

5.2. Дисперсия 9

5.3. СКО 9

5.4 Исправленная дисперсия 9

5.4 Исправленная СКО 9

6. ГИПОТЕЗА О НОРМАЛЬНОМ РАСПРЕДЕЛЕНИИ 10

7. ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ ДЛЯ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ 13

7.1. При известной дисперсии 13

7.2. При неизвестной дисперсии 13

8. ТАБЛИЦА СО ВСЕМИ ОТДЕЛЬНО НАЙДЕННЫМИ ЗНАЧЕНИЯМИ 14

1. Ранжирование и размах выборки

Массив данных		
Столбец 1	Столбец 2	Столбец 3
555	559	567
572	574	542
543	565	568
546	582	568
562	558	545
566	576	560
548	562	556
563	552	562
569	557	555
549	562	552
567	566	547
550	552	571
560	549	544
553	572	544
564	538	552
548	549	543
543	538	552
561	547	587
562	554	548
560	560	535

Рисунок 1 – Исходные данные по варианту

Массив данных с рангами					
Столбец 1		Столбец 2		Столбец 3	
555	32	559	28	567	11
572	5	574	4	542	57
543	54	565	15	568	9
546	50	582	2	568	9
562	18	558	29	545	51
566	26	576	3	560	24
548	45	562	18	556	31
563	17	552	36	562	18
569	8	557	30	555	32
549	42	562	18	552	36
567	11	566	13	547	48
550	41	552	36	571	7
560	24	549	42	544	52
553	35	572	5	544	52
564	16	538	58	552	36
548	23	549	42	543	54
543	54	538	58	552	36
561	23	547	48	587	1
562	18	554	34	548	45
560	52	560	24	535	60

Рисунок 2 – Ранжированная таблица исходных данных

Всего дано 60 вариаций. На ранжированной таблице (рис 2.) видно, что минимальная варианта = 535, а максимальная = 587.

Размах выборки = 587 – 535 = 45.

2. Преобразование в интервальный ряд

Длина интервала = *Размах выборки* / *Число интервалов* = 52/8 = 6.5

Интервал	x	x^2	n	w
535 - 541,5	538,25	289713,063	3	0,050
541,5 - 548	544,75	296752,563	13	0,217
548 - 554,5	551,25	303876,563	11	0,183
554,5 - 561	557,75	311085,063	11	0,183
561 - 567,5	564,25	318378,063	12	0,200
567,5 - 574	570,75	325755,563	7	0,117
574 - 580,5	577,25	333217,563	1	0,017
580,5 - 587	583,75	340764,063	2	0,033
		Сумма:	60	1

Рисунок 3 – Интервальный вариационный ряд

Формулы:

1. Средние значения интервалов - x:

=(N3+L3)/2;

2. Квадрат от x:

=O3*O3;

3. Частота - n:

=СЧЁТЕСЛИМН($A$3:$C$22;">="&L3;$A$3:$C$22;"<="&N3),

$A$3:$C$22 – массив исходных данных.

4. Плотность - w:

=Q3/60

3. Полигон и гистограмма

Рисунок 4 – Полигон частот

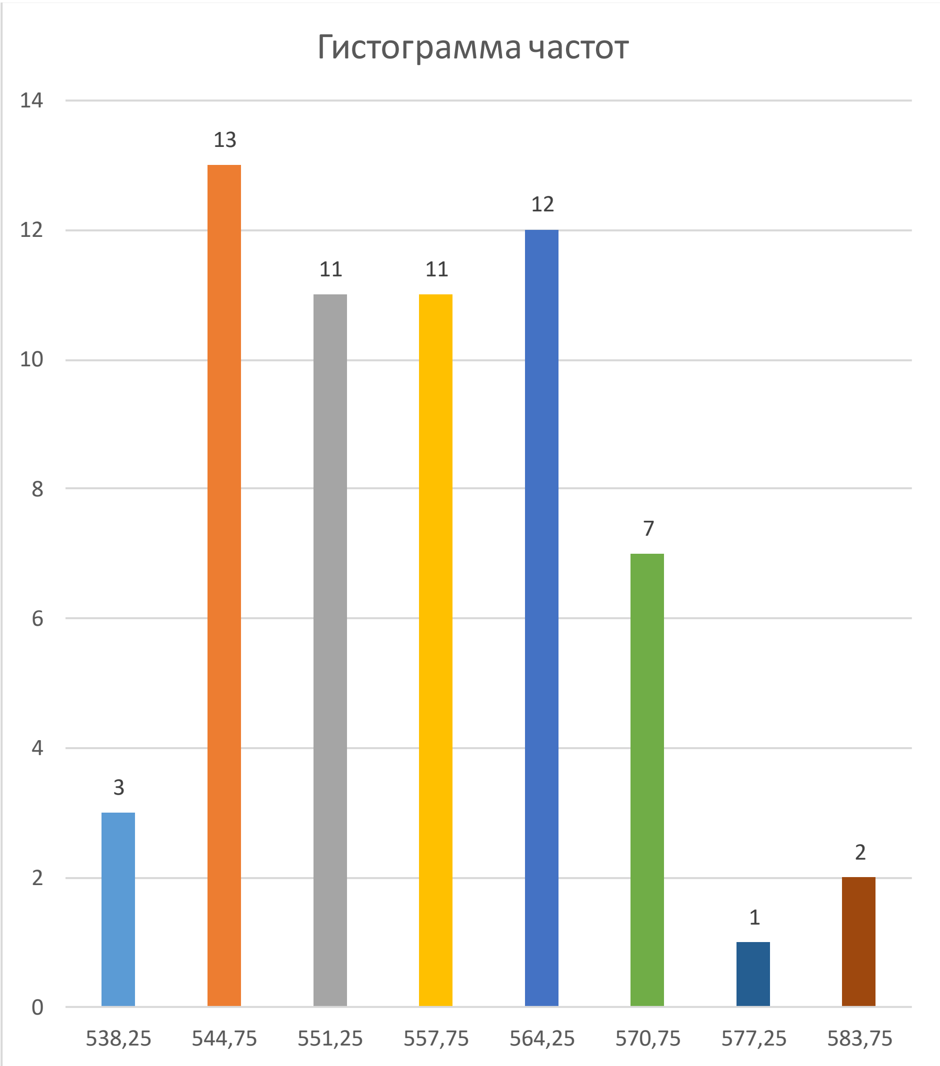

Рисунок 5 – Гистограмма частот

4. Выборочные мода и медиана

4.1. Выборочная мода

Выборочная мода находится на основании интервала с наибольшей частотой – в данном случае это 541.5 – 548, где n = 13. Для этого берется формула $M_0 = x_0 + \frac{n_M - n_{M-1}}{(n_M - n_{M-1}) + (n_M - n_{M+1})} \cdot h$, где

x_0 – нижняя граница модального интервала

n_M – частота модального интервала

n_{M-1} – частота предыдущего интервала

n_{M+1} – частота следующего интервала

h – длина модального интервала

В результате получается следующая формула '=L4+(Q4-Q3)/((Q4-Q3)*(Q4-Q5))*$C$24, где $C$24 – это длина интервала, т.к. частота, записанная в Q6 – наибольшая. И в итоге M_0 = 554.750.

4.1. Выборочная медиана

Для нахождения медианы таблица интервального вариационного ряда была расширена (рис.6), чтобы включать в себя накопленные частоты и промежуточные значения, которые будут использоваться в дальнейшем.

n	w	x*n	(x^2)*n	Накопление частот
3	0,050	1614,750	869139,188	3
13	0,217	7081,750	3857783,313	16
11	0,183	6063,750	3342642,188	27
11	0,183	6135,250	3421935,688	38
12	0,200	6771,000	3820536,750	50
7	0,117	3995,250	2280288,938	57
1	0,017	577,250	333217,563	58
2	0,033	1167,500	681528,125	60
60	1	33406,500	18607071,750	

Рисунок 6 – Интервальный вариационный ряд с накоплением частот и промежуточными значениями

Формула медианы: $m_e = x_0 + \frac{0,5n - n^{н}_{m-1}}{n_m} \cdot h$, где

n – объем статистической совокупности

x_0 – нижняя граница медианного интервала

n_m – частота медианного интервала

$n_{m-1}^{н}$ – накопленная частота следующего интервала

h – длина медианного интервала

Для этого в соответствующую ячейку была записана следующая формула: =L4+((0,5*Q11-V3)/Q4)*C24. И получилось, что m_e = 555.

5. Выборочное среднее, дисперсия и СКО

5.1. Выборочное среднее

Т.к. формула выборочной средней: $\overline{x}_в = \frac{\sum_{i=1}^{k} x_i n_i}{n}$, то все значения уже найдены, поэтому в соответствующую ячейку вписывается эта формула '=S11/Q11' с результатом 557,156.

5.2. Дисперсия

Из формулы для дисперсии $D_в = \frac{\sum_{i=1}^{k} x_i^2 n_i}{n} - \left(\frac{\sum_{i=1}^{k} x_i n_i}{n} \right)^2$ выводится следующая формула для Excel: '=U11/Q11-СТЕПЕНЬ((S11)/Q11;2)' с результатом 119.462

5.3. СКО

Среднее квадратичное отклонение равно квадратному корню из дисперсии, поэтому СКО = $\sqrt{119.462} \approx 10.93$

5.4 Исправленная дисперсия

$$S^2 = \frac{n}{n-1} D_в$$

, формула для ячейки "=(O39/(O39-1))*Q26" с результатом 121.487.

5.4 Исправленная СКО

Среднее исправленное квадратичное отклонение равно квадратному корню из исправленной дисперсии, поэтому СКО = $\sqrt{121.487} \approx 11.02$

6. Гипотеза о нормальном распределении

Для начала по формуле $z_1 = \frac{x_1 - \overline{x}_в}{\sigma_в}$ находится промежуточное значение для f(z), которая для ячейки W3 выглядит так: '=(O3-$M$28)/$Q$27', где $M$28 – выборочная средняя, а $Q$27 – СКО. После чего в соответствующие ячейки (рис.7) записывается формула, которая для ячейки X3 имеет следующий вид: '=НОРМ.РАСП(W3;0;1;0)'. И наконец по формуле $n_1' = \frac{h \cdot n}{\sigma_в} \cdot f(z_1)$ находятся теоретические частоты. Для Y3 она выглядит так: '=$C$24*$Q$11*X3/$Q$27'. Напоминаю, что $C$24 – это длина интервала, а $Q$24 – число данных. В результате получилась таблица, изображенная на рисунке 7.

z	f(z)	n'
-1,6949	0,0949	3,39
-1,1002	0,2178	7,77
-0,5055	0,3511	12,53
0,0892	0,3974	14,18
0,6839	0,3158	11,27
1,2786	0,1762	6,29
1,8733	0,0690	2,46
2,4680	0,0190	0,68
		58,55

Рисунок 7 – Таблица промежуточных значений и теоретических частот

Для наглядности, строится график (рис.8) для сравнения теоретических и эмпирических частот в виде наложения графика теоретических частот на полигон эмпирических частот.

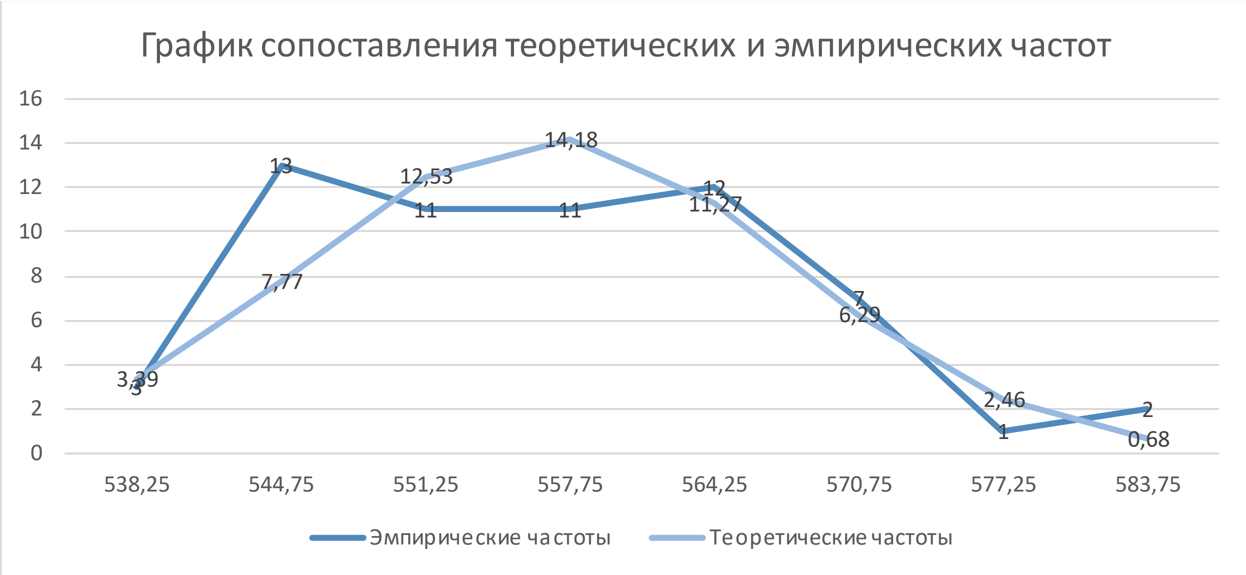

Рисунок 8 – График сопоставления теоретических и эмпирических частот

Далее следует объединить интервалы с частотой ниже 5, в данном случае 1 и 2, 7 и 8. По формуле $\frac{(n_i - n_i')^2}{n_i'}$, для каждого интервала находится хи-квадрат, и в результате получается таблица на рисунке 9. Для ячейки Q33 записывается формула '=СТЕПЕНЬ(O33-P33;2)/P33', где O33 – это эмпирическая частота для первого интервала после объединения, а P33 – теоретическая.

Интервал			n	n'	хи-квадрат
(-∞)	-	548	16	11,16	2,10
548	-	554,5	11	12,53	0,19
554,5	-	561	11	14,18	0,71
561	-	567,5	12	11,27	0,05
567,5	-	574	7	6,29	0,08
574	-	(+∞)	3	3,14	0,01
Сумма:			60	58,55	3,14

Рисунок 9 – Диаграмма сопоставления теоретических и эмпирических частот

Чтобы проверить гипотезу находим степень свободы k по формуле: k = m – r – 1, где m – число интервалов, а r – число параметров. Оценивается два параметра – n и n', поэтому r = 2. Число интервалов теперь равно 6, поэтому k = 6 – 2 – 1 = 3. Уровень значимости по варианту равен 0,05. В соответствующую ячейку вписывается формула: '=ХИ2.ОБР.ПХ($C$25;Q28)', и получается $\chi^2_{набл} < \chi^2_{кр}$.

теоретический хи-квадрат = 7.81. Т.к. при гипотезе не отвергается, а 3.14 < 7.81, то гипотеза о нормальном распределении генеральной совокупности не отвергается.

7. Доверительный интервал для математического ожидания

7.1. При известной дисперсии

Чтобы найти доверительный интервал для математического ожидания α при известной дисперсии, необходимо вычислить точность оценки по формуле $\delta = \frac{t_\gamma \sigma}{\sqrt{n}}$, где t – коэффициент доверия, который отыскивается из функции Лапласа $2\Phi(t_\gamma) = \gamma$, , поэтому в сответствующую ячейку записывается формула ‘=$C$26/2’, где C26 – это надёжность. По варианту надёжность равна 0.8, поэтому Ф(t) = 0.400. t находится по формуле ‘=НОРМСТОБР($T$27+0,5)’, где T27 – Ф(t). Для Ф(t) = 0.400, t = 1.28. В ячейку для точности оценки записывается формула ‘=T28*Q27/КОРЕНЬ(Q11)’, где Q27 – это СКО, T28 – t, а Q11 – число данных. $\delta = \frac{1,28*10,930}{\sqrt{60}} = 1.808$. Границы интервала для математического ожидания равны разности и суммы выборочной средней и точности оценки. В результате получается следующий интервал: 554,967 < α < 558,583, т.к. выборочная средняя равна 556.775.

7.2. При неизвестной дисперсии

Для нахождения этого же интервала для неизвестной дисперсии стоит воспользоваться формулой для EXCEL – ‘=СРЗНАЧ(A3:C22)-ДОВЕРИТ.СТЬЮДЕНТ(0,03;СТАНДОТКЛОН.В(A3:C22);Q11)’, где A3:C22 – исходные данные. В результате получился интервал 553,643 < α < 560,057.

8. Таблица со всеми отдельно найденными значениями

Мода: 544,750	Дисперсия: 119,462	хи-квадрат: 7,815
Медиана: 555,000	СКО: 10,930	Ф(t): 0,400
Средняя: 556,775	Степень свободы k: 3	t: 1,28
Точность оценки: 1,808	Доверительный интервал с известной дисперсией: 554,967	< α < 558,583
Исправленная дисперсия: 121,487	Доверительный интервал с неизвестной дисперсией: 553,643	< α < 560,057

Рисунок 10 – Таблица со всеми отдельно найденными значениями