**Соленок Анастасия , ПИ - 10/2**

**Лабораторная работа №1**

**Критерий согласия Пирсона**

**Вариант 26**

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?

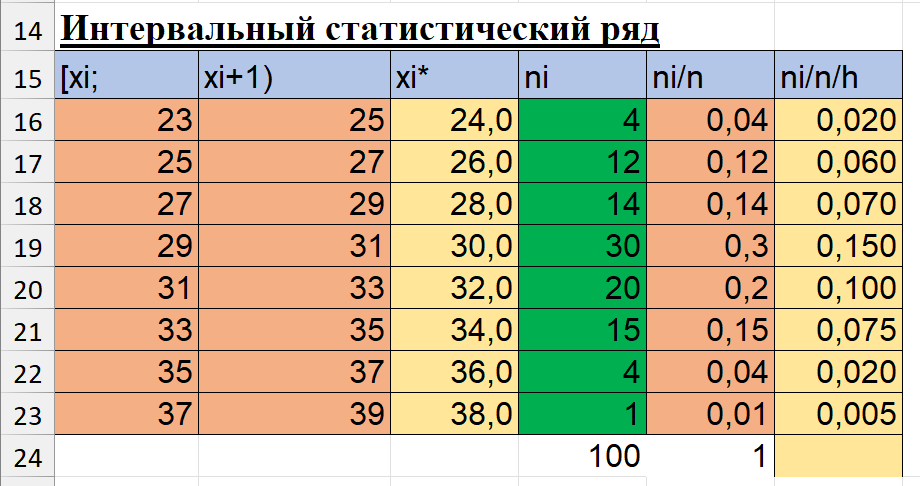
4. Определить выборочное среднее и несмещенную оценку дисперсии по сгруппированному статистическому ряду.

5. Записать предполагаемую плотность закона распределения.

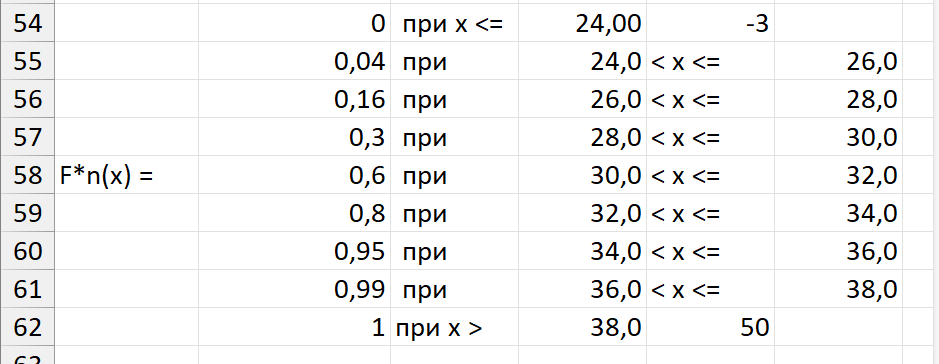
6. Проверить по критерию  Пирсона гипотезу о законе распределения. Уровень значимости принять равным α = 0,05.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Вариант 26  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | 59 | 44 | 52 | 56 | 33 | 29 | 28 | 36 | 39 | 30 | | 45 | 56 | 42 | 29 | 29 | 31 | 32 | 26 | 31 | 32 | | 29 | 29 | 31 | 31 | 33 | 33 | 33 | 31 | 31 | 32 | | 31 | 26 | 28 | 29 | 33 | 28 | 29 | 26 | 29 | 27 | | 28 | 25 | 30 | 24 | 30 | 36 | 26 | 33 | 29 | 33 | | 33 | 28 | 27 | 30 | 26 | 32 | 23 | 28 | 26 | 25 | | 32 | 24 | 30 | 32 | 32 | 28 | 35 | 28 | 34 | 24 | | 29 | 33 | 27 | 29 | 32 | 31 | 30 | 25 | 31 | 33 | | 30 | 34 | 34 | 33 | 32 | 27 | 28 | 26 | 36 | 30 | | 30 | 30 | 32 | 28 | 30 | 30 | 29 | 30 | 29 | 30 | |

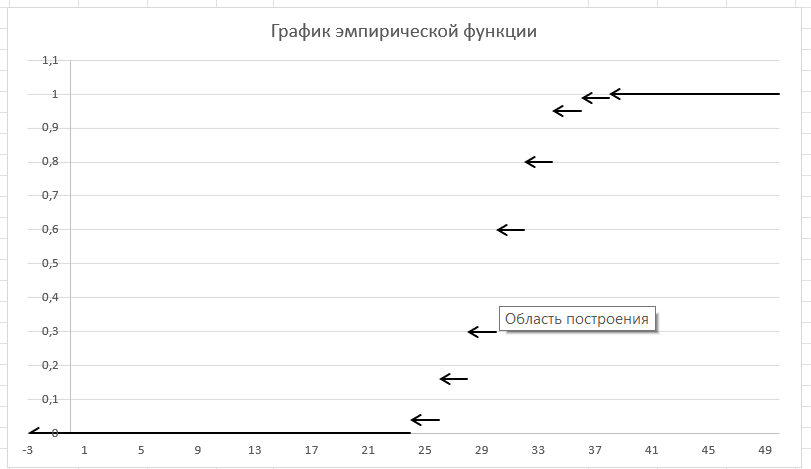
1. Объем выборки n = 100. Построим интервальный статистический ряд. Количество интервалов определим по формуле Стерджесса k ≈ 1 + log2(n) = 1+ log2(100) = 7,6439. Принимаем k = 8. Размах выборки W = xmax – xmin = 31 – 76 = 45. Длина каждого интервала будет h = W/k ≈ 16/8 = 2. Находим количество элементов выборки в каждом интервале.



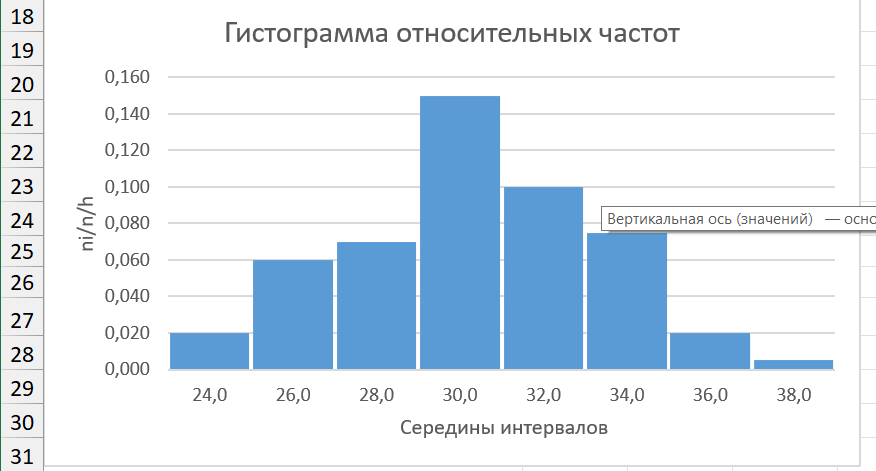
2. Для построения эмпирической функции распределения и гистограммы относительных частот дополним интервальный статистический ряд столбцами ni/n (относительные частоты нужны для построения эмпирической функции распределения) и ni/nh (высоты прямоугольников гистограммы). Запишем эмпирическую функцию распределения, накапливая относительные частоты ni/n (отметим, что при построении эмпирической функции распределения по интервальному статистическому ряду изменения ее значений (скачки) происходят в точках, соответствующих серединам интервалов группировки):



Построим график Fn\* (x) .

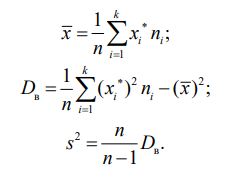


3. Построим гистограмму относительных частот, состоящую из прямоугольников шириной h = 6,5 и высотой ni/nh , По виду гистограммы можно выдвинуть гипотезу о том, что выборка взята из нормального распределения. Для проверки этой гипотезы по критерию согласия X2 Пирсона нужно рассчитать оценки параметров распределения по сгруппированному статистическому ряду.



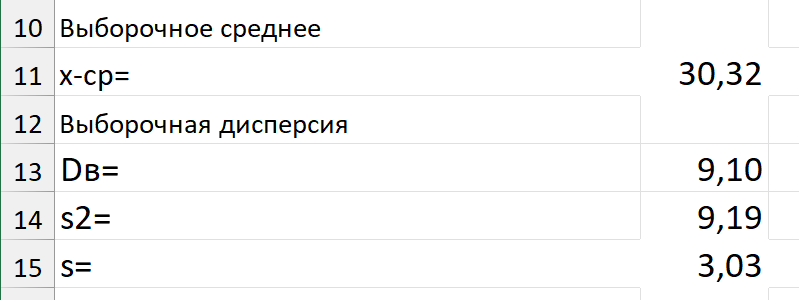
4. Рассчитаем оценки параметров предполагаемого нормального закона распределения по сгруппированному статистическому ряду. Данный закон содержит два параметра a и σ, которые имеют смысл математического ожидания и среднего квадратического отклонения СВ ξ: Mξ= a, D ξ = σ2.

В качестве оценок для математического ожидания a и дисперсии σ2 наблюдаемой случайной величины рассчитаем соответственно выборочное среднее и несмещенную оценку дисперсии s2, для вычисления s2 предварительно найдем выборочную дисперсию в Dв:

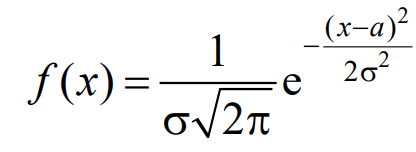


Используя интервальный статистический ряд, получим:

Тогда оценкой для среднего квадратического отклонения σ будет



5. Функция плотности нормального закона распределения имеет вид



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

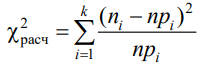
6. Проверим с помощью критерия согласия χ2 Пирсона гипотезу

H0: наблюдаемая СВ имеет нормальное распределение с параметрами

a =, σ = при альтернативе

: *наблюдаемая СВ имеет другое распределение.*

Для расчета статистики критерия Пирсона

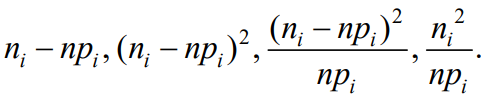


составим новую таблицу, содержащую следующие столбцы:

интервалы [xi-1; xi) − (при этом крайние интервалы должны быть расширены до −∞ и +∞ соответственно; а интервалы с количеством наблюдений меньше 5 объединяются с соседними);

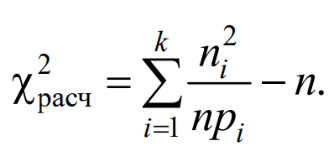
ni – эмпирическая частота наблюдения значений из интервала [xi-1; xi);

pi = P(ξ∈[xi-1; xi)) – теоретическая вероятность попадания СВ в интервал [xi-1; xi), в случае нормального распределения с параметрами a =, σ = эта вероятность рассчитывается как разность значений функции Лапласа:

npi – теоретическое значение соответствующей частоты,

а также столбцы со значениями

Последний столбец используется для контроля вычислений по формуле

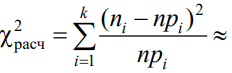


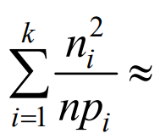
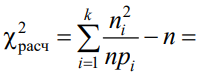
Все вычисления заносим в таблицу.



Суммирования значения в предпоследнем столбце, вычисляем

выборочное значение статистики критерия χ2 Пирсона:

3.326065. Сумма элементов последнего столбца равна

 103.3261. Это позволяет провести контроль вычислений: x`

102.618 – 100 = 2.618.

Определим критическое значение χ2крит = χ2α; k – r –1, где α = 0,05 – заданный уровень значимости; k = 8 – число интервалов после объединения малочисленных групп с соседними; r = 2, поскольку при расчете теоретических вероятностей pi использовались две полученные по выборке оценки x и s параметров нормального распределения. По таблице квантилей распределения χ2 получаем χ2крит = χ20.05; 2 = 9.487729.

Таким образом, χ2расч = 3.326065 < χ2крит = 9.487729, поэтому на уровне значимости α = 0,05 нет оснований отвергнуть гипотезу H0, согласно которой выборка взята из нормального распределения с параметрами a =, σ = .

Ответы на вопросы

1. Что называется выборкой? Что называется объемом выборки?

Выборка — это подмножество данных, полученное из генеральной совокупности, которое используется для анализа и получения статистических выводов.

Объем выборки — это количество элементов в выборке (обозначается ( n )).

2. Что называется частотой выборочного значения? Что называется относительной частотой?

Частота выборочного значения — это количество раз, которое определенное значение встречается в выборке.

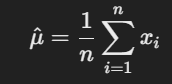
Относительная частота — это доля данного значения в выборке, рассчитывается как отношение частоты к объему выборки:

f = Частота значения/n

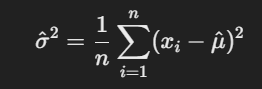
3. Как оценить по выборке математическое ожидание и дисперсию наблюдаемой СВ?

Для выборки ( x1, x2, …, xn ):

- Математическое ожидание :

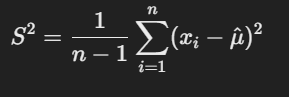


- Дисперсия (смещенная оценка):



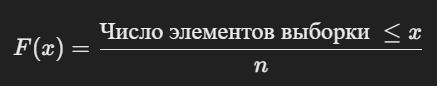
4. Как рассчитать несмещенную оценку дисперсии?

Несмещенная оценка дисперсии (( S^2 )) рассчитывается как:

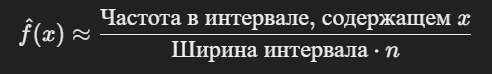


Мю-выборочное среднее

5. Как оценить по выборке функцию и плотность распределения наблюдаемой СВ?

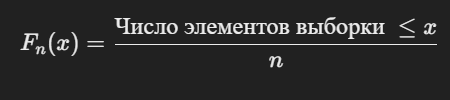


Плотность распределения может быть оценена с использованием гистограммы или ядерного сглаживания:



6. Что называется эмпирической функцией распределения?

Эмпирическая функция распределения (( F\_n(x))) — это функция, которая показывает долю выборочных значений, не превышающих ( x ). Она вычисляется как:



7. Что называется гистограммой относительных частот?

Гистограмма относительных частот — это графическое представление относительных частот значений выборки. Каждая колонка гистограммы имеет высоту, пропорциональную относительной частоте значений, попадающих в соответствующий интервал.

8. Чему равна площадь гистограммы относительных частот?

Площадь гистограммы относительных частот равна 1, так как сумма всех относительных частот равна 1.

9. Что называется статистической гипотезой?

Статистическая гипотеза — это предположение о свойствах распределения генеральной совокупности (например, математическом ожидании, дисперсии или форме распределения).

10. В каком случае статистическая гипотеза называется простой? сложной?

- Простая гипотеза полностью определяет параметры распределения.

- Сложная гипотеза оставляет неопределенными некоторые параметры.

11. В чем разница между нулевой и альтернативной гипотезами?

- Нулевая гипотеза H\_0 — это предположение, которое проверяется. Обычно утверждает, что «изменений нет» или «данные соответствуют ожиданиям».

- Альтернативная гипотеза H\_1 — это предположение, противоположное нулевой гипотезе, утверждает, что «изменения есть».

12. Что называется уровнем значимости статистического критерия?

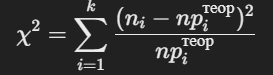
Уровень значимости alpha — это вероятность ошибочного отклонения нулевой гипотезы H\_0, если она верна. Обычно выбирается значение alpha = 0.05 или alpha = 0.01.

13. Что называется критерием согласия?

Критерий согласия — это метод проверки гипотезы о соответствии выборочного распределения теоретическому распределению (например, нормальному).

14. В чем суть критерия согласия Пирсона?

Критерий проверяет гипотезу о соответствии наблюдаемого распределения теоретическому. Расчетная статистика:



15. Какие достоинства и недостатки имеет критерий согласия Пирсона?

Достоинства:

1. Простота вычислений.

2. Подходит для любых типов распределений.

3. Основан на стандартных статистических подходах.

Недостатки:

1. Требует большого объема выборки n > 50.

2. Чувствителен к выбору числа интервалов.

3. Грубо аппроксимирует непрерывное распределение через дискретные интервалы.