

Задание. Дана выборка

37	34	42	38	31	41	40	35	32	34
37	37	26	39	45	37	40	40	45	31
39	42	47	37	42	40	29	35	40	36
34	33	31	28	37	40	41	41	49	41
37	29	43	43	39	35	42	42	39	50
31	33	38	42	38	35	32	37	45	42
44	34	34	34	38	38	38	30	39	35
42	33	35	31	35	53	48	39	47	41
37	48	41	43	42	29	33	48	39	42
41	41	36	43	37	33	38	43	37	34

1. Составить интервальный статистический ряд. Величину интервалов округлить с точностью до 0,1 в большую сторону.
2. Найти эмпирическую функцию распределения и построить ее график.
3. Построить гистограмму относительных частот. Можно ли предположить, что данная выборка взята из нормального распределения?
4. Определить выборочное среднее и несмещенную оценку дисперсии по сгруппированному статистическому ряду.
5. Записать предполагаемую плотность закона распределения.
6. Проверить по критерию χ^2 Пирсона гипотезу о законе распределения. Уровень значимости принять равным $\alpha = 0,05$.

Контрольные вопросы

1. Что называется выборкой? Что называется объемом выборки?
2. Что называется частотой выборочного значения? Что называется относительной частотой?
3. Как оценить по выборке математическое ожидание и дисперсию наблюдаемой СВ?
4. Как рассчитать несмещенную оценку дисперсии?
5. Как оценить по выборке функцию и плотность распределения наблюдаемой СВ?
6. Что называется эмпирической функцией распределения?
7. Что называется гистограммой относительных частот?
8. Чему равна площадь гистограммы относительных частот?
9. Что называется статистической гипотезой?
10. В каком случае статистическая гипотеза называется простой? сложной?
11. В чем разница между нулевой и альтернативной гипотезами?
12. Что называется уровнем значимости статистического критерия?
13. Что называется критерием согласия?
14. В чем суть критерия согласия χ^2 Пирсона?
15. Какие достоинства и недостатки имеет критерий согласия χ^2 Пирсона?

Методические указания по выполнению лабораторной работы

1. Объем выборки $n = 100$.

Построим интервальный статистический ряд.

Количество интервалов определим по формуле Стерджесса

$$k \approx 1 + \log_2 n = 1 + \log_2 100 = 7,644.$$

Принимаем $k = 8$.

Размах выборки

$$W = x_{\max} - x_{\min} = 53 - 26 = 27.$$

Длина каждого интервала

$$h \approx \frac{W}{k} = \frac{27}{8} = 3,375.$$

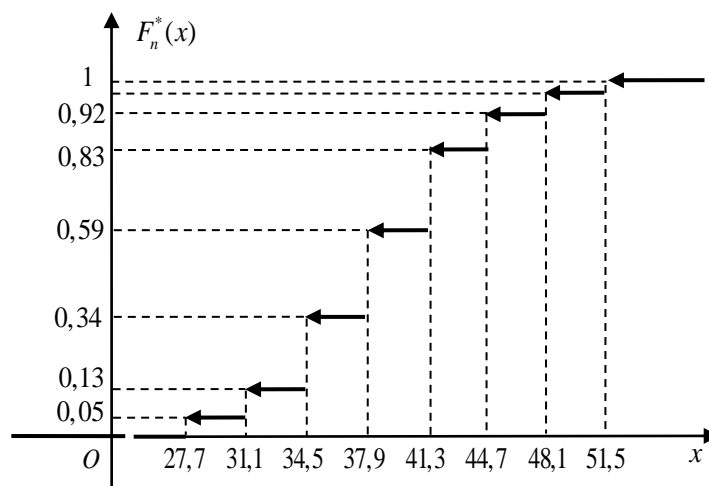
Округлив с точностью до 0,1 в большую сторону, принимаем $h = 3,4$.

Интервальный статистический ряд

$[x_i; x_{i-1})$	x_i^*	n_i	$\frac{n_i}{n}$	$\frac{n_i}{nh}$
[26; 29,4)	27,7	5	0,05	0,015
[29,4; 32,8)	31,1	8	0,08	0,024
[32,8; 36,2)	34,5	21	0,21	0,062
[36,2; 39,6)	37,9	25	0,25	0,074
[39,6; 43)	41,3	24	0,24	0,071
[43; 46,4)	44,7	9	0,09	0,026
[46,4; 49,8)	48,1	6	0,06	0,018
[49,8; 53,2]	51,5	2	0,02	0,006

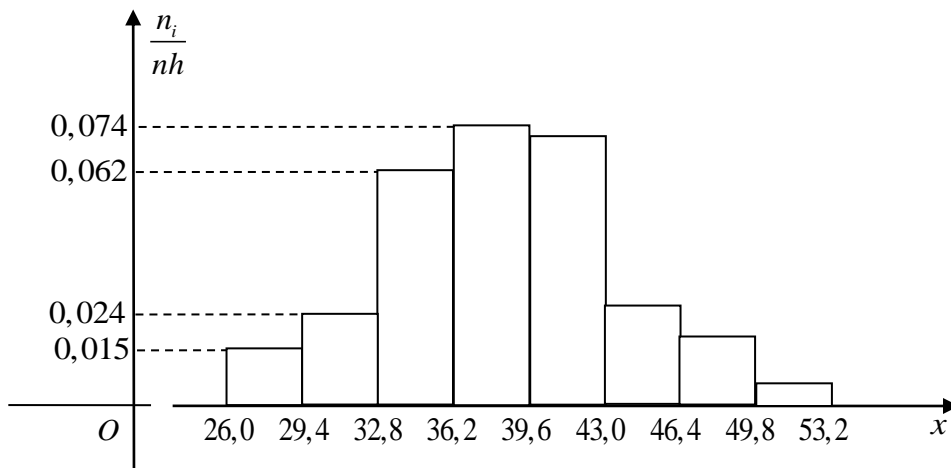
2. Эмпирическая функция распределения:

$$F_n^*(x) = \begin{cases} 0 & \text{при } x \leq 27,7, \\ 0,05 & \text{при } 27,7 < x \leq 31,1, \\ 0,13 & \text{при } 31,1 < x \leq 34,5, \\ 0,34 & \text{при } 34,5 < x \leq 37,9, \\ 0,59 & \text{при } 37,9 < x \leq 41,3, \\ 0,83 & \text{при } 41,3 < x \leq 44,7, \\ 0,92 & \text{при } 44,7 < x \leq 48,1, \\ 0,98 & \text{при } 48,1 < x \leq 51,5, \\ 1 & \text{при } x > 51,5. \end{cases}$$



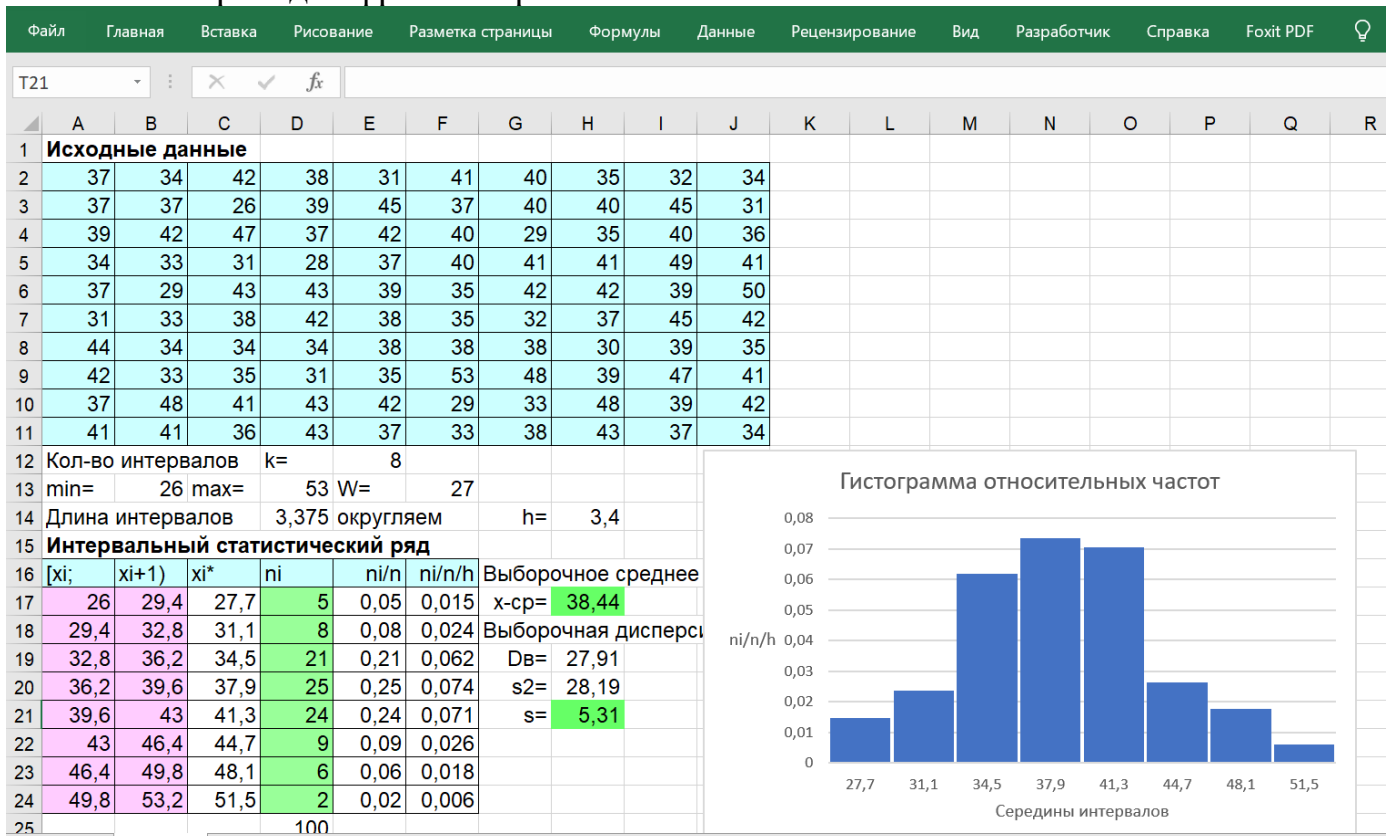
Отметим, что при построении эмпирической функции распределения по интервальному статистическому ряду изменения ее значений (скачки) происходят в точках, соответствующих серединам интервалов группировки.

3. **Гистограмма относительных частот** состоит из прямоугольников шириной $h = 3,4$ и высотой $\frac{n_i}{nh}$



По виду гистограммы можно выдвинуть гипотезу о том, что выборка взята из нормального распределения.

Ниже приведен фрагмент рабочего листа в Excel.



Для построения гистограммы относительных частот, использован *Мастер диаграмм*, который вызывается с помощью команды *Вставка* → *Диаграмма* или при нажатии соответствующей кнопки на панели инструментов.

4. **Рассчитаем оценки параметров** предполагаемого нормального закона распределения по сгруппированному статистическому ряду.

Данный закон содержит два параметра a и σ , которые имеют смысл математического ожидания и среднего квадратического отклонения СВ ξ : $M\xi = a$, $D\xi = \sigma^2$.

В качестве оценок для математического ожидания a и дисперсии σ^2 наблюдаемой случайной величины рассчитаем соответственно выборочное среднее \bar{x} и несмещенную оценку дисперсии s^2 , для вычисления s^2 предварительно найдем выборочную дисперсию D_B :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k x_i^* n_i;$$

$$D_B = \frac{1}{n} \sum_{i=1}^k (x_i^*)^2 n_i - (\bar{x})^2;$$

$$s^2 = \frac{n}{n-1} D_B.$$

Используя интервальный статистический ряд, получим:

$$\bar{x} = \frac{1}{100} \cdot (27,7 \cdot 5 + 31,1 \cdot 8 + 34,5 \cdot 21 + 37,9 \cdot 25 + 41,3 \cdot 24 +$$

$$+ 44,7 \cdot 9 + 48,1 \cdot 6 + 51,5 \cdot 2) \approx 38,44;$$

$$D_B = \frac{1}{100} \cdot (27,7^2 \cdot 5 + 31,1^2 \cdot 8 + 34,5^2 \cdot 21 + 37,9^2 \cdot 25 + 41,3^2 \cdot 24 +$$

$$+ 44,7^2 \cdot 9 + 48,1^2 \cdot 6 + 51,5^2 \cdot 2) - 38,44^2 \approx 27,91;$$

$$s^2 = \frac{100}{99} \cdot 27,91 \approx 28,19.$$

Тогда оценкой для среднего квадратического отклонения σ будет $s = \sqrt{28,19} \approx 5,31$.

5. **Функция плотности нормального закона распределения** имеет вид

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Следовательно, выдвигаем гипотезу о том, что выборка взята из нормального распределения с плотностью

$$f(x) = \frac{1}{5,31\sqrt{2\pi}} e^{-\frac{(x-38,44)^2}{56,38}}.$$

6. **Критерий согласия χ^2 Пирсона.**

Проверяемая гипотеза представляет собой предположение о распределении наблюдаемой СВ и является простой:

H_0 : функция распределения наблюдаемой СВ совпадает с $F(x)$;

\bar{H} : функция распределения наблюдаемой СВ не совпадает с $F(x)$.

Критерий согласия χ^2 Пирсона основан на сравнении эмпирических и теоретических частот попадания СВ в рассматриваемые группы (интервалы):

n_i – **эмпирическая частота** наблюдения значений из интервала $[x_{i-1}; x_i)$;

$np_i = n P(\xi \in [x_{i-1}; x_i)) = n(F(x_i) - F(x_{i-1}))$ – **теоретическое значение соответствующей частоты.**

$$\chi^2_{\text{расч}} = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}.$$

Замечание. Контроль вычислений можно осуществить по формуле

$$\chi^2_{\text{расч}} = \sum_{i=1}^k \frac{n_i^2}{np_i} - n.$$

Пусть r – число неизвестных параметров теоретического распределения, оцененных по выборке. **Сколько параметров характеризуют нормальное распределение?**

Критерий согласия χ^2 Пирсона заключается в следующем:

если $\chi^2_{\text{расч}} < \chi^2_{\alpha; k-r-1}$, где $\chi^2_{\alpha; k-r-1}$ определяется по таблице квантилей распределения χ^2 , то гипотеза H_0 принимается (признается непротиворечащей экспериментальным данным; нет оснований отвергнуть гипотезу H_0) на уровне значимости α , а если $\chi^2_{\text{расч}} \geq \chi^2_{\alpha; k-r-1}$, то гипотеза H_0 отвергается (не согласуется с данными эксперимента).

Проверим с помощью критерия согласия χ^2 Пирсона гипотезу

H_0 : наблюдаемая СВ имеет нормальное распределение с параметрами $a = 38,44, \sigma = 5,31$

при альтернативе

\bar{H} : наблюдаемая СВ имеет другое распределение.

Для расчета статистики критерия Пирсона

$$\chi^2_{\text{расч}} = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}$$

составим новую таблицу, содержащую следующие столбцы:

- интервалы $[x_{i-1}; x_i)$ (при этом крайние интервалы должны быть расширены до $-\infty$ и $+\infty$ соответственно; а интервалы с количеством наблюдений меньше 5 объединяются с соседними);
- n_i – эмпирическая частота наблюдения значений из интервала $[x_{i-1}; x_i)$;
- теоретическая вероятность попадания СВ в интервал $[x_{i-1}; x_i)$, в случае нормального распределения с параметрами $a = 38,44, \sigma = 5,31$

$$p_i = \Phi\left(\frac{x_i - 38,44}{5,31}\right) - \Phi\left(\frac{x_{i-1} - 38,44}{5,31}\right);$$
- np_i – теоретическое значение соответствующей частоты,
- а также столбцы со значениями $n_i - np_i, (n_i - np_i)^2, \frac{(n_i - np_i)^2}{np_i}, \frac{n_i^2}{np_i}$.

Последний столбец используется для контроля вычислений по формуле

$$\chi^2_{\text{расч}} = \sum_{i=1}^k \frac{n_i^2}{np_i} - n.$$

Интервалы	n_i	p_i	np_i	$n_i - np_i$	$(n_i - np_i)^2$	$\frac{(n_i - np_i)^2}{np_i}$	$\frac{n_i^2}{np_i}$
$[-\infty; 29,4)$	5	0,0443	4,425	0,575	0,33	0,075	5,649
$[29,4; 32,8)$	8	0,0996	9,964	-1,96	3,858	0,387	6,423
$[32,8; 36,2)$	21	0,1924	19,24	1,761	3,103	0,161	22,92
$[36,2; 39,6)$	25	0,2499	24,99	0,011	0,0001	0,000	25,01
$[39,6; 43)$	24	0,2184	21,84	2,16	4,668	0,214	26,37
$[43; 46,4)$	9	0,1284	12,84	-3,84	14,76	1,149	6,308
$[46,4; +\infty)$	8	0,067	6,701	1,299	1,686	0,252	9,55
Суммы	100	1	100			$\chi^2_{\text{расч}} = 2,2376$	102,24

$$\chi^2_{\text{расч}} = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} \approx 2,24.$$

Сумма элементов последнего столбца – $\sum_{i=1}^k \frac{n_i^2}{np_i} \approx 102,24$. Контроль вычислений:

$$\chi^2_{\text{расч}} = \sum_{i=1}^k \frac{n_i^2}{np_i} - n = 102,24 - 100 = 2,24.$$

Определим критическое значение $\chi^2_{\text{крит}} = \chi^2_{\alpha; k-r-1}$, где $\alpha = 0,05$ – заданный уровень значимости; $k = 7$ – число интервалов после объединения малочисленных групп с соседними; $r = 2$, поскольку при расчете теоретических вероятностей p_i использовались две полученные по выборке оценки \bar{x} и s параметров нормального распределения. По таблице квантилей распределения χ^2 получаем

$$\chi^2_{\text{крит}} = \chi^2_{0,05; 4} = 9,4877.$$

Таким образом, $\chi^2_{\text{расч}} = 2,24 < \chi^2_{\text{крит}} = 9,4877$, поэтому на уровне значимости $\alpha = 0,05$ нет оснований отвергнуть гипотезу H_0 , согласно которой выборка взята из нормального распределения с параметрами $a = 38,44$, $\sigma = 5,31$.

Указание. Для вычисления вероятностей p_i использована функция, возвращающая нормальную функцию распределения для указанного среднего и стандартного отклонения:

НОРМРАСП(х;среднее; стандартное_откл;интегральная),

где x — значение, для которого строится распределение;

среднее — среднее выборочное;

стандартное_откл — стандартное отклонение распределения;

интегральная — логическое значение, определяющее форму функции.

Если интегральная имеет значение ИСТИНА, то функция НОРМРАСП возвращает интегральную функцию распределения; если аргумент имеет значение ЛОЖЬ, то возвращается функция плотности распределения.

Для вычисления выборочного значения статистики критерия χ^2 используется функция

ХИ2ОБР(уровень_значимости;степени_свободы)

категории *Статистические*.

Фрагмент рабочего листа в Excel.

