

ПРИМЕР ВЫПОЛНЕНИЯ ДОМАШНЕЙ КОНТРОЛЬНОЙ РАБОТЫ

НОРМАЛЬНОЕ РАСТРЕДЕЛЕНИЕ

1,60	2,63	2,57	3,56	0,90	3,05	2,01	1,44	2,08	2,45
2,62	2,19	2,42	3,94	2,28	2,01	3,16	3,03	2,82	2,34
3,18	3,40	2,25	1,44	2,44	1,08	0,73	1,49	2,70	3,39
2,76	4,38	2,05	3,31	1,90	2,74	1,85	2,68	2,79	2,37
2,40	1,90	3,47	3,19	2,76	2,49	3,40	1,80	2,60	1,60

По заданной выборке из 50 значений:

1. Построить интервальный вариационный ряд в виде:

Номер интервала	Нижняя граница интервала	Верхняя граница интервала	Середина интервала	Частота	Накопленная частота
-----------------	--------------------------	---------------------------	--------------------	---------	---------------------

2. Построить гистограмму и ящик с усами. Сделать предварительный вывод о нормальности распределения.
3. Вычислить среднюю арифметическую \bar{x} и стандартное отклонение σ .
4. Проверить гипотезу о нормальном распределении с помощью критерия Жарка-Бера на уровне значимости $\alpha = 0,05$.
5. В случае нормальности интерпретировать среднюю арифметическую \bar{x} и интервал $\bar{x} \pm 2 \cdot \sigma$. В случае ненормальности – медиану и квартильный размах.
6. Разбить выборку на две части по 25 наблюдений (в соответствии с разделительной линией в данных). Вычислить коэффициент корреляции (Пирсона или Спирмена – в зависимости от результатов 4-го задания) между выборками. Сделать вывод о тесноте связи. Построить диаграмму рассеяния.

Решение

1. Построить интервальный вариационный ряд.

Объем выборки $N = 50$.

Число интервалов определим по формуле Стерджесса:

$$n = 1 + 3,322 \cdot \lg N = 1 + 3,322 \cdot \lg 50 \approx 7.$$

$$\text{Длина интервала: } h = \frac{R}{n-1} = \frac{x_{\max} - x_{\min}}{n-1} = \frac{4,38 - 0,73}{7-1} \approx 0,61 \approx 0,6.$$

$$\text{Границы первого интервала: } \left(x_{\min} - \frac{h}{2}; x_{\min} + \frac{h}{2} \right) = \left(0,73 - \frac{0,6}{2}; 0,73 + \frac{0,6}{2} \right) = (0,43; 1,03).$$

Интервальный вариационный ряд представлен в Табл. 1.

Табл. 1. Интервальный вариационный ряд

Номер интервала	Нижняя граница интервала	Верхняя граница интервала	Середина интервала x_i	Частота ν_i	Накопленная частота
1	0,43	1,03	0,73	2	2
2	1,03	1,63	1,33	6	8
3	1,63	2,23	1,93	9	17
4	2,23	2,83	2,53	20	37
5	2,83	3,43	3,13	9	46
6	3,43	4,03	3,73	3	49
7	4,03	4,63	4,33	1	50
Сумма:	—	—	—	50	—

2. Построить гистограмму и ящик с усами.

Гистограмма, построенная по данным Табл. 1, представлена на Рис. 1.

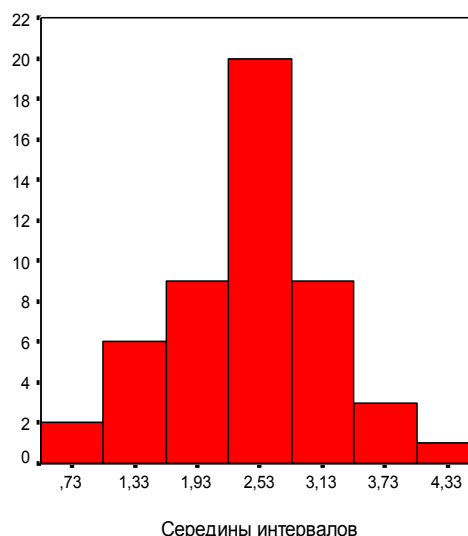


Рис. 1. Гистограмма

Для построения ящика с усами найдем медиану и квартили:

1) медианный интервал – 4-ый, т.к. накопленная частота (37) первая превышает половину общей суммы частот (25)

$$Me = x_{Me} + h \cdot \frac{\frac{1}{2} \cdot \sum v_i - S_{Me-1}}{v_{Me}} = 2,23 + 0,6 \cdot \frac{\frac{1}{2} \cdot 50 - 17}{20} = 2,47$$

2) интервал, содержащий нижнюю квартиль – 3-ий, т.к. накопленная частота (17) первая превышает одну четвертую общей суммы частот (12,5)

$$Q_n = x_{Q_n} + h \cdot \frac{\frac{1}{4} \cdot \sum v_i - S_{Q_n-1}}{v_{Q_n}} = 1,63 + 0,6 \cdot \frac{\frac{1}{4} \cdot 50 - 8}{9} = 1,93$$

3) интервал, содержащий верхнюю квартиль – 5-ый, т.к. накопленная частота (46) первая превышает три четвертых общей суммы частот (37,5)

$$4) Q_6 = x_{Q_6} + h \cdot \frac{\frac{3}{4} \cdot \sum v_i - S_{Q_6-1}}{v_{Q_6}} = 2,83 + 0,6 \cdot \frac{\frac{3}{4} \cdot 50 - 37}{9} = 2,86$$

Построенный по вычисленным значениям ящик с усами представлен на Рис. 2.

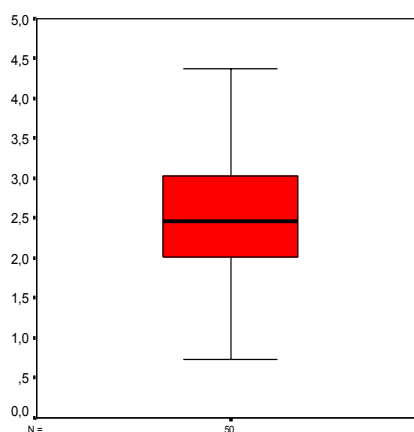


Рис. 2. Ящик с усами

Вывод. По анализу гистограммы и ящика с усами можно сделать предварительный вывод о том, что данные распределены нормально.

3. Вычислить среднюю арифметическую \bar{x} и стандартное отклонение σ . Составим расчетную таблицу (Табл. 2).

Табл. 2. Расчетная таблица

Номер интервала	Середина интервала x_i	Частота v_i	$x_i v_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 v_i$
1	0,73	2	1,46	-1,69	2,86	5,71
2	1,33	6	7,98	-1,09	1,19	7,13
3	1,93	9	17,37	-0,49	0,24	2,16
4	2,53	20	50,60	0,11	0,01	0,24
5	3,13	9	28,17	0,71	0,50	4,54
6	3,73	3	11,19	1,31	1,72	5,15
7	4,33	1	4,33	1,91	3,65	3,65
Сумма:	—	50	121,1	—	—	28,58

$$\text{Средняя арифметическая: } \bar{x} = \frac{\sum x_i v_i}{\sum v_i} = \frac{121,1}{50} \approx 2,42.$$

$$\text{Стандартное отклонение: } \sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2 v_i}{N - 1}} = \sqrt{\frac{28,58}{50 - 1}} = 0,76.$$

4. Проверить гипотезу о нормальном распределении с помощью критерия Жарка-Бера на уровне значимости $\alpha = 0,05$.

Для начала вычислим асимметрию и эксцесс. Будем производить вычисления по действиям:

$$1) \tilde{\sigma} = \sigma \cdot \sqrt{\frac{N-1}{N}} = 0,76 \cdot \sqrt{\frac{50-1}{50}} = 0,75$$

- 2) Составим расчетную таблицу для вычисления асимметрии и эксцесса:

№	Середина интервала x_i	Частота v_i	$x_i - \bar{x}$	$\frac{x_i - \bar{x}}{\tilde{\sigma}}$	$v_i \cdot \left(\frac{x_i - \bar{x}}{\tilde{\sigma}}\right)^3$	$v_i \cdot \left(\frac{x_i - \bar{x}}{\tilde{\sigma}}\right)^4$
1	0,73	2	-1,69	-2,25	-22,88	51,56
2	1,33	6	-1,09	-1,45	-18,42	26,77
3	1,93	9	-0,49	-0,65	-2,51	1,64
4	2,53	20	0,11	0,15	0,06	0,01
5	3,13	9	0,71	0,95	7,64	7,23
6	3,73	3	1,31	1,75	15,99	27,92
7	4,33	1	1,91	2,55	16,52	42,06
Сумма:	—	50	—	—	-3,6	157,19

$$Sk = \frac{1}{N} \cdot \sum \left[v_i \cdot \left(\frac{x_i - \bar{x}}{\tilde{\sigma}}\right)^3 \right] = \frac{1}{50} \cdot (-3,6) = -0,072,$$

$$K = \frac{1}{N} \cdot \sum \left[v_i \cdot \left(\frac{x_i - \bar{x}}{\tilde{\sigma}}\right)^4 \right] = \frac{1}{50} \cdot 157,19 = 3,144.$$

- 3) Выдвинем гипотезу H_0 о нормальном распределении выборки.

- 4) Вычислим фактическое значение критерия Жарка-Бера:

$$J - B = \frac{N}{6} \cdot \left(Sk^2 + \frac{(K-3)^2}{4} \right) = \frac{50}{6} \cdot \left((-0,072)^2 + \frac{(3,144-3)^2}{4} \right) = 0,086$$

- 5) $J - B = 0,086 < 5,991$, значит гипотеза H_0 о нормальном распределении выборки принимается, т.е. распределение является нормальным.

5. Так как было выяснено, что выборка имеет нормальное распределение, то будем интерпретировать среднюю арифметическую \bar{x} и интервал $\bar{x} \pm 2 \cdot \sigma$:

1) половина выборки меньше 2,42, а половина - больше.

2) 95% значений выборки лежат в интервале:

$$(\bar{x} - 2 \cdot \sigma; \bar{x} + 2 \cdot \sigma) = (2,42 - 2 \cdot 0,76; 2,42 + 2 \cdot 0,76) = (0,9; 3,94).$$

6. Разбить выборку на две части по 25 наблюдений (в соответствии с разделительной линией в данных). Вычислить коэффициент корреляции (Пирсона или Спирмена – в зависимости от результатов 4-го задания) между выборками. Сделать вывод о тесте связи. Построить диаграмму рассеяния.

Диаграмма рассеяния представлена на Рис. 3. По рисунку – связь не сильная, обратная.

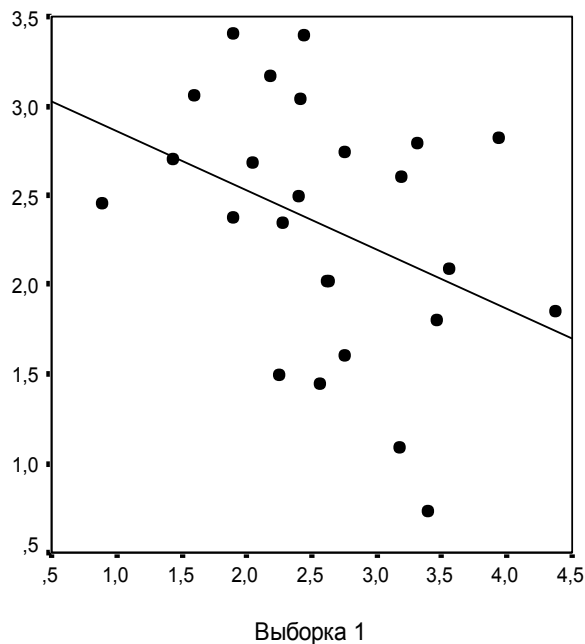


Рис. 3. Диаграмма рассеяния

По результатам задания 4, обе выборки извлечены из нормально распределенной совокупности, поэтому вычислим коэффициент корреляции Пирсона. Для этого составим расчетную таблицу (Табл. 3).

Табл. 3. Расчетная таблица

№	Выборка 1 (x)	Выборка 2 (y)	xy	$(x - \bar{x})^2$	$(y - \bar{y})^2$
1	1,60	3,05	4,88	1,04	0,53
2	2,62	2,01	5,27	0	0,10
3	3,18	1,08	3,43	0,31	1,55
4	2,76	2,74	7,56	0,02	0,17
5	2,40	2,49	5,98	0,05	0,03
6	2,63	2,01	5,29	0,0001	0,10
7	2,19	3,16	6,92	0,19	0,70
8	3,40	0,73	2,48	0,61	2,54
9	4,38	1,85	8,10	3,09	0,22
10	1,90	3,40	6,46	0,52	1,16
11	2,57	1,44	3,70	0,003	0,78
12	2,42	3,03	7,33	0,04	0,50
13	2,25	1,49	3,35	0,14	0,70
14	2,05	2,68	5,49	0,33	0,13
15	3,47	1,80	6,25	0,72	0,27
16	3,56	2,08	7,40	0,88	0,06
17	3,94	2,82	11,11	1,74	0,25
18	1,44	2,70	3,89	1,40	0,14
19	3,31	2,79	9,23	0,47	0,22
20	3,19	2,60	8,29	0,32	0,08
21	0,90	2,45	2,21	2,96	0,02
22	2,28	2,34	5,34	0,12	0,0003
23	2,44	3,39	8,27	0,03	1,14
24	1,90	2,37	4,50	0,52	0,002
25	2,76	1,60	4,42	0,02	0,52
Сумма	65,54	58,1	147,15	15,52	11,91

Средние значения:

$$\bar{x} = \frac{\sum x_i}{N} = \frac{65,54}{25} = 2,62, \quad \bar{y} = \frac{\sum y_i}{N} = \frac{58,1}{25} = 2,32, \quad \overline{xy} = \frac{\sum x_i y_i}{N} = \frac{147,15}{25} = 5,89.$$

Стандартные отклонения:

$$\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N-1}} = \sqrt{\frac{15,52}{25-1}} = 0,80, \quad \sigma_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{N-1}} = \sqrt{\frac{11,91}{25-1}} = 0,70.$$

Корреляция Пирсона:

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} = \frac{5,89 - 2,62 \cdot 2,32}{0,80 \cdot 0,70} = -0,336.$$

Связь слабая, обратная.