ПРИМЕР ВЫПОЛНЕНИЯ ДОМАШНЕЙ КОНТРОЛЬНОЙ РАБОТЫ

НЕНОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

5	2,12	2,2	2,22	2,69	3,77	2,79	2,51	2,99	2,14
2,28	2,34	2,3	2,26	3,58	2,92	2,54	2,44	2,37	2,17
2,46	2,99	2,86	2,76	3,84	2,72	3,49	2,42	3,19	2,05
2,21	3,38	2,5	2,55	2,74	2,47	2,5	2,42	4,31	2,11
2,15	3,1	3,19	4,24	3,44	3,14	2,95	2,05	2,29	3,03

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:

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

Длина интервала: $h = \dfrac{R}{n-1} = \dfrac{x_{\max} - x_{\min}}{n-1} = \dfrac{5 - 2,05}{7-1} = 0,5$.

Границы первого интервала: $\left(x_{\min} - \dfrac{h}{2}; x_{\min} + \dfrac{h}{2}\right) = \left(2,05 - \dfrac{0,5}{2}; 2,05 + \dfrac{0,5}{2}\right) = (1,8; 2,3)$.

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	1,8	2,3	2,05	14	14
2	2,3	2,8	2,55	17	31
3	2,8	3,3	3,05	10	41
4	3,3	3,8	3,55	5	46
5	3,8	4,3	4,05	2	48
6	4,3	4,8	4,55	1	49
7	4,8	5,3	5,05	1	50
Сумма:	—	—	—	50	—

2. Построить гистограмму и ящик с усами.

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

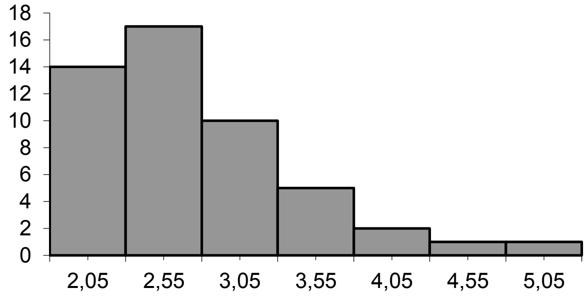

Рис. 1. *Гистограмма*

Для построения ящика с усами найдем медиану и квартили:

1) медианный интервал – 2-ой, т.к. накопленная частота (31) первая превышает половину общей суммы частот (25)

$$Me = x_{Me} + h \cdot \frac{\frac{1}{2} \cdot \sum \nu_i - S_{Me-1}}{\nu_{Me}} = 2,3 + 0,5 \cdot \frac{\frac{1}{2} \cdot 50 - 14}{17} \approx 2,62$$

2) интервал, содержащий нижнюю квартиль – 1-ый, т.к. накопленная частота (14) первая превышает одну четвертую общей суммы частот (12,5)

$$Q_н = x_{Q_н} + h \cdot \frac{\frac{1}{4} \cdot \sum \nu_i - S_{Q_н - 1}}{\nu_{Q_н}} = 1,8 + 0,5 \cdot \frac{\frac{1}{4} \cdot 50 - 0}{14} \approx 2,25$$

3) интервал, содержащий верхнюю квартиль – 5-ый, т.к. накопленная частота (47) первая превышает три четвертых общей суммы частот (37,5)

$$Q_в = x_{Q_в} + h \cdot \frac{\frac{3}{4} \cdot \sum \nu_i - S_{Q_в - 1}}{\nu_{Q_в}} = 2,8 + 0,5 \cdot \frac{\frac{3}{4} \cdot 50 - 31}{10} \approx 3,13$$

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

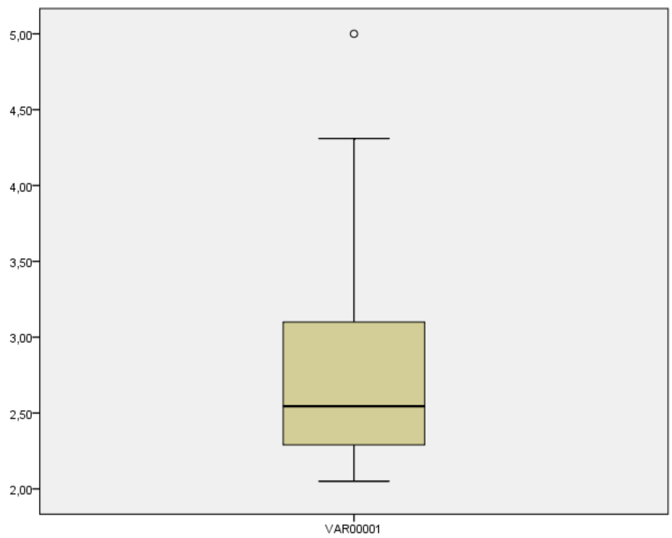

Рис. 2. *Ящик с усами*

Замечание. На Рис. 2 максимальное значение представлено в виде точки, график нарисован с использованием статистического пакета PASW Statistics 18.0. При изображении ящика «от руки» верхний «ус» будет продлен до максимального значения.

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

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 ν_i	$x_i\nu_i$	$x_i - \bar{x}$	$\left(x_i - \bar{x}\right)^2$	$\left(x_i - \bar{x}\right)^2 \nu_i$
1	2,05	14	28,7	-0,71	0,50	7,06
2	2,55	17	43,35	-0,21	0,04	0,75
3	3,05	10	30,5	0,29	0,08	0,84
4	3,55	5	17,75	0,79	0,62	3,12
5	4,05	2	8,1	1,29	1,66	3,33
6	4,55	1	4,55	1,79	3,20	3,20
7	5,05	1	5,05	2,29	5,24	5,24
Сумма:	—	50	138	—	—	23,54

Средняя арифметическая: $\bar{x} = \dfrac{\sum x_i\nu_i}{\sum \nu_i} = \dfrac{138}{50} = 2,76$.

Стандартное отклонение: $\sigma = \sqrt{\dfrac{\sum\left(x_i - \bar{x}\right)^2 \nu_i}{N-1}} = \sqrt{\dfrac{23,54}{50-1}} = 0,69$.

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) $\tilde{\sigma} = \sigma \cdot \sqrt{\dfrac{N-1}{N}} = 0,69 \cdot \sqrt{\dfrac{50-1}{50}} = 0,68$

2) Составим расчетную таблицу для вычисления асимметрии и эксцесса:

№	Середина интервала x_i	Частота ν_i	$x_i - \bar{x}$	$\dfrac{x_i - \bar{x}}{\tilde{\sigma}}$	$\nu_i \cdot \left(\dfrac{x_i - \bar{x}}{\tilde{\sigma}}\right)^3$	$\nu_i \cdot \left(\dfrac{x_i - \bar{x}}{\tilde{\sigma}}\right)^4$
1	2,05	14	-0,71	-1,04	-15,94	16,64
2	2,55	17	-0,21	-0,31	-0,50	0,15
3	3,05	10	0,29	0,43	0,78	0,33
4	3,55	5	0,79	1,16	7,84	9,11
5	4,05	2	1,29	1,90	13,65	25,90
6	4,55	1	1,79	2,63	18,24	48,01
7	5,05	1	2,29	3,37	38,19	128,62
Сумма:	—	50	—	—	62,26	228,76

$$Sk = \frac{1}{N} \cdot \sum\left[\nu_i \cdot \left(\frac{x_i - \bar{x}}{\tilde{\sigma}}\right)^3\right] = \frac{1}{50} \cdot 62,26 = 1,245,$$

$$K=\frac{1}{N}\cdot\sum\left[\nu_i\cdot\left(\frac{x_i-\bar{x}}{\tilde{\sigma}}\right)^4\right]=\frac{1}{50}\cdot 228,76=4,575.$$

3) Выдвинем гипотезу H_0 о нормальном распределении выборки.
4) Вычислим фактическое значение критерия Жарка-Бера:

$$J-B=\frac{N}{6}\cdot\left(Sk^2+\frac{(K-3)^2}{4}\right)=\frac{50}{6}\cdot\left(1,245^2+\frac{(4,575-3)^2}{4}\right)=18,085.$$

5) $J-B=18,085>5,991$, значит, гипотеза H_0 о нормальном распределении выборки отклоняется, т.е. распределение является ненормальным.

5. Так как было выяснено, что выборка имеет ненормальное распределение, то будем интерпретировать медиану и квартильный размах
   1) половина выборки меньше 2,62, а половина – больше.
   2) 50% значений выборки лежат в интервале $(Q_н;Q_в)=(2,25;3,13)$, т.е. в полосе шириной $H=Q_в-Q_н=3,13-2,25=0,88$.
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 тестоне связи. Построить диаграмму рассеяния.

   Диаграмма рассеяния представлена на Рис. 3. По рисунку – связи нет.

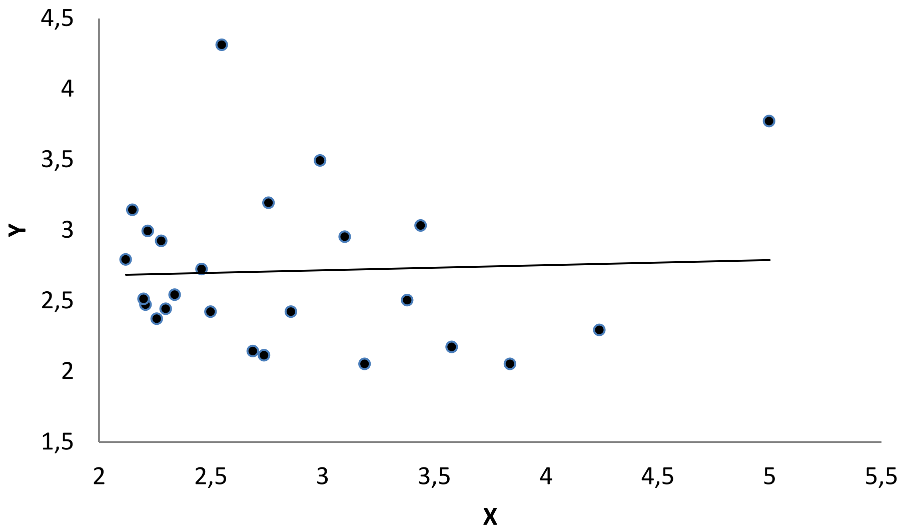

Рис. 3. *Диаграмма рассеяния*

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

Табл. 3. *Расчетная таблица*

№	x	y	Ранг x	Ранг y	d	d^2
1	5	3,77	25	24	1	1
2	2,28	2,92	7	17	-10	100
3	2,46	2,72	10	15	-5	25
4	2,21	2,47	4	11	-7	49
5	2,15	3,14	2	21	-19	361
6	2,12	2,79	1	16	-15	225
7	2,34	2,54	9	14	-5	25
8	2,99	3,49	17	23	-6	36
9	3,38	2,5	20	12	8	64
10	3,1	2,95	18	18	0	0
11	2,2	2,51	3	13	-10	100
12	2,3	2,44	8	10	-2	4
13	2,86	2,42	16	8,5	7,5	56,25
14	2,5	2,42	11	8,5	2,5	6,25
15	3,19	2,05	19	1,5	17,5	306,25
16	2,22	2,99	5	19	-14	196
17	2,26	2,37	6	7	-1	1
18	2,76	3,19	15	22	-7	49
19	2,55	4,31	12	25	-13	169
20	4,24	2,29	24	6	18	324
21	2,69	2,14	13	4	9	81
22	3,58	2,17	22	5	17	289
23	3,84	2,05	23	1,5	21,5	462,25
24	2,74	2,11	14	3	11	121
25	3,44	3,03	21	20	1	1
Сумма:	—	—	—	—	—	3052

Коэффициент корреляции Спирмена: $r_S = 1 - \dfrac{6 \cdot \sum_{i=1}^{N} d_i^2}{N\left(N^2 - 1\right)} = 1 - \dfrac{6 \cdot 3052}{25 \cdot \left(25^2 - 1\right)} = -0{,}174$.

Полученный коэффициент корреляции по абсолютному значению не превышает 0,3, значит, существенной статистической связи между фактором и откликом не наблюдается.