|  |  |  |
| --- | --- | --- |
|  | | |
| Федеральное государственное бюджетное  образовательное учреждение высшего образования «Новосибирский государственный технический университет» | | |
|  | | |
| Кафедра прикладной математики | | |
| Практическое задание № 4 | | |
| по дисциплине «Методы принятия оптимальных решений» | | |
| **Критерий . Влияние группирование на мощность критерия** | | |
|  | | |
|  | Бригада | ПМ-13 Буданцев дмитрий |
| . | ПМ-13 Форкин Кирилл |
|  |  |
|  |  |
|  |  |
|  |  |
| Преподаватели | Лемешко Борис Юрьевич |
|  |  |
| Новосибирск,2024 | | |

**1. Цель исследования**

*Цель занятия* – проследить, как влияет выбор числа интервалов и способ разбиения области определения случайной величины на вид гистограммы, как отражаются эти факторы на мощности критерия  Пирсона.

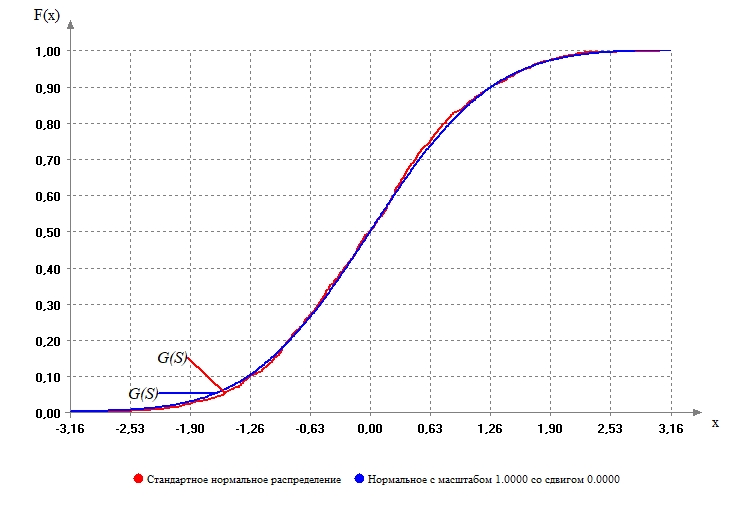
**2. Этапы исследования**

1. Смоделировать выборку в соответствии со стандартным нормальным законом объёмом 
2. Проследить, как меняется вид гистограммы при изменении интервалов в зависимости от способа группирования:
   * при равномерном группировании и числе интервалов 
   * при равночастотном группировании и числе интервалов 
   * при равновероятном группировании и числе интервалов 
   * при асимптотически оптимальном группировании и числе интервалов 
3. Исследование мощности критерия  Пирсона **при проверке простых** гипотез в случае проверки нормальности относительно наиболее близкого логистического закона с параметром масштаба 
4. Исследовать мощность критерия **при интервалах Неймана-Пирсона**
5. Исследовать мощность критерия при **оптимальном группировании** для данной пары конкурирующих законов.
6. Сформировать таблицу мощностей. Сделать выводы.

**3. Выполнение исследования**

**Исследование №1**

Эмпирическая функция распределения *(см. график 1)* выборки *model\_1*, смоделированная в соответствии со *стандартным нормальным законом* с объёмом 

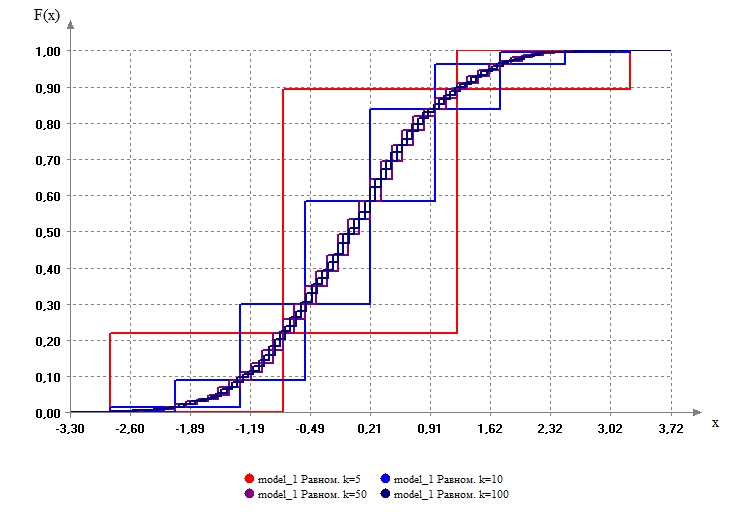


*График 1*. Эмпирическая функция распределения выборки *model\_1*

**Исследование №2**

**Равномерный метод группирования**

Результаты группирования выборки *model\_1* при использовании **равномерного метода****группирования** *(см. график 2)* при числе интервалов 

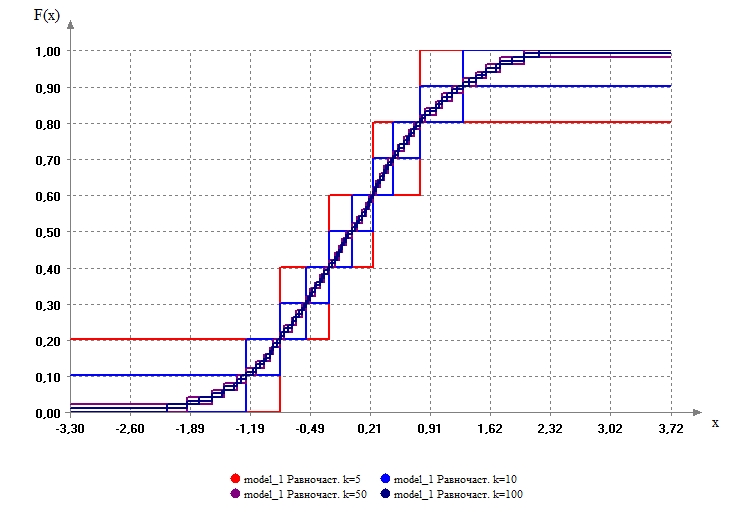


*График 2*. Гистограмма группированных выборок при использовании равномерного метода группирования

Из графика *(см. график 2)* видно, длина интервалов у всех одинаковая. Очевидно, что здесь прослеживается *обратная зависимость* числа интервалов от длины интервала.

**Равночастотный метод группирования**

Результаты группирования выборки *model\_1* при использовании **равночастотного метода группирования** *(см. график 3)* при числе интервалов 

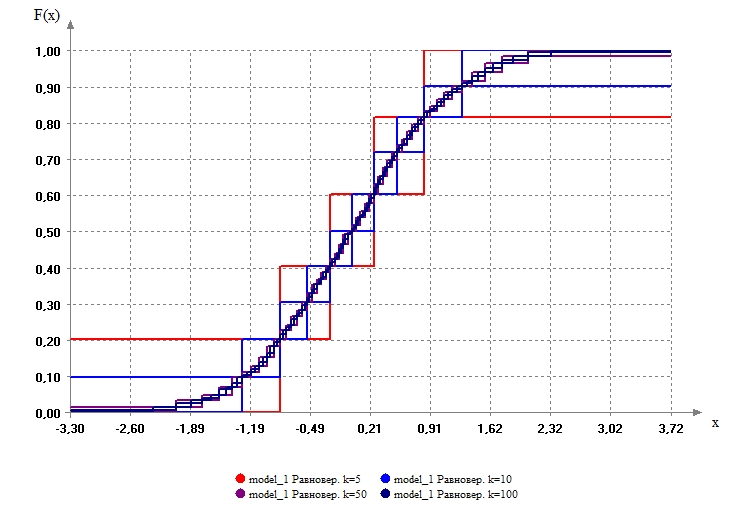


*График 3*. Гистограмма группированных выборок при использовании **равночастотного метода группирования**

Из графика *(см. график 3)* видно, что длина интервалов отличается. Данный метод основан на том, что в один интервал должно попасть равное количество значений выборки. Так как, выборка моделировалась в соответствии с нормальным законом, можно заметить, что большее число значений выборки находится вблизи среднего значения, как и должно быть. Аналогично объясняется то, почему граничные интервалы имеют большую длину, чем остальные.

**Равновероятный метод группирования**

Результаты группирования выборки *model\_1* при использовании **равновероятного метода группирования** *(см. график 4)* при числе интервалов 



*График 4*. Гистограмма группированных выборок при использовании равновероятного метода группирования

Из *графика 4* видно, что он идентичен *графику 3*. Из этого можно сделать ложный вывод о том, что данные методы группируют **одинаково**. На самом деле это далеко не так. В данном примере так совпало.

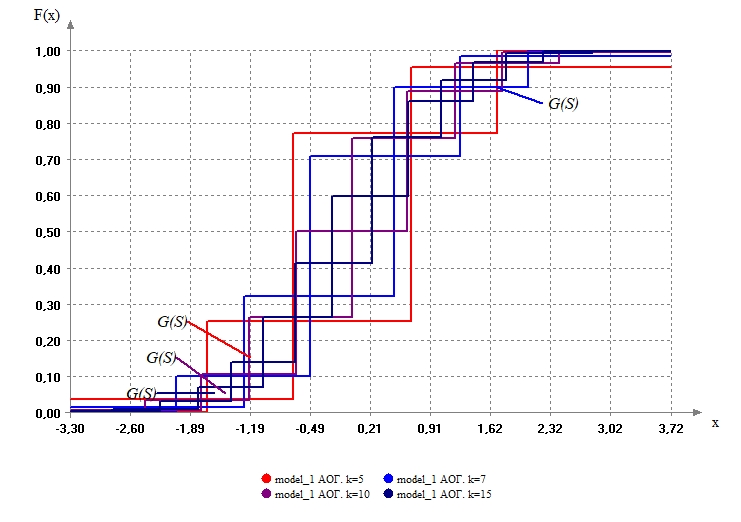
Метод **равновероятного группирования** работает по другому принципу. В процессе группирования данных он вычисляет *квантили* параметрической модели (в нашем случае *стандартного нормального закона*) с *"равновероятностным шагом"* , где -количество интервалов. Например, для  это будет выглядеть так:

|  |  |  |
| --- | --- | --- |
|  |  | Квантили |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |
| 5 |  |  |

Если сопоставить значения квантилей с *графиком 4*, то можно заметить, что границы интервалов ровно такие же. Для остальных значений  интервалы можно получить аналогичным образом.

**Асимптотически оптимальный метод группирования**

Результаты группирования выборки *model\_1* при использовании **асимптотически оптимального метода группирования** *(см. график 5)* при числе интервалов 



*График 5*. Гистограмма группированных выборок при использовании асимптотически оптимального метода группирования

Из графика *(см. график 5)*, тяжело найти какую-то зависимость построения интервалов от их числа. Сам метод основан на **мере информации Фишера**. Очевидно, что группирование наблюдений приводит к **потере информации**. Таким образом, если - матрица потерь информации, вызванных группированием, где - информационная матрица Фишера по негруппированным наблюдениям, -информационная матрица Фишера по группированным данным, то становится очевидным, что чем ближе матрица  к , тем меньше **потеря информации** от группирования и тем **выше мощность** критериев при *близких конкурирующих гипотезах*.

Таким образом, задача метода заключается к *приближению*  к . От сюда следует, что граничные точки интервалов получаются в результате максимизации определителя информационной матрицы т.е. задача сводится к

.

Для достаточно широкого ряда распределений уже были получены **оптимальные граничные точки интервалов**, которые занесены в таблицы. Следовательно воспользовавшись ими можем получить интервалы полученные на *графике 5* для данной параметрической модели. Например, для  они будут выглядеть так:

|  |  |
| --- | --- |
|  | Интервалы |
| 1 |  |
| 2 |  |
| 3 |  |
| 4 |  |
| 5 |  |

Если сопоставить эти интервалы с *графиком 5* можно понять, что это в точности они. Аналогично, можно получить для другого числа интервалов *(в таблице представлено только до 15 интервалов)*

Таким образом, из всего вышесказанного можно сделать вывод, что *граничные точки* подбираются так, чтобы группирование данных приводила к меньшей **потери информации**. Исходя из этого, можно заметить, что для данной параметрической модели, в отличие от других методов, интервалы, которые близки к *среднему значению*, по длине больше остальных.

**Исследование №3**

Результаты исследования *мощности* критерия Хи-квадрат при близко конкурирующих гипотезах *(см. таблица 1)* используя выборки с объёмом  наблюдений,  симуляций и  при **равночастотном**, **равновероятном** и **асимптотически оптимальном** группированиях.

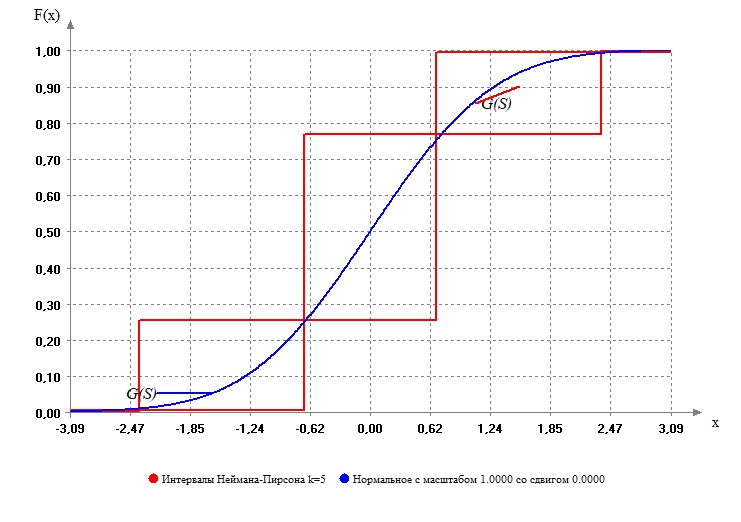
|  |  |
| --- | --- |
| Метод группирования |  |
| Равночастотный |  |
| Равновероятный |  |
| АОГ |  |

*Таблица 1*. Мощность критерия *Хи-квадрат Пирсона* при **равночастотном**, **равновероятном** и **асимптотически оптимальном** методах группированиях

Таким образом, из полученных результатов *(см. таблица 1)*, видно, что мощность при использовании **АОГ** существенно выше, чем при **равночастотном** и **равновероятном** методах группирования. Как было сказано ранее, при выполнении *пункта №2*, граничные точки интервалов для данной параметрической модели при *равночастотном* и *равновероятном* методах группирования **идентичны**. По этой причине исследование разности мощности при этих методах на данной параметрической модели **не имеет смысла**.

**Исследование №4**

**Эмпирическая функция** распределения при интервалах **Неймана-Пирсона** *(см. график 6)*



*График 6*. **Эмпирическая функция** распределения при *интервалах* **Неймана-Пирсона**

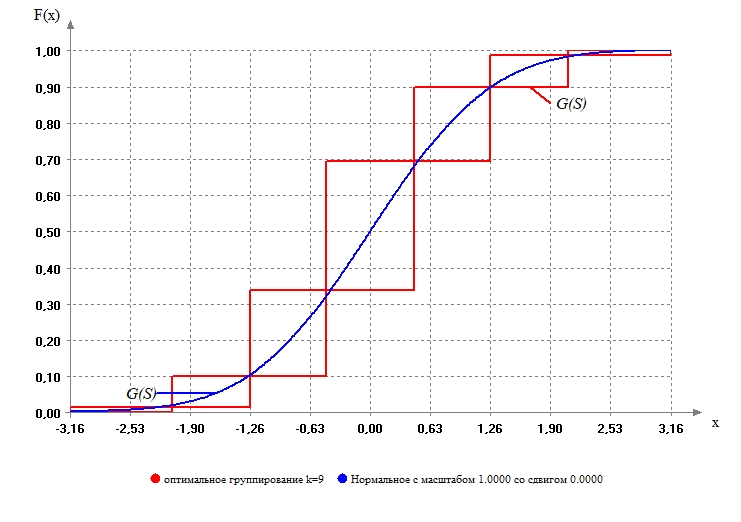
Результаты исследования *мощности* критерия Хи-квадрат при близко конкурирующих гипотезах *(см. таблица 2)* используя выборки с объёмом  наблюдений,  симуляций и  при **интервалах Неймана-Пирсона**.

|  |  |
| --- | --- |
| Метод группирования |  |
| Неймана-Пирсона |  |

*Таблица 2*. Мощность критерия *Хи-квадрат Пирсона* при интервалах **Неймана-Пирсона**

**Исследование №5**

**Эмпирическая функция** распределения при **оптимальном группировании** *(см. график 7)*



*График 7*. **Эмпирическая функция** распределения при *интервалах* **оптимальном группировании**

Результаты исследования *мощности* критерия **Хи-квадрат** при близко конкурирующих гипотезах *(см. таблица 3)* используя выборки с объёмом  наблюдений,  симуляций и  при **оптимальном группировании**.

|  |  |
| --- | --- |
| Метод группирования |  |
| Оптимальный |  |

*Таблица 3*. Мощность критерия *Хи-квадрат Пирсона* при **оптимальном группировании**

**Исследование №6**

Итоговая таблица *(см. таблица 4)* сформированная по результату выполнений **исследований №3-5**, где производилась исследование *мощности* критерия **Хи-квадрат** при близко конкурирующих гипотезах используя выборки с объёмом  наблюдений,  симуляций и  при **различных методах группирования**

|  |  |
| --- | --- |
| Метод группирования |  |
| Равночастотный |  |
| Равновероятный |  |
| АОГ |  |
| Неймана-Пирсона |  |
| Оптимальный |  |

*Таблица 4*. Мощность критерия *Хи-квадрат Пирсона* при **различных методах группирования**

Из результатов *(см. таблицу 4)* можно сделать вывод, что *оптимальное группирование*, является самым лучшим методом среди методов на  интервалах, поскольку при данном методе мощность критерия **Хи-квадрат** является *одним из самых наибольших* среди других. Удивительно, что при  метод **Неймана-Пирсона** показывает *наибольшее значение мощности*, хотя количество разбиений меньше, чем у других, из-за чего потери информации должно быть больше, но на мощность это не отразилось. Исходя из этого, можно сделать вывод, что метод **Неймана-Пирсона** из всех применяемых методов справился **лучше всех**.