Математическая статистика.

Индивидуальное задание. Вариационный ряд.

Выполнил студент 2-го курса ИИТТО, факультета ИВТ-2, Фролов Андрей Алексеевич.

Дан ряд распределения зарплат рабочих в определенном предприятии (n=100):

45	38	42	51	47	53	44	56	50	39
58	61	49	66	52	44	37	63	55	60
52	55	48	61	57	59	46	62	68	41
64	53	45	59	48	70	54	48	56	42
67	71	58	63	49	65	54	60	43	69
47	55	72	50	57	62	46	65	51	58
73	44	59	68	52	49	61	55	47	63
58	66	53	42	70	56	48	64	57	45
62	51	48	69	54	60	43	65	59	52
50	67	55	46	71	58	49	63	56	41

Задание 1. Построить интервальный вариационный ряд, результат изобразить графически.

1) Первым делом получаем значения Xmin – минимальный элемент ряда и Xmax – максимальный элемент ряда.
2) Считаем количество интервалов и шаг по формулам:

$$\boldsymbol{k = 1 + (3,322 * \lg(n))}, \quad (1)$$

$$\boldsymbol{h = (Xmax - Xmin)k}, \quad (2)$$

Должны получиться следующие промежуточные результаты:

Xmin	37
Xmax	73
n	100
k	8
h	4,5

3) Делим на интервалы, первый интервал [Xmin, Xmin+h], второй [Xmin+h, Xmin+h+h] и т. д., для каждого интервала считаем количество попавших элементов, левую границу включаем, правую нет, но для последнего интервала включаем обе границы.
Пример формулы: *=СЧЁТЕСЛИМН(B3:K12;">=37";B3:K12;"<41,5")*

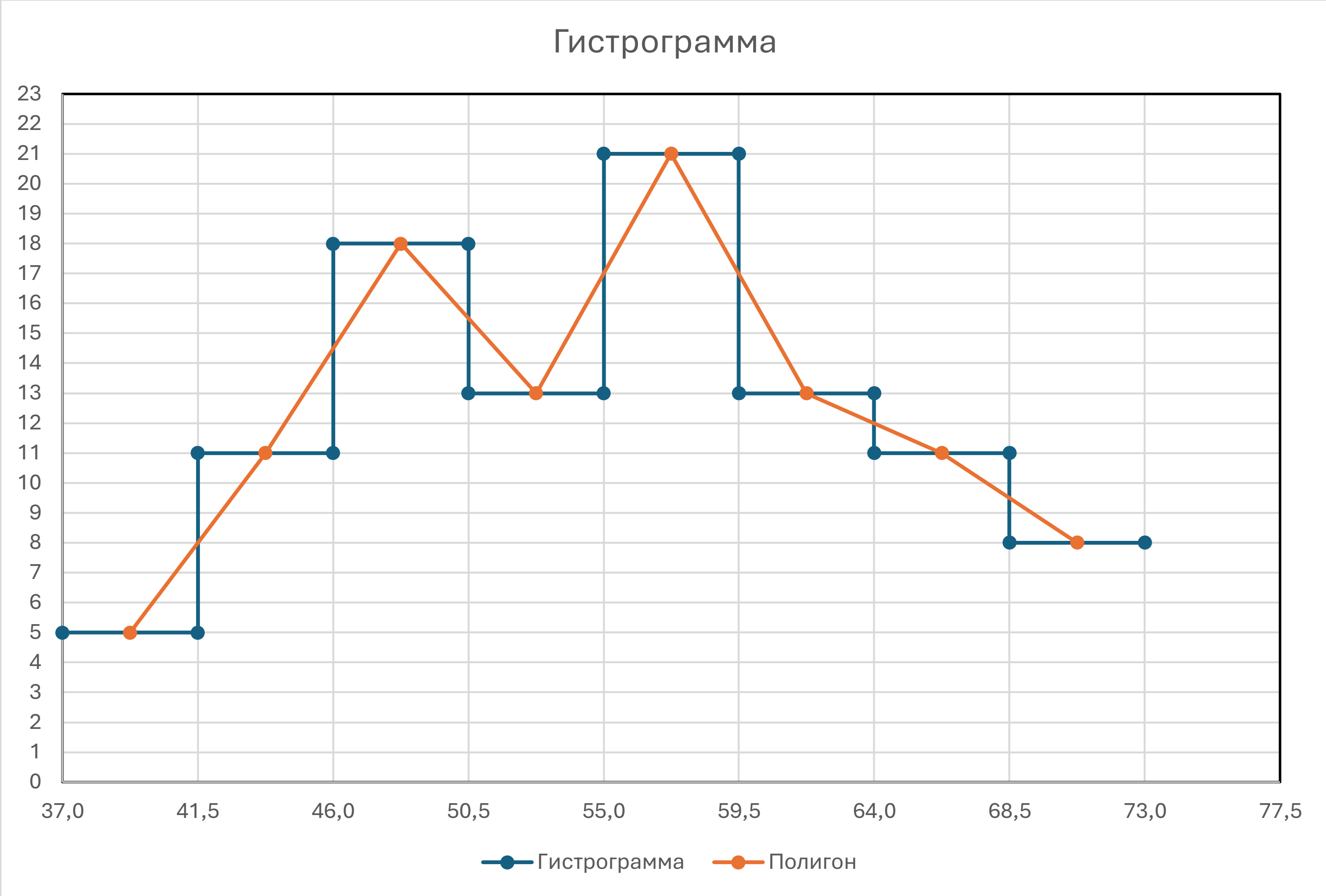
X	37,0	41,5	41,5	46,0	46,0	50,5	50,5	55,0	55,0	59,5	59,5	64,0	64,0	68,5	68,5	73,0
M	5	5	11	11	18	18	13	13	21	21	13	13	11	11	8	8

Для удобства построения графика я предпочел дублировать значение М под каждой границей, что позволить использовать точечный график для построения **гистограммы**, в таком случае формулу можно прописать только в одной ячейке, а во вторую просто занести значение.

4) Так же составим таблицу со средними значениями интервалов, для построения полигона распределения.

Xcp	39,25	43,75	48,25	52,75	57,25	61,75	66,25	70,75
M	5	11	18	13	21	13	11	8

5) По данным первой таблицы строим точечный график, а затем добавляем данные для полигона из второй таблицы. Получаем гистограмму и полигон на одном графике:



Задание 2. Вычислить числовые характеристики вариационного ряда.

1. Среднее значение признака

Вычисляется по формуле:

$$\overline{\boldsymbol{X}} = \frac{\mathbf{1}}{\boldsymbol{n}} * \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{k}} \boldsymbol{x}_{\text{ср}} \boldsymbol{m}_{\boldsymbol{i}} \,, \tag{3}$$

где $x_{\text{ср}}$- среднее значение интервала

Для вычисления используем формулу = СУММПРОИЗВ() и следующую таблицу:

Xср	39,25	43,75	48,25	52,75	57,25	61,75	66,25	70,75
M	5	11	18	13	21	13	11	8

Получаем результат $\overline{\boldsymbol{X}} \approx 55{,}3$

Вывод:

Средняя зарплата рабочих составляет примерно **55,3 единицы**. Это усреднённое значение показывает общий уровень зарплат и служит центральной характеристикой распределения.

2. Дисперсия и среднее квадратичное отклонение

Дисперсия вычисляется по формуле:

$$S^2 = \frac{1}{n} * \sum_{i=1}^{n} (x_i - x_{\text{ср}})^2 , \qquad (4)$$

Среднее квадратичное отклонения вычисляется по формуле: $S = \sqrt{S^2}$.

Будем использовать формулу =СУММКВРАЗН(), но так как она принимает два равных массива значений (равные матрицы), достроим таблицу 10x10 состоящую из среднего значения:

55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3
55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3
55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3
55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3
55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3
55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3
55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3
55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3
55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3
55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3	55,3

Получаем значения: $S^2 \approx 77,3$, $S \approx 8,79$

Вывод:

Отклонения индивидуальных зарплат от среднего достаточно заметные — около **9 единиц**.

Дисперсия и СКО показывают, что **вариация зарплат умеренная**, но не слишком большая: значения относительно стабильно сосредоточены вокруг среднего.

3. Коэффициент вариации

Вычисляется по формуле:

$$V = \frac{\overline{x}}{S} * 100\%, \qquad (5)$$

Получаем значение: $V \approx 15,9$ %

Вывод:

Уровень относительной изменчивости — **низкий (до 20%)**.
Следовательно, зарплаты на предприятии можно считать **достаточно однородными**, без резких перепадов.

4. Промежуток колебания

Отражает колебания плотности. Для этого вычислим плотность каждого интервала по формуле: $\frac{m}{h}$

Дополним этими данными уже имеющуюся таблицу:

Xср	39,25	43,75	48,25	52,75	57,25	61,75	66,25	70,75
M	5	11	18	13	21	13	11	8
Плотн	1,1	2,4	4,0	2,9	4,7	2,9	2,4	1,8

Получаем промежуток колебания плотности **от 1,1 до 4,7**.

Вывод: Плотность распределения в разных интервалах меняется почти в 4 раза.
То есть количество рабочих в интервалах распределено **неравномерно**, есть интервалы, где рабочих значительно больше.

5. Коэффициент асимметрии

Вычисляется по формуле:

$$A_s = \frac{\sum m(x_i - \overline{x})^3}{n * s^3}, \qquad (6)$$

Для вычисления этой величины удобнее всего будет построить вспомогательную таблицу:

Xср	Ср.Ар	M	Xср-Ср.А	4ст^3	5ст*M
39,25	55,3	5	-16,05	-4134,52	-20672,6
43,75	55,3	11	-11,55	-1540,8	-16948,8
48,25	55,3	18	-7,05	-350,403	-6307,25
52,75	55,3	13	-2,55	-16,5814	-215,558
57,25	55,3	21	1,95	7,414875	155,7124
61,75	55,3	13	6,45	268,3361	3488,37
66,25	55,3	11	10,95	1312,932	14442,26
70,75	55,3	8	15,45	3687,954	29503,63

СУММ(последний столбец) – это числитель, далее пишем общую формулу вычисляем значение.
Получаем результат: $A_s \approx 0,05$

Вывод:

Асимметрия почти нулевая.
Распределение **практически симметричное**, без выраженного «хвоста» вправо или влево. Это сближает его с нормальным.

6. Эксцесс

Вычисляется по формуле:

$$E = \frac{\sum m(x_i - \overline{x})^4}{n * s^4} - 3, \qquad (7)$$

Аналогично с коэффициентом асимметрии строим вспомогательную таблицу и затем подставляем все в итоговую формулу и вычисляем значение.

Xср	Ср.Ар	M	Xср-Ср.А	4ст^4	5ст*M
39,25	55,3	5	-16,05	66359,05	331795,2
43,75	55,3	11	-11,55	17796,23	195758,5
48,25	55,3	18	-7,05	2470,339	44466,09
52,75	55,3	13	-2,55	42,28251	549,6726
57,25	55,3	21	1,95	14,45901	303,6391
61,75	55,3	13	6,45	1730,768	22499,98
66,25	55,3	11	10,95	14376,61	158142,7
70,75	55,3	8	15,45	56978,88	455831,1

Получаем результат: $\boldsymbol{E \approx -0,98}$

Вывод:

Отрицательный эксцесс показывает, что распределение **более плоское**, чем нормальное. То есть значения зарплат **более растянуты** по интервалам, меньше выражен пик в центре.

Итоговая таблица с результатами:

Средняя арифметическая	55,3
Дисперсия	77,3
Ср. кв. отклонение	8,79
Коэфф. Вариации, %	15,9
Колебание плотности	от 1,1 до 4,7
Коэфф Ассиметрии	0,051
Эксесс	-0,98

Задание 3. Проверить данные на совпадение закону нормального распределения.

1. Из исходных данных берем среднее значение и среднее квадратическое отклонение.
2. Для каждой варианты вычисляем величину $\mathbf{t_i}$

$$\boldsymbol{t_i = \frac{x_i - x_{\text{ср}}}{\delta}}, \qquad (\mathbf{8})$$

3. Для каждого $\mathbf{t_i}$ находим значение функции:

$$\boldsymbol{\phi(t) = \frac{1}{\sqrt{2\pi}} * e^{-\frac{t^2}{2}}}, \qquad (\mathbf{9})$$

4. Определяем теоретические частоты f_m

$$\boldsymbol{f_m = \phi(t) * \frac{Nd}{\delta}}, \qquad (10)$$

Где N – объем совокупности (сумма всех эмпирических частот), d – длинна интервала.

5. Вычисляем накопленные эмпирические и теоретические частоты.
6. Вычисляем D_i по формуле $\boldsymbol{D_i = |F_i - F_m|}$
7. Из вычисленных данных находим **Dmax.**
8. Расчитываем величину $\boldsymbol{\lambda}$

$$\lambda = \frac{Dmax}{\sqrt{N}}$$

9. По таблице приложения 1 находим вероятность $P(\boldsymbol{\lambda})$ того, что исходные данные нормально распределены. Вероятность изменяется от 0 до 1.

Таблица:

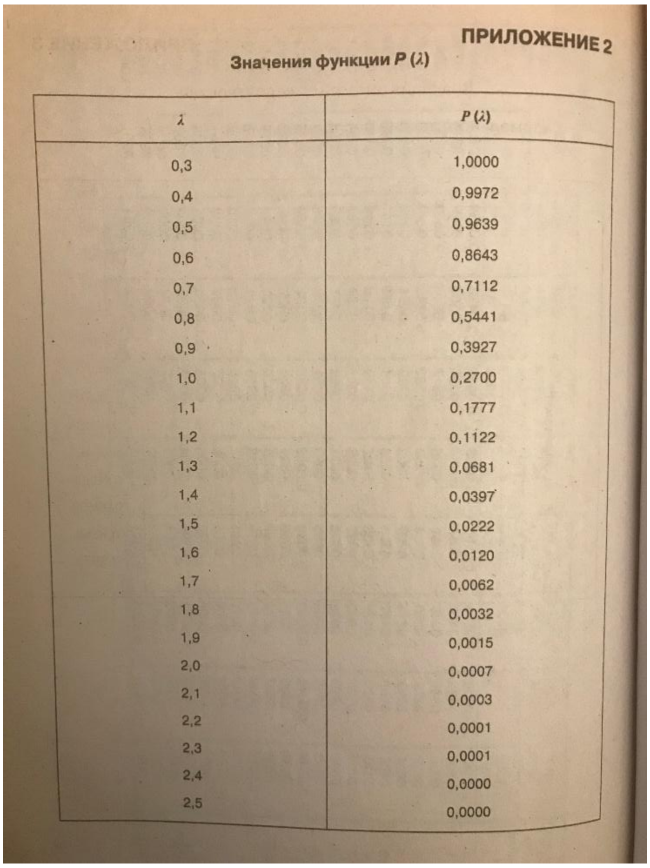
p	3,1415
e	2,7182

Интервал	Кол-во f_i	Середина интервала x_i	ср знач	(x_i - ср знач)^2	t_i	фи(t)	Теор. Част. fm	Накопл. Эмпир. Част. Fi	Накопл. Теор. Част. Fm	D_i
37-41,5	5	39,25	55,3	257,6025	-1,83	0,0753	4	5	4	1
41,5-46	11	43,75	55,3	133,4025	-1,31	0,1683	9	16	12	4
46-50,5	18	48,25	55,3	49,7025	-0,80	0,2892	15	34	27	7
50,5-55	13	52,75	55,3	6,5025	-0,29	0,3825	20	47	47	0
55-59,5	21	57,25	55,3	3,8025	0,22	0,3893	20	68	67	1
59,5 - 64	13	61,75	55,3	41,6025	0,73	0,3048	16	81	82	1
64 - 68,5	11	66,25	55,3	119,9025	1,25	0,1836	9	92	92	0
68,5-73	8	70,75	55,3	238,7025	1,76	0,0851	4	100	96	4
Итого	100						96			

D_max	7
alpha	0,67

ср знач	55,3
ср отклон	8,79

Приложение 1:



ПРИЛОЖЕНИЕ 2

Значения функции $P(\lambda)$

λ	$P(\lambda)$
0,3	1,0000
0,4	0,9972
0,5	0,9639
0,6	0,8643
0,7	0,7112
0,8	0,5441
0,9	0,3927
1,0	0,2700
1,1	0,1777
1,2	0,1122
1,3	0,0681
1,4	0,0397
1,5	0,0222
1,6	0,0120
1,7	0,0062
1,8	0,0032
1,9	0,0015
2,0	0,0007
2,1	0,0003
2,2	0,0001
2,3	0,0001
2,4	0,0000
2,5	0,0000

Вывод:

Вероятность того, что наблюдаемое распределение **не противоречит** нормальному — около 70%.
То есть данные **достаточно хорошо согласуются с нормальным распределением**, хотя совпадение не идеальное.