Лабораторная работа.
Критерий Пирсона

Цель работы – научить применять критерий Пирсона для решения статистических задач посредством MS Excel.

Рассмотрим, как критерий Пирсона (Пирсона χ^2) применяется для сравнения экспериментального распределения с теоретическим, например, для проверки гипотезы о нормальном распределении генеральной совокупности. Здесь можем выделить два случая.

1. Эмпирическое распределение задано в виде последовательности равноотстоящих вариант и соответствующих им частот.

Пусть по выборочной совокупности объема n получено эмпирическое распределение:

Варианты x_i	x_1	x_2	…	x_i	…	x_k
Эмпирические частоты m_i	m_1	m_2	…	m_i	…	m_k

Алгоритм

1. Сформулировать гипотезы:

H_0: гипотеза о нормальном распределении генеральной совокупности.

H_1: генеральная совокупность не распределена по нормальному закону.

Выбрать уровень значимости α.

2. Вычислить теоретические частоты по формуле:

$$m_i^* = \frac{n \cdot h}{\sigma_в} \cdot \varphi(u_i),$$

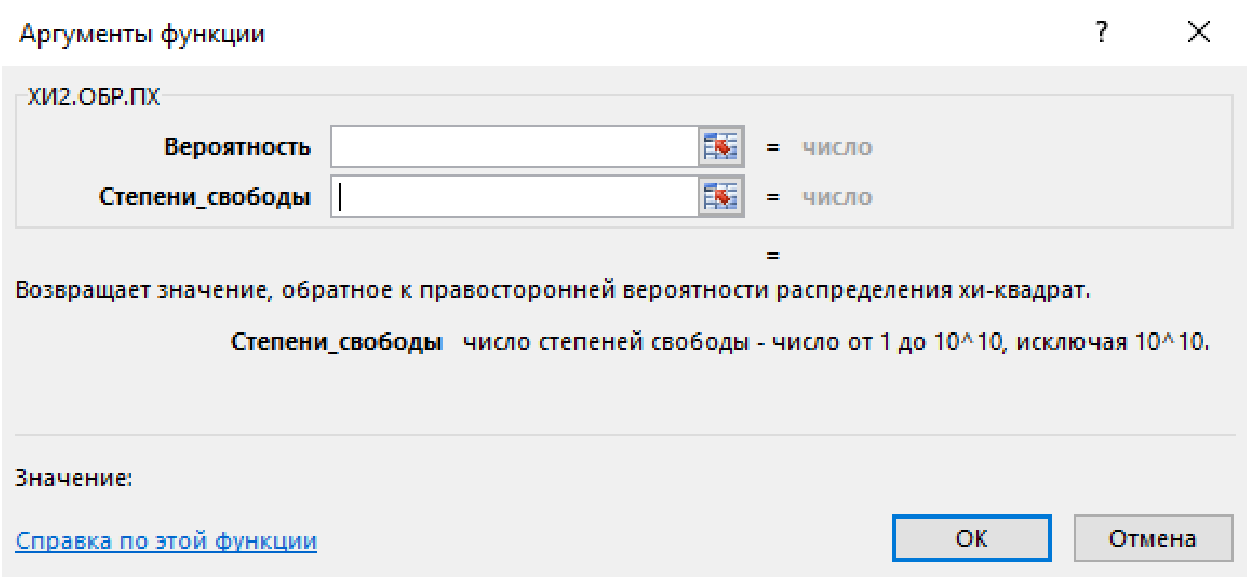
где n – объем выборочной совокупности, h – шаг (разность между двумя соседними вариантами),

$$u_i = \frac{x_i - \overline{x}}{\sigma_в}, \ \ \varphi(u_i) = \frac{1}{\sqrt{2\pi}} \cdot e^{-u_i^2/2} \quad .$$

3. Найти эмпирическое значение критерия по формуле:

$$\chi^2_{эмп} = \sum_{i=1}^{n} \frac{\left(m_i - m_i^*\right)^2}{m_i^*}.$$

4. Определить критические значения критерия «хи-квадрат» с помощью статистической функции **ХИ2.ОБР.ПХ**, для уровня значимости α и степени свободы $k = s - r - 1$, где s – количество групп выборочной совокупности, r –количество параметров в распределении (например, в нормальном распределении $r = 2$).



5. Если $\chi^2_{эмп} < \chi^2_{крит}$, то на уровне значимости α принимается нулевая гипотеза, в ином случае – альтернативная.

Пример. Установить при уровне значимости 0,05, согласуется ли гипотеза о нормальном распределении генеральной совокупности с данными выборки объёмом n = 200, если известны значения вариант и их эмпирические частоты:

x_i	0,3	0,5	0,7	0,9	1,1	1,3	1,5	1,7	1,9	2,1	2,3
m_i	6	9	26	25	30	26	21	24	20	8	5

Решение. На основании данных задачи составляем таблицу для вычисления теоретических частот:

1	2	3	4	5	6	7	8
i	x_i	m_i	$\lvert\Delta x_i\rvert$	$\lvert\Delta x_i\rvert^2$	u_i	$\varphi(u_i)$	m_i^*
1	0,3	6	0,962	0,9254	1,963	0,0581	4,7
2	0,5	9	0,762	0,5806	1,555	0,1191	9,7
3	0,7	26	0,562	0,3158	1,147	0,2067	16,9
4	0,9	25	0,362	0,1310	0,739	0,3037	24,8
5	1,1	30	0,162	0,0262	0,331	0,3777	30,8
6	1,3	26	0,038	0,0014	0,078	0,3977	32,5
7	1,5	21	0,238	0,0566	0,486	0,3546	28,9
8	1,7	24	0,438	0,1918	0,894	0,2676	21,8
9	1,9	20	0,638	0,4070	1,302	0,1709	14,0
10	2,1	8	0,838	0,7022	1,710	0,0925	7,5
11	2,3	5	1,038	1,0774	2,118	0,0423	3,5

По 1-му столбцу таблицы определяем число различных вариант s = 11 и $k=s-3=11-3=8$.

По 3-му столбцу:

$$n = \sum_{i=1}^{k} m_i = 6 + 9 + 26 + 25 + 30 + 26 + 21 + 24 + 20 + 8 + 5 = 200.$$

По 2-му и 3-му столбцам: $|\Delta x_i| = |x_i - \bar{x}|$

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{k} x_i \cdot m_i = \frac{1}{200} \cdot (0{,}3 \cdot 6 + 0{,}5 \cdot 9 + 0{,}7 \cdot 26 + \ldots + 2{,}3 \cdot 5) = \frac{1}{200} \cdot 252{,}4 = 1{,}262.$$

По 3-му и 5-му столбцам: $\sigma^2 = \frac{1}{n} \cdot \sum_{i=1}^{k} |\Delta x_i|^2 \cdot m_i = \frac{1}{200} \cdot (0{,}9254 \cdot 6 + \ldots + 1{,}0774 \cdot 5) =$

$= \frac{1}{200} \cdot 48{,}0312 = 0{,}24.$ $\sigma = \sqrt{0{,}24} = 0{,}49$.

По 4-му: $u_i = \frac{|\Delta x_i|}{\sigma}$.

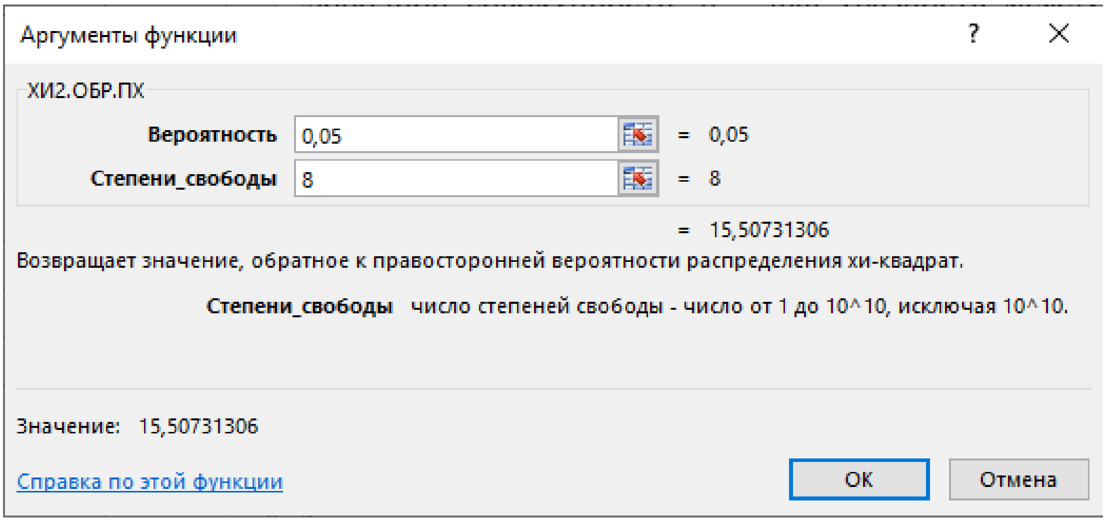
Затем рассчитываем значения функции $\varphi(u_i) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{u_i^2}{2}}$. Шаг вариант определяем по 2-му столбцу: $h = x_{i+1} - x_i = 0{,}2$. Теоретические частоты определяем как $m_i^* = \frac{n \cdot h}{\sigma} \cdot \varphi(u_i)$.

Для вычисления критерия χ^2 строим следующую таблицу:

1	2	3	4	5	6
i	m_i	m_i^*	$m_i - m_i^*$	$(m_i - m_i^*)^2$	$\frac{(m_i - m_i^*)^2}{m_i^*}$
1	6	4,7	1,3	1,6	0,33
2	9	9,7	-0,7	0,5	0,05
3	26	16,9	9,1	83,3	4,94
4	25	24,8	0,2	0,0	0,00
5	30	30,8	-0,8	0,7	0,02
6	26	32,5	-6,5	41,8	1,29
7	21	28,9	-7,9	63,1	2,18
8	24	21,8	2,2	4,7	0,21
9	20	14,0	6,0	36,6	2,62
10	8	7,5	0,5	0,2	0,03
11	5	3,5	1,5	2,4	0,69

По 6-му столбцу, суммируя, находим значение $\chi^2_{эмп}$=12,37.

Найдем с помощью статистической функции **ХИ2.ОБР.ПХ** значение $\chi^2_{крит}$=15,51 для уровня значимости 0,05 и степени свободы 8.



Это значение больше, чем вычисленное по данным выборки, поэтому можно сделать вывод о принятии нулевой гипотезы, то есть, гипотеза о нормальном распределении генеральной совокупности согласуется с данными выборки.

2. Эмпирическое распределение задано в виде последовательности интервалов и соответствующих им частот.

Алгоритм

1. Сформулировать гипотезы:

H_0: гипотеза о нормальном распределении генеральной совокупности.

H_1: генеральная совокупность не распределена по нормальному закону.

Выбрать уровень значимости α.

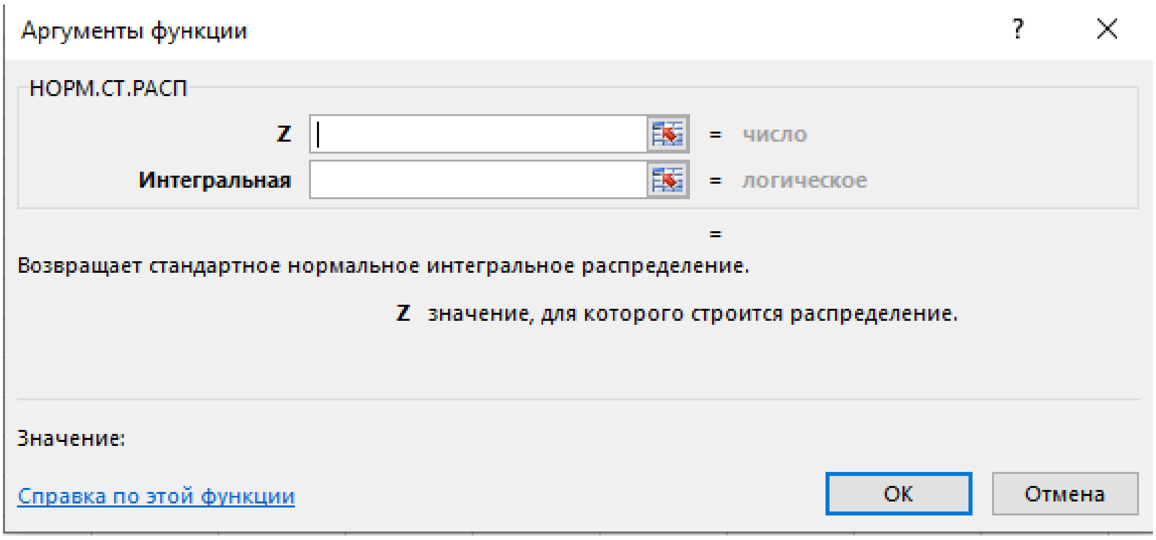
2. Пронормировать значения выборочной совокупности X, то есть перейти к случайной величине $Z = \dfrac{X - \overline{x^*}}{\sigma^*}$, где $\overline{x^*}$ – выборочные среднее значение и σ^* – среднее квадратическое отклонение, для нахождения которых необходимо в расчетных формулах взять за значения вариант середины интервалов.

3. Вычислить концы интервалов: $z_i = \dfrac{x_i - \overline{x^*}}{\sigma^*}$, $z_{i+1} = \dfrac{x_{i+1} - \overline{x^*}}{\sigma^*}$.

4. Вычислить теоретические частоты по формуле:

$$m_i^* = n \cdot p_i,$$

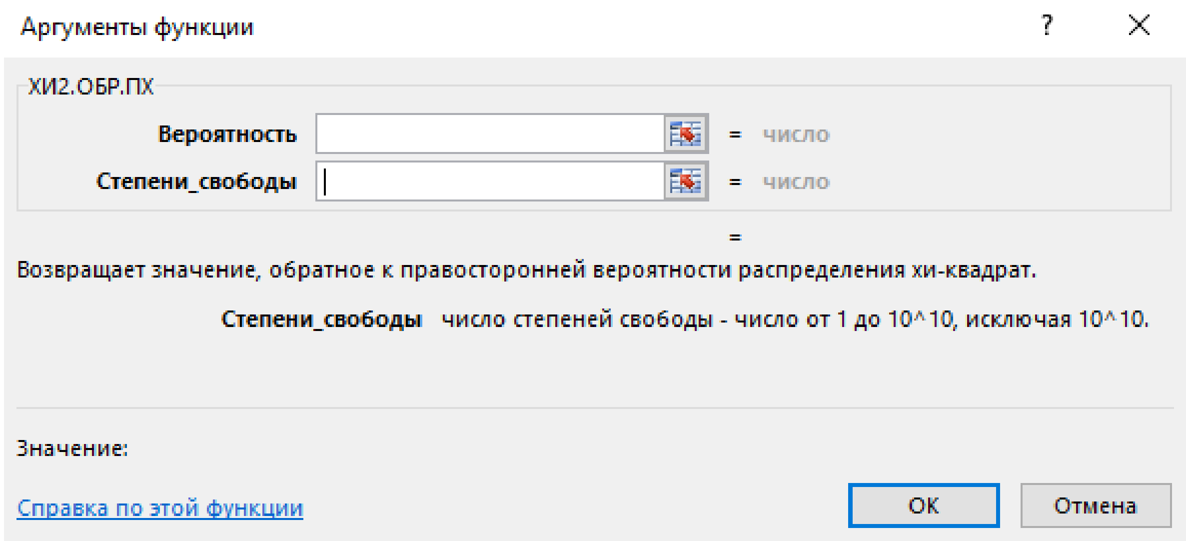
где n – объем выборочной совокупности, $p_i = \Phi(z_{i+1}) - \Phi(z_i)$ – вероятности попадания X в интервал $(x_i;\ x_{i+1})$, $\Phi(z)$ – функция Лапласа с помощью статистической функции **НОРМ.СТ.РАСП(x)-0,5.**



3. Найти эмпирическое значение критерия по формуле:

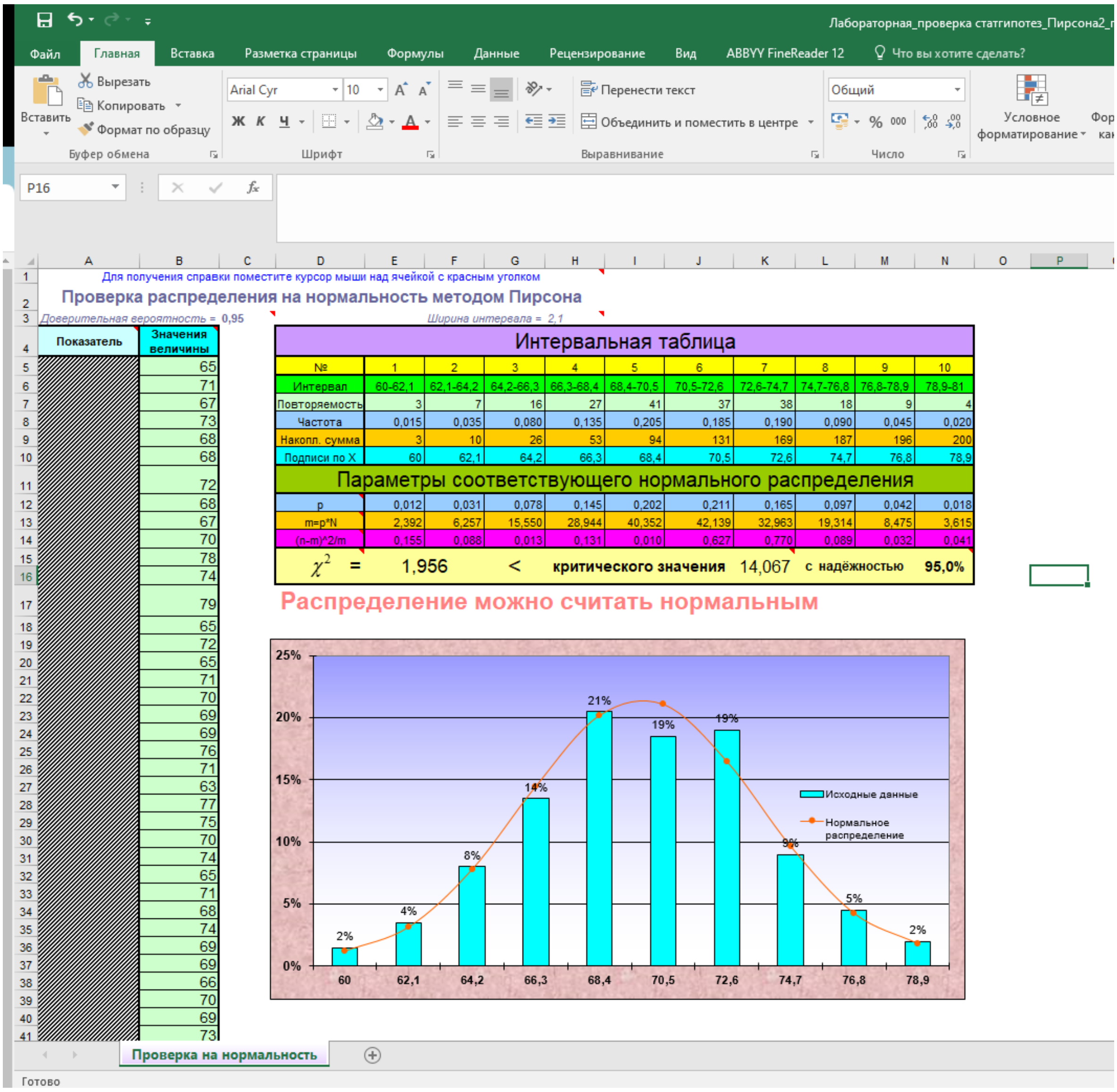
$$\chi^2_{эмп} = \sum_{i=1}^{n} \frac{\left(m_i - m_i^*\right)^2}{m_i^*}.$$

4. Определить критические значения критерия «хи-квадрат» с помощью статистической функции **ХИ2.ОБР.ПХ**, для уровня значимости α и степени свободы $k = s - 3$, где s – количество интервалов выборочной совокупности.



5. Если $\chi^2_{эмп} < \chi^2_{крит}$, то на уровне значимости α принимается нулевая гипотеза, в ином случае – альтернативная.

Замечание. В MS Excel авторами составлен шаблон, позволяющий выборочную совокупность сводить к последовательности интервалов и соответствующих им частот, проверять гипотезу о нормальном распределении генеральной совокупности и строить графическое изображение **(2 случай).**



Задачи для самостоятельного решения

Задача 1. Суд рассматривает жалобу посетителей казино. По их мнению, игральная кость, используемая в казино, фальшива и некоторые числа очков выпадают чаще, чем другие, этим пользуются крупье и обирают игроков.

Число очков	Количество выпадений $m_{факт}$
1	101
2	86
3	107
4	94

5	97
6	117

Задача 2. В некотором городе произошла вспышка инфекционного заболевания. Есть предположение, что источником заражения явилась питьевая вода. Проверить это предположение решили с помощью выборочного опроса городского населения, по которому необходимо установить, влияет ли количество выпиваемой воды на количество заболевших.

Количество выпиваемой в среднем за день воды	Число заболевших	Число не-заболевших
Менее одного стакана	140	220
От одного стакана до одного литра	360	350
Более одного литра	365	245