

Лабораторная работа 1.

Критерий согласия Пирсона

Теоретические сведения

Теория вероятностей и математическая статистика занимаются анализом закономерностей случайных массовых явлений. Предметом математической статистики является изучение случайных величин (или случайных событий, процессов) по результатам наблюдений.

Множество значений результатов наблюдений над одной и той же СВ ξ при одних и тех же условиях называется **выборкой**. Элементы выборки называются **выборочными значениями**. Количество проведенных наблюдений называется **объемом выборки**. Разность W между максимальным и минимальным элементами называется **размахом выборки**: $W = x_{\max} - x_{\min}$.

Пусть имеется выборка объема n : $x_1; x_2; \dots; x_n$. Если в выборке объема n элемент x_i встречается n_i раз, число n_i называется **частотой** выборочного значения x_i , а $\frac{n_i}{n}$ —

относительной частотой. Очевидно, что $\sum_{i=1}^k n_i = n$, где k — число различных элементов выборки.

Последовательность пар $(x_i^*; n_i)$, где $x_1^*, x_2^*, \dots, x_k^*$ — различные выборочные значения, а n_1, n_2, \dots, n_k — соответствующие им частоты, называется **статистическим рядом**. Обычно статистический ряд записывают в виде таблицы, первая строка которой содержит различные выборочные значения x_i^* , а вторая — их частоты n_i (или относительные частоты $\frac{n_i}{n}$, иногда и те, и другие).

В случае, когда число значений признака (СВ ξ) велико или признак является непрерывным (т. е. когда СВ ξ может принимать любое значение в некотором интервале), составляют **интервальный статистический ряд**. Для этого весь диапазон выборочных значений от x_{\min} до x_{\max} разбивают на k интервалов (обычно от 5 до 20; для определения количества интервалов можно использовать полуэмпирическую формулу Стерджесса $k \approx 1 + \log_2 n$) одинаковой длины $h = \frac{W}{k}$ и определяют частоты n_i — количество элементов выборки, попавших в i -й интервал (элемент, совпадающий с верхней границей интервала, относится к последующему интервалу). Полученные данные сводят в таблицу:

$[x_i; x_{i-1})$	$[x_0; x_1)$	$[x_1; x_2)$...	$[x_{k-1}; x_k]$
$x_i^* = \frac{x_{i-1} + x_i}{2}$	x_1^*	x_2^*	...	x_k^*
n_i	n_1	n_2	...	n_k
$\frac{n_i}{n}$	$\frac{n_1}{n}$	$\frac{n_2}{n}$...	$\frac{n_k}{n}$

Пусть выборка объема n представлена в виде интервального статистического ряда. Оценками для математического ожидания и дисперсии наблюдаемой случайной величины являются **выборочное среднее**

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

и **выборочная дисперсия**

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

При этом выборочная дисперсия дает всегда немного заниженную оценку для дисперсии, поэтому вместо нее используют несмещенную оценку дисперсии

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

Эмпирической функцией распределения называется функция $F_n^*(x)$, определяющая для каждого значения x относительную частоту наблюдения значений, меньших x :

$$F_n^*(x) = \sum_{x_i^* < x} \frac{n_i}{n}.$$

Гистограммой относительных частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длины h , а высоты равны $\frac{n_i}{nh}$. Площадь гистограммы относительных частот равна 1.

При достаточно большом объеме выборки n и достаточно малых интервалах группировки h гистограмма относительных частот является хорошим приближением графика плотности распределения наблюдаемой случайной величины. Поэтому по виду гистограммы можно выдвинуть предположение (гипотезу) о распределении изучаемой случайной величины.

Процедура сопоставления высказанного предположения (гипотезы) с выборочными данными называется **проверкой гипотез**.

Под **статистической гипотезой** понимают всякое высказывание (предположение) о виде (**непараметрическая гипотеза**) или параметрах (**параметрическая гипотеза**) неизвестного распределения. Статистическая гипотеза называется **простой**, если она полностью определяет функцию распределения. В противном случае гипотеза называется **сложной**.

Одну из гипотез выделяют в качестве **основной** (или **нулевой**) H_0 , а другую, являющуюся логическим отрицанием H_0 , – в качестве **конкурирующей** (или **альтернативной**) гипотезы \bar{H} .

Правило, по которому принимается решение принять или отклонить проверяемую гипотезу, называется **критерием проверки статистической гипотезы** (**статистическим критерием**). При этом заранее выбирают допустимое значение ошибки вывода, которое называется **уровнем значимости** статистического критерия и обозначается α (это вероятность отвергнуть нулевую гипотезу, когда она верна).

Статистические критерии, с помощью которых проверяются гипотезы о виде распределения, называются **критериями согласия** или **непараметрическими критериями**.

Критерий согласия χ^2 Пирсона. Пусть имеется выборка объема n и сгруппированный статистический ряд, в котором k групп. Например, в случае непрерывной СВ это будут k интервалов $[x_{i-1}; x_i]$.

Группы должны выбираться так, чтобы охватывать весь диапазон значений предполагаемой СВ. Если диапазон значений СВ не ограничен (к примеру, нормальная СВ принимает любые значения из $(-\infty; +\infty)$), то крайние интервалы должны быть расширены до $-\infty$ и $+\infty$ соответственно.

Кроме того, интервалы (группы) должны быть не очень маленькими, чтобы в каждый из них входило не менее 5 наблюдений. Группы с малым количеством наблюдений объединяют с соседними.

Проверяемая гипотеза представляет собой предположение о распределении наблюдаемой СВ и является простой (конкретно указывает предполагаемое распределение):

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.

Рассмотрим статистику

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

Для вычисления статистики $\chi_{\text{расч}}^2$ нужно знать сгруппированный статистический ряд и теоретическую функцию распределения $F(x)$ для расчета вероятностей p_i . При этом теоретическое распределение $F(x)$ может зависеть от одного или нескольких параметров. В этом случае вместо значений параметров используются их оценки, рассчитанные по сгруппированному статистическому ряду до объединения групп.

Замечание. Контроль вычислений можно осуществить по формуле

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

Пусть r – число неизвестных параметров теоретического распределения, оцененных по выборке. **Критерий согласия χ^2 Пирсона** заключается в следующем: если $\chi_{\text{расч}}^2 < \chi_{\alpha; k-r-1}^2$, где $\chi_{\alpha; k-r-1}^2$ 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 $\chi_{\text{расч}}^2 \geq \chi_{\alpha; k-r-1}^2$, то гипотеза H_0 отвергается (не согласуется с данными эксперимента).

Основное достоинство критерия согласия χ^2 Пирсона – его универсальность, т. е. применимость для любого закона распределения, в том числе с неизвестными параметрами. Основной недостаток – необходимость большого объема выборки (не менее 60–100 наблюдений) и произвольность группировки, влияющая на величину $\chi_{\text{расч}}^2$.

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 распределения 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?

Пример и методические указания по выполнению лабораторной работы в Excel

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 $\alpha = 0,05$.

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. 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 будет $h \approx \frac{W}{k} = \frac{27}{8} = 3,375$. Округлив с точностью до 0,1 в большую сторону, принимаем $h = 3,4$. Находим количество элементов выборки в каждом интервале.

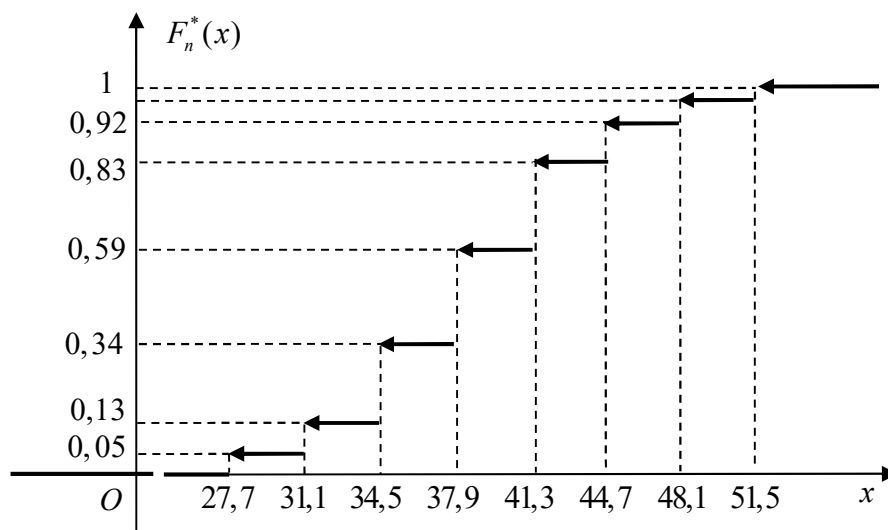
$[x_i; x_{i-1})$	x_i^*	n_i	$\frac{n_i}{n}$	$\frac{n_i}{nh}$
[26; 29,4)	27,7	5	0,05	0,015
[29,4; 32,8)	31,1	8	0,08	0,024
[32,8; 36,2)	34,5	21	0,21	0,062
[36,2; 39,6)	37,9	25	0,25	0,074
[39,6; 43)	41,3	24	0,24	0,071
[43; 46,4)	44,7	9	0,09	0,026
[46,4; 49,8)	48,1	6	0,06	0,018
[49,8; 53,2]	51,5	2	0,02	0,006

2. Для построения эмпирической функции распределения и гистограммы относительных частот дополним интервальный статистический ряд столбцами $\frac{n_i}{n}$ (относительные частоты нужны для построения эмпирической функции распределения) и $\frac{n_i}{nh}$ (высоты прямоугольников гистограммы).

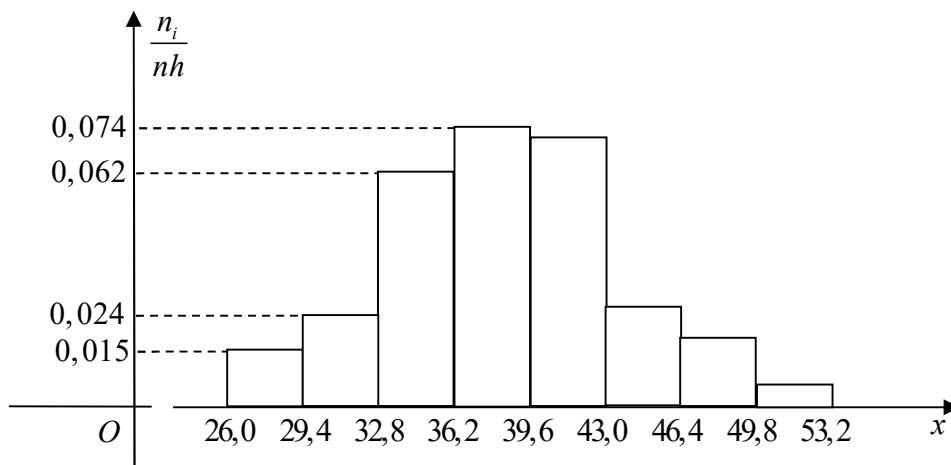
Запишем эмпирическую функцию распределения, накапливая относительные частоты $\frac{n_i}{n}$ (отметим, что при построении эмпирической функции распределения по интервальному статистическому ряду изменения ее значений (скачки) происходят в точках, соответствующих серединам интервалов группировки):

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

Построим график $F_n^*(x)$.

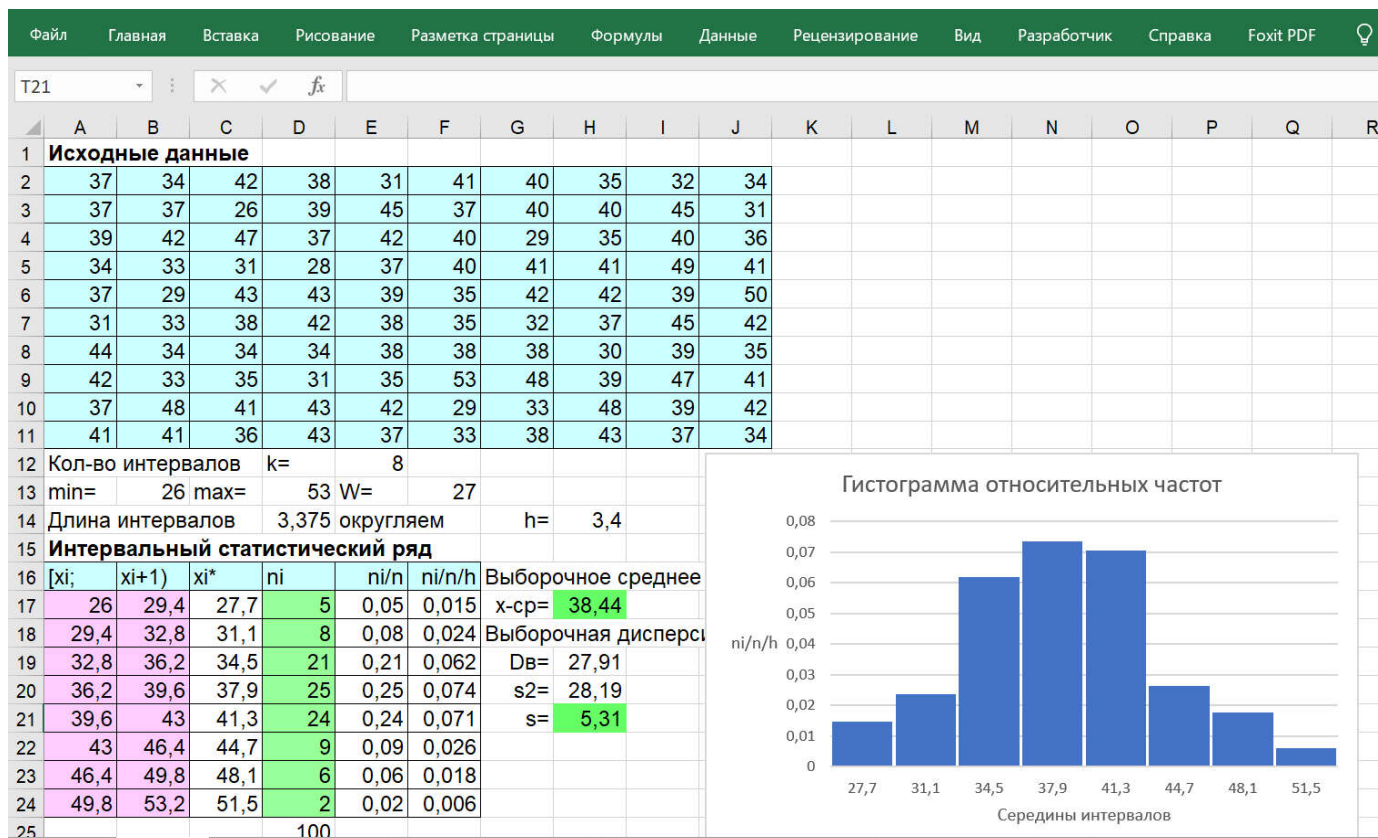


3. Построим гистограмму относительных частот, состоящую из прямоугольников шириной $h = 3,4$ и высотой $\frac{n_i}{nh}$,



По виду гистограммы можно выдвинуть гипотезу о том, что выборка взята из нормального распределения. Для проверки этой гипотезы по критерию согласия χ^2 Пирсона нужно рассчитать оценки параметров распределения по сгруппированному статистическому ряду.

Ниже приведен фрагмент рабочего листа и даны рекомендации по выполнению пунктов 1, 3 в Excel.



1. 1) В ячейке A1: "Исходные данные", в ячейки A2:J11 введите или скопируйте выборку.

2) В ячейке A12: "Кол-во интервалов"; в ячейке D12: "k="; в ячейке E12 вычислите количество интервалов по формуле Стерджесса и округлите до целого: =ОКРУГЛ(1+LOG(100;2);0) (в Excel каждая формула начинается со знака =; в формулах недопустимы пробелы). Получается $k = 8$.

В ячейке A13: "min="; в ячейке B13 =МИН(A2:J11); в ячейке C13: "max="; в ячейке D13 =МАКС(A2:J11); в ячейке E13: "W="; в ячейке F13 =D13-B13. В ячейке A14: "длина интервалов"; в ячейке D14 =F13/E12; в ячейке E14: "округляем"; в ячейке G14: "h="; в ячейке H14 =ОКРВВЕРХ(D14;0,1).

3) В ячейке A15: "Интервальный статистический ряд". В ячейке A16: "[xi;"; в ячейке B16 "xi+1)"; в ячейке C16: "xi*"; в ячейке D16: "ni"; в ячейке E16: "ni/n"; в ячейке F16: "ni/n/h".

4) Для нахождения концов интервалов в ячейку A17 запишите минимальное выборочное значение (используйте формулу или ссылку на вычисленное значение), а в ячейки A18 и B17 — формулу =A17+\$H\$14, затем скопируйте (протяните) эту формулу в ячейки соответствующих столбцов до строки 24 включительно (т. к. должно быть 8 интервалов).

При копировании формула в Excel перенастраивается на новые адреса (относительные ссылки). Чтобы адрес какой-либо ячейки был абсолютным (т. е. не перенастраивался), нужно после его указания в процессе формирования формулы нажать клавишу F4 или записать адрес в виде \$A\$1. Клавиша F4 действует в этом случае как переключатель, преобразуя адрес последовательно в \$A\$1, A\$1, \$A1, A1. Знаком \$ обозначается та часть адреса, которая должна оставаться абсолютной. При перемещении формулы в новое место таблицы ссылки в формуле не изменяются.

5) В столбце С вычислите середины интервалов.

6) Для вычисления частот в ячейке D18 запишите формулу $=СЧЁТЕСЛИМН(\$A\$2:\$J\$11;">="&A17;\$A\$2:\$J\$11;"<"&B17)$ и протяните ее на ячейки E19:E24. В последнем интервале нужно включить правый конец, поэтому в ячейке D24 формулу нужно записать в виде:

$=СЧЁТЕСЛИМН(\$A\$2:\$J\$11;">="&A17;\$A\$2:\$J\$11;"<"&B17)$.

Для контроля вычислений рассчитайте в ячейке D25 сумму частот.

7) Вычислите относительные частоты и высоты прямоугольников гистограммы в столбцах Е и F.

3. Построим гистограмму относительных частот, используя вкладку Вставка → Диаграмма.

1) Выделите данные, которые будут отображаться по оси Ox , т. е. ячейки C17:C24 (обозначим интервалы их серединами) и, нажав клавишу **Ctrl**, выберите данные, которые будут включены в гистограмму по оси Oy , т. е. ячейки F17:F24.

2) Выберите вкладку Вставка → Диаграмма, выберите вариант Гистограмма.

3) Щелкнув по прямоугольникам гистограммы правой кнопкой мыши, выбираем Формат ряда данных и во вкладке Параметры устанавливаем Боковой зазор 5%.

4) Далее можно изменить название гистограммы и подписи осей.

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

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

Е 4. 1) В ячейке G16: "Выборочное среднее"; в ячейке G17: "х-ср="; в ячейке
Х N17 вычислите выборочное среднее по формуле
С =СУММПРОИЗВ(C17:C24;D17:D24)/100.

2) В ячейке N18: "Выборочная дисперсия"; в ячейке G19: "Дв="; в ячейке
Е N19 вычислите выборочную дисперсию по формуле
Л =СУММПРОИЗВ(C17:C24;C17:C24;D17:D24)/100-N17*N17.

3) В ячейке G20: "s2="; в ячейке N20 вычислите несмещенную оценку
дисперсии: =N19*100/99.

4) В ячейке G21: "s="; в ячейке N21 вычислите оценку среднего
квадратического отклонения: =КОРЕНЬ(N20).

6. 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 $[x_{i-1}; x_i)$;

$p_i = P(\xi \in [x_{i-1}; x_i))$ – теоретическая вероятность попадания СВ в интервал $[x_{i-1}; x_i)$, в случае нормального распределения с параметрами $a = 38,44, \sigma = 5,31$ эта вероятность рассчитывается как разность значений функции Лапласа:

$$p_i = \Phi\left(\frac{x_i - 38,44}{5,31}\right) - \Phi\left(\frac{x_{i-1} - 38,44}{5,31}\right);$$

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

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

Все вычисления заносим в таблицу.

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

Суммирования значения в предпоследнем столбце, вычисляем выборочное значение статистики критерия χ^2 Пирсона: $\chi^2_{\text{расч}} = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} \approx 2,24$. Сумма

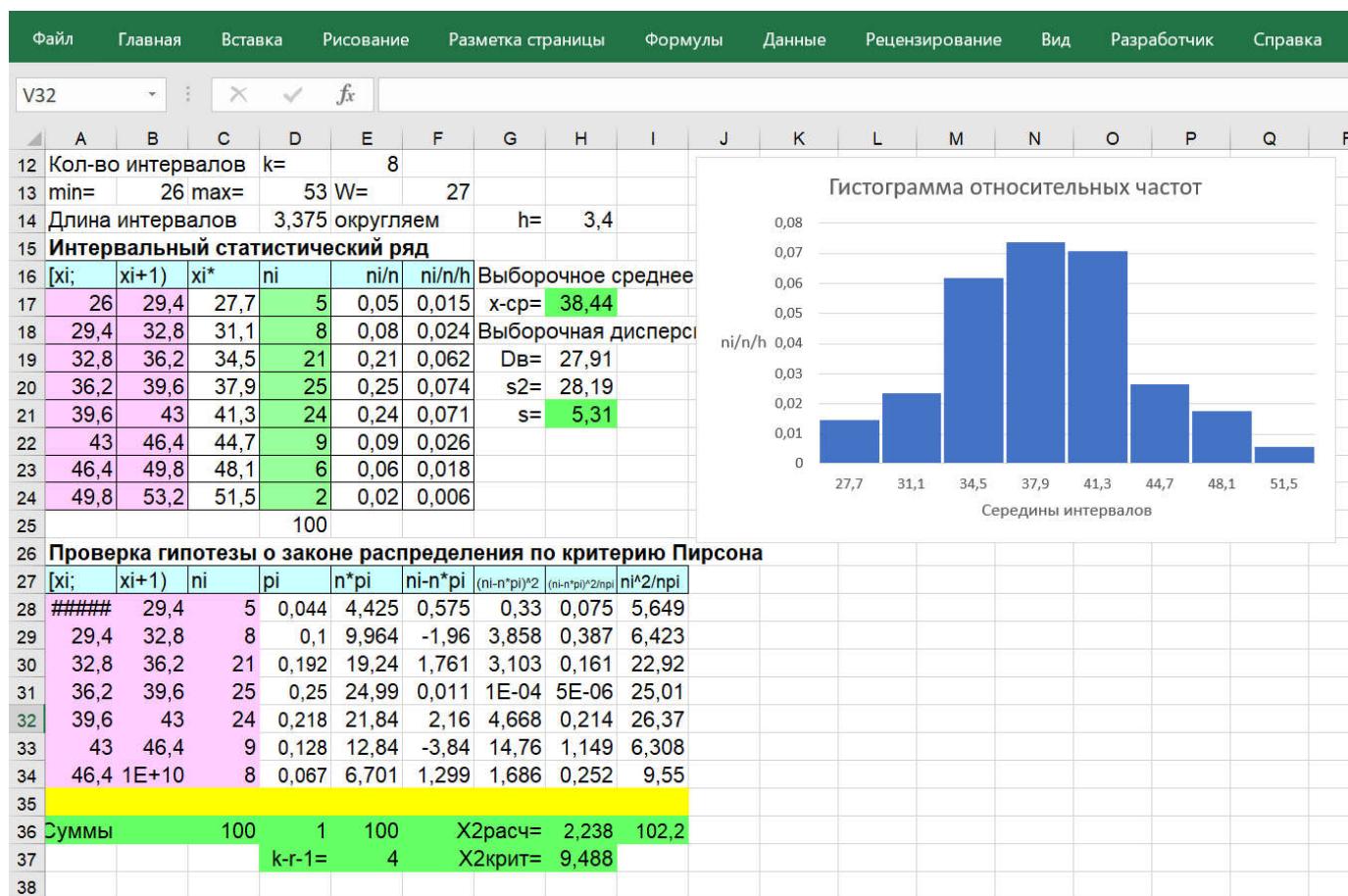
элементов последнего столбца равна $\sum_{i=1}^k \frac{n_i^2}{np_i} \approx 102,24$. Это позволяет провести

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 $a = 38,44$, $\sigma = 5,31$.

Ниже приведен фрагмент рабочего листа и даны рекомендации по выполнению пункта 6 в Excel.



6. 1) В ячейке A26: "Проверка гипотезы о законе распределения по критерию Пирсона". В ячейках A27:G36 будет расположена таблица,

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

2) Подпишите столбцы таблицы в строке 27: в ячейке A27: "[xi;"; в ячейке B27 "xi+1)"; в ячейке C27: "ni"; в ячейке D27: "pi"; в ячейке E27: "n*pi"; в ячейке F27: "ni-npi "; в ячейке G27: "(ni-npi)^2"; в ячейке H27: "(ni-npi)^2/np_i"; в ячейке I27: "ni^2/np_i".

3) Запишите в ячейки A28:B35 интервалы группировки, задав в ячейке A28 формулу =A17 и протянув ее на остальные ячейки. Аналогично скопируйте в массив C28:C35 частоты из массива D17:D24.

4) **Помните**, что при использовании критерия χ^2 Пирсона интервалы числом наблюдений $n_i < 5$ объединяют с соседними. Исправьте интервальный статистический ряд в соответствии с этим замечанием. Учтите, что первый интервал нужно продлить до $-\infty$, а последний — до $+\infty$. В ячейке C36 введите формулу =СУММ(C28:C35) и проконтролируйте: сумма частот должна быть равна объему выборки. Скопируйте эту формулу на ячейки D36, E36, H36, I36.

5) В столбце D вычислите вероятности попадания нормальной случайной величины в соответствующие интервалы. Для этого в ячейке D29 задайте формулу
=НОРМ.РАСП(B29;\$H\$17;\$H\$21;ИСТИНА)-
НОРМ.РАСП(A29;\$H\$17;\$H\$21;ИСТИНА) и протяните ее на ячейки D30:D34.
Функция

Е НОРМ.РАСП(х;среднее;стандартное откл;интегральная)
Л вычисляет значение функции нормального распределения

$F(x) = 0,5 + \Phi\left(\frac{x-a}{\sigma}\right)$. Здесь x — значение, для которого вычисляется

Е значение функции, среднее — математическое ожидание a (задаем
Х выборочное среднее \bar{x}), стандартное откл — среднеквадратическое
С отклонение σ (задаем s), интегральная — логическое значение,
определяющее форму функции. Для того чтобы получить нужное значение
интегральной функции распределения, задаем Интегральная: ИСТИНА.
Если задать значение ЛОЖЬ, то получится значение плотности
нормального распределения.

Для первого интервала вероятность вычисляется в ячейке D28 как
=НОРМ.РАСП(B28;\$H\$17;\$H\$21;ИСТИНА), для последнего — в ячейке D34
как =1-НОРМ.РАСП(A34;\$H\$17;\$H\$21;ИСТИНА). В ячейке D36
проконтролируйте, что сумма вероятностей равна 1.

Е 6) В столбцах E, F, G, H, I вычислите для каждого интервала значения np_i ,

Л $n_i - np_i, (n_i - np_i)^2, \frac{(n_i - np_i)^2}{np_i}, \frac{n_i^2}{np_i}$ и посчитайте соответствующие суммы:

$$\sum_{i=1}^k np_i = n; \quad \chi_{\text{расч}}^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}; \quad \sum_{i=1}^k \frac{n_i^2}{np_i} = \chi_{\text{расч}}^2 + n.$$

Е 7) Вычислите критическое значение $\chi^2_{\text{крит}} = \chi^2_{\alpha; k-r-1}$, где k — число
Х интервалов группировки после объединения малочисленных с соседними,
С $r = 2$, поскольку при расчете теоретических вероятностей использовались
две полученные по выборке оценки \bar{x} и s параметров нормального
распределения. Для этого в ячейке E37 запишите значение $k - r - 1$ с
учетом объединения интервалов, а в ячейке H37 найдите $\chi^2_{\text{крит}}$ по формуле
=ХИ2.ОБР.ПХ(0,05;E37).

С 8) Сделайте вывод о соответствии или несоответствии проверяемой
гипотезы экспериментальным данным при заданном уровне значимости.