МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ

ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ

ВЫСШЕГО ОБРАЗОВАНИЯ

«НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра систем сбора и обработки данных

**ЛАБОРАТОРНАЯ РАБОТА №7**

по дисциплине: Компьютерные технологии моделирования и анализа данных

на тему:Исследование критериев проверки отклонения от нормального закона. Часть 1.

Вариант №2

Факультет: ФПМИ

Группа: ПММ-21

Выполнил: Сухих А.С., Черненко Д.А.

Проверил: д.т.н., профессор Лемешко Борис Юрьевич

Дата выполнения: 22.12.22

Отметка о защите:

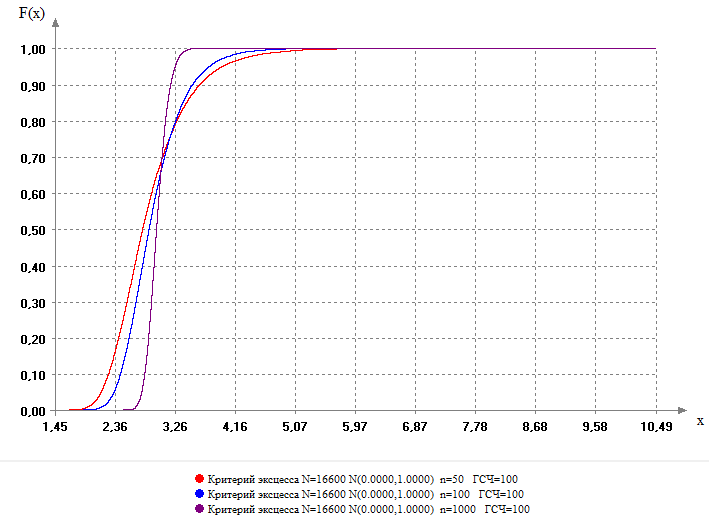
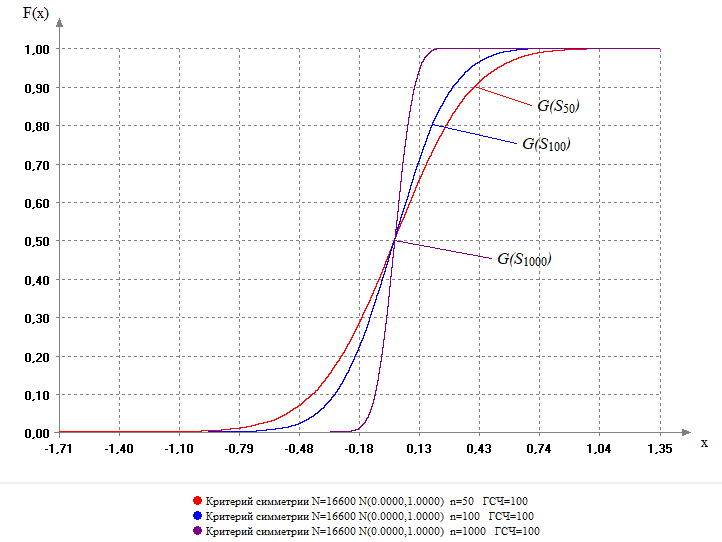
Новосибирск 2022

**Цель работы:** Исследование распределений статистик критериев, используемых при проверке отклонения эмпирических распределений наблюдаемых величин от нормального закона (в том числе критериев проверки гипотез о симметричности и о значении эксцесса при различных наблюдаемых законах). Исследование распределений статистик критериев Шапиро–Уилка, Эппса–Палли, Д’Агостино. Исследование и сравнение мощности критериев относительно близких конкурирующих гипотез.

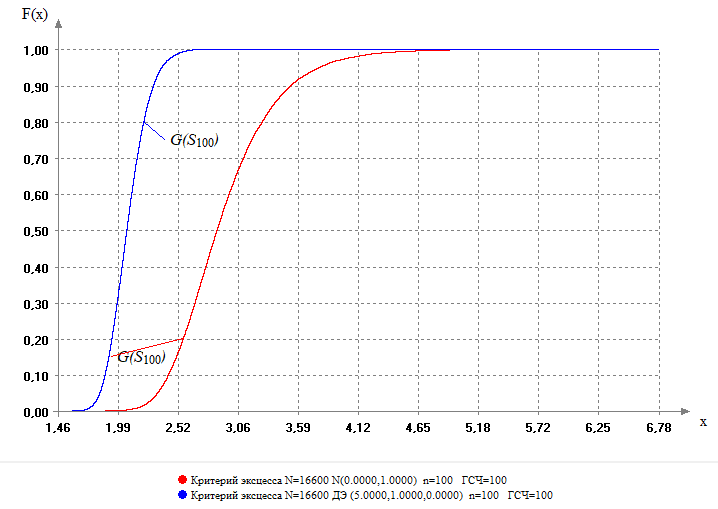
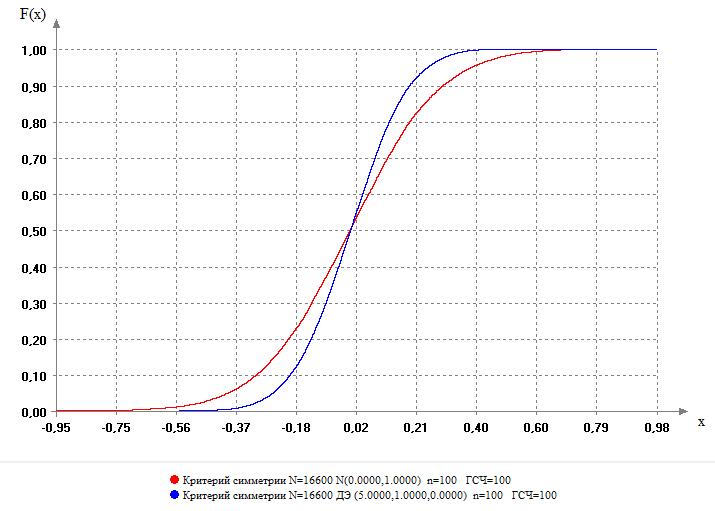
|  |  |
| --- | --- |
| № п/п |  |
| 2 | Двустороннее экспоненциальное с параметром формы 5 |

1. Исследовать зависимость распределений статистик критерия симметричности и критерия эксцесса от объема выборок в случае принадлежности наблюдений нормальному закону. При некотором объеме выборок смоделировать эмпирические распределения статистик критерия симметричности и критерия эксцесса при законе двустороннего экспоненциального распределения. Сравнить распределения статистик со случаем нормальности наблюдаемого закона.

По объемам выборок n = 50, 100, 1000 были смоделированы статистики критериев симметричности и эксцесса в случае принадлежности нормальному закону.

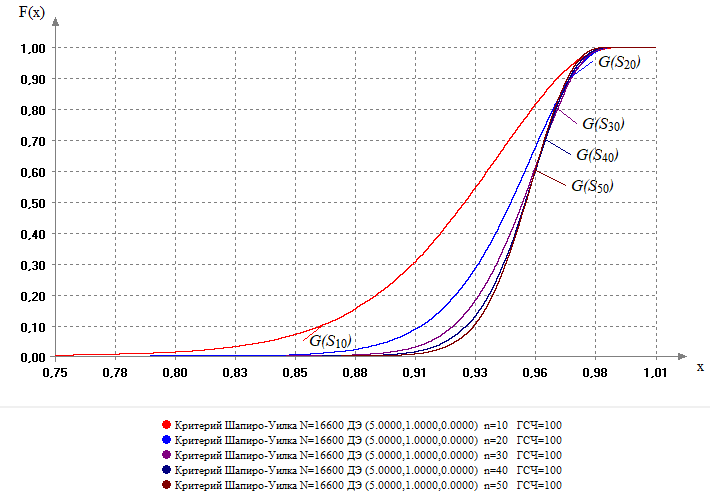


На примере объема выборок n = 100 сравним распределения статистик критериев симметричности и эксцесса нормального распределения и двустороннего экспоненциального с параметром формы 5.



2. Исследовать распределения статистик критерия Шапиро-Уилка при различных объемах выборок. Оценить близость получаемых эмпирических распределений статистик к «теоретическим» по процентным точкам таблиц, соответствующим данному критерию. При некотором объеме выборок (*n*=10) смоделировать распределение статистики критерия при обобщённом нормальном законе (двустороннее экспоненциальное) при параметре формы, равном 4÷7. Сравнить с ситуацией, соответствующей справедливой проверяемой гипотезе о нормальном законе. Оценить мощность критерия.

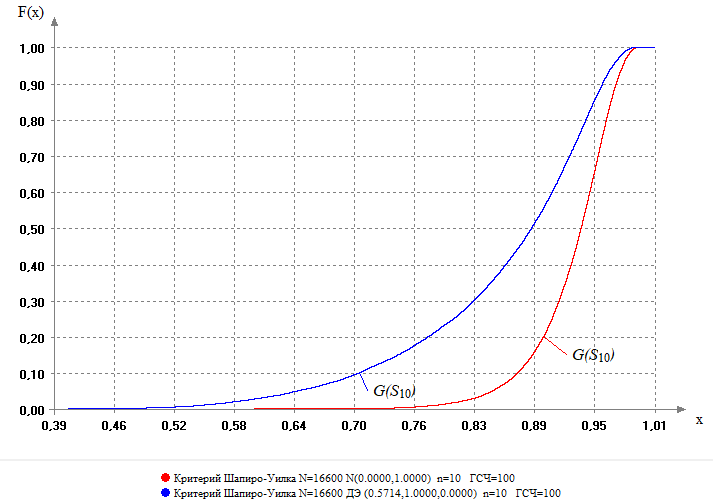
Распределения статистик критерия Шапиро-Уилка при двустороннем экспоненциальном распределении с объемами выборок n = 10,20,30,40 и 50 приведены на рисунке.



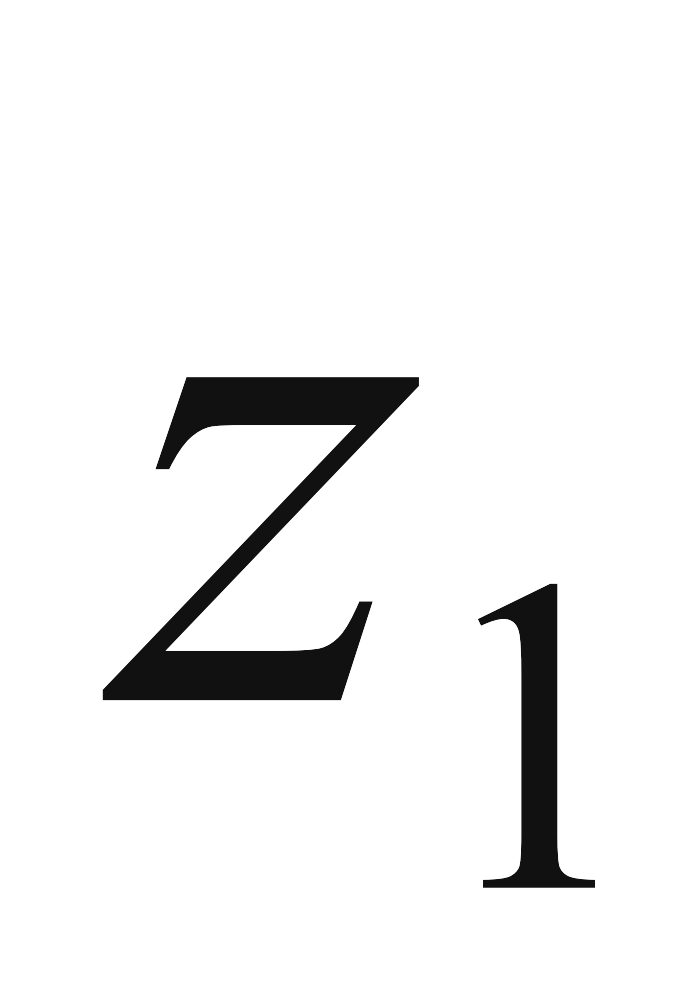
Экспериментальные и «теоретические» значения статистик критерия Шапиро-Уилка в процентных точках при разных объемах выборок приведены в таблице.

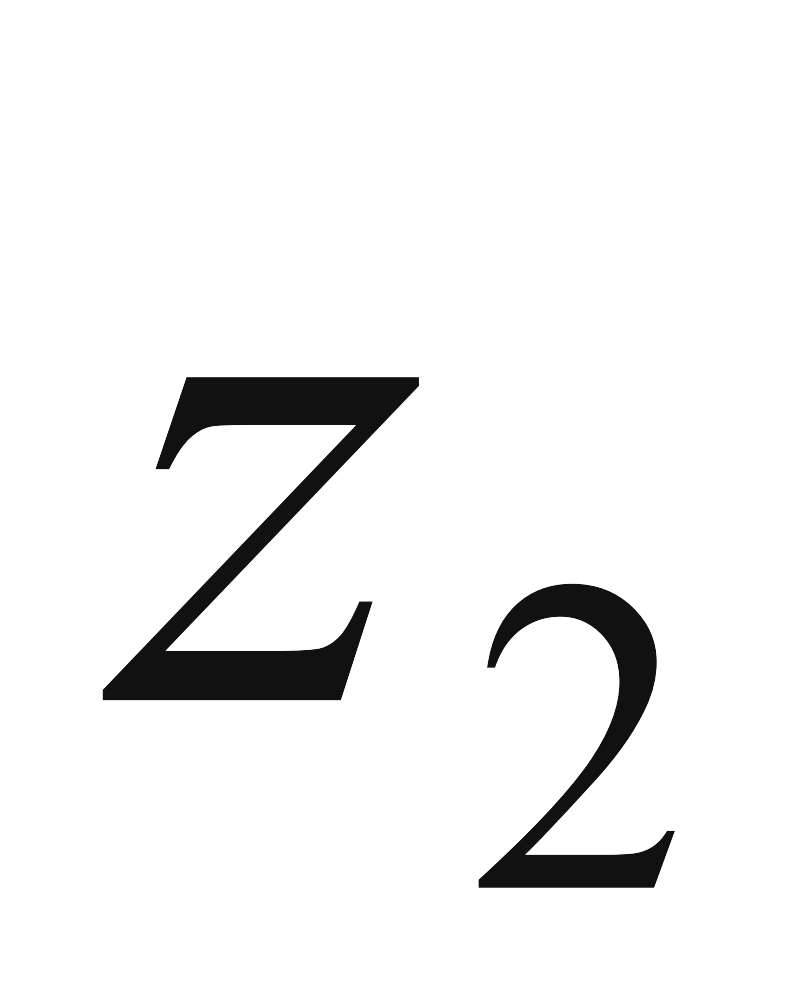
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | p=0.01 | | p=0.05 | |
| S, теор | S, эксп | S, теор | S, эксп |
| n=10 | 0.781 | 0.793 | 0.842 | 0.844 |
| n=20 | 0.868 | 0.867 | 0.905 | 0.895 |
| n=30 | 0.9 | 0.892 | 0.927 | 0.913 |
| n=40 | 0.919 | 0.902 | 0.94 | 0.92 |
| n=50 | 0.93 | 0.91 | 0.947 | 0.924 |

Были смоделированы распределения статистики критерия при обобщенном нормальном распределении и нормальном распределении. Их вид показан на графике. Мощность критерия при 𝛼 = 0,05 и при H0 – обобщенное нормальное распределение, H1 – нормальное распределение равна 1 - 𝛽 = 0,12747.



3. Аналогично предыдущему пункту исследовать распределения статистики критерия Эппса-Палли.

4. Исследовать распределения статистики . Проверить близость эмпирических распределений статистики стандартному нормальному закону.

5. Исследовать распределения статистики ******. Проверить близость эмпирических распределений статистики стандартному нормальному закону.

6. Оценить мощность критериев со статистиками (7.3), (7.4), (7.5), (7.6) относительно заданной альтернативы.

**Для Пирсона АОГ, k = 10:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Простая гипотеза** | | | **Сложная гипотеза** | | |
|  | | | | | |
|  |  |  |  |  |  |
| 0.15 | 0.199639 | 0.248614 | 0.902229 | 0,153373 | 0,207892 | 0,896024 |
| 0.1 | 0.139699 | 0.17994 | 0.857169 | 0,101988 | 0,141566 | 0,845361 |
| 0.05 | 0.073253 | 0.101386 | 0.772831 | 0,0512048 | 0,0759036 | 0,743133 |
| 0.025 | 0.0372892 | 0.0570482 | 0.685241 | 0,0246988 | 0,0392169 | 0,63241 |
| 0.01 | 0.0155422 | 0.025 | 0.560964 | 0,0109036 | 0,0157831 | 0,476928 |

**Для Никулина АОГ, k = 10:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Простая гипотеза** | | | **Сложная гипотеза** | | |
|  | | | | | |
|  |  |  |  |  |  |
| 0.15 | 0.199699 | 0.248494 | 0.902169 | 0.148373 | 0.190361 | 0.87512 |
| 0.1 | 0.139639 | 0.17988 | 0.857048 | 0.0968675 | 0.128494 | 0.818072 |
| 0.05 | 0.0733133 | 0.101265 | 0.772711 | 0.0453012 | 0.0672892 | 0.707952 |
| 0.025 | 0.0372289 | 0.056988 | 0.68506 | 0.0213855 | 0.0328916 | 0.585 |
| 0.01 | 0.0155422 | 0.0249398 | 0.560663 | 0.00777108 | 0.0121687 | 0.439157 |

Как можно увидеть, мощность критерия Никулина, для данного метода группирования, выше как для сложных гипотез, так и для простых, правда, для простой гипотезы их отличие в плане мощности не столь критично.

4. Исследовать мощность критерия Никулина в зависимости от числа интервалов и выполнить сравнение с аналогичными исследованиями по критерию Пирсона.

**Пирсон:** простая гипотеза, АОГ при k = 5, 10, 15:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | | | | | | | | |
| k = 5 | | | k = 10 | | | k = 15 | | |
| n = 50 | n=100 | n=1000 | n=50 | n=100 | n=1000 | n = 50 | n=100 | n=1000 |
| 0.15 | 0.212651 | 0.24259 | 0.754157 | 0.199639 | 0.248614 | 0.902229 | 0.15006 | 0.20512 | 0.895542 |
| 0.1 | 0.153675 | 0.177651 | 0.67988 | 0.139699 | 0.17994 | 0.857169 | 0.1 | 0.14168 | 0.848494 |
| 0.05 | 0.086506 | 0.106988 | 0.56012 | 0.073253 | 0.101386 | 0.772831 | 0.05006 | 0.07487 | 0.750843 |
| 0.025 | 0.050241 | 0.065 | 0.446566 | 0.0372892 | 0.0570482 | 0.685241 | 0.02506 | 0.03927 | 0.642048 |
| 0.01 | 0.0237952 | 0.0304217 | 0.323916 | 0.0155422 | 0.025 | 0.560964 | 0.01006 | 0.01385 | 0.494699 |

**Никулин:** простая гипотеза, АОГ при k = 5, 10, 15:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | | | | | | | | |
| k = 5 | | | k = 10 | | | k = 15 | | |
| n = 50 | n=100 | n=1000 | n=50 | n=100 | n=1000 | n = 50 | n=100 | n=1000 |
| 0.15 | 0.212651 | 0.24247 | 0.753976 | 0,199699 | 0.248494 | 0.902169 | 0.15868 | 0.205 | 0.895361 |
| 0.1 | 0.153675 | 0.177651 | 0.679819 | 0,139639 | 0.17988 | 0.857048 | 0.10422 | 0.14163 | 0.848434 |
| 0.05 | 0.0864458 | 0.106988 | 0.559759 | 0,0733133 | 0.101265 | 0.772711 | 0.05127 | 0.07488 | 0.750663 |
| 0.025 | 0.050241 | 0.0648193 | 0.446506 | 0,0372289 | 0.056988 | 0.68506 | 0.02386 | 0.03928 | 0.641807 |
| 0.01 | 0.0237349 | 0.0304819 | 0.323916 | 0,0155422 | 0.0249398 | 0.560663 | 0.00807 | 0.01380 | 0.494578 |

**Пирсон:** простая гипотеза, равновероятное, k = 5, 10, 15:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | | | | | | | | |
| k = 5 | | | k = 10 | | | k = 15 | | |
| n = 50 | n=100 | n=1000 | n=50 | n=100 | n=1000 | n = 50 | n=100 | n=1000 |
| 0.15 | 0.15006 | 0.238494 | 0.820843 | 0.180723 | 0.21741 | 0.724518 | 0.17735 | 0.20855 | 0.684157 |
| 0.1 | 0.1 | 0.180301 | 0.757169 | 0.123253 | 0.153072 | 0.644217 | 0.1228 | 0.14 | 0.600181 |
| 0.05 | 0.0500602 | 0.100602 | 0.638012 | 0.0624699 | 0.0825301 | 0.513916 | 0.06584 | 0.079217 | 0.468012 |
| 0.025 | 0.0250602 | 0.0593976 | 0.51994 | 0.0346386 | 0.0458434 | 0.411084 | 0.03548 | 0.039939 | 0.353735 |
| 0.01 | 0.0100602 | 0.0290964 | 0.39494 | 0.0143373 | 0.0207831 | 0.280422 | 0.01259 | 0.018374 | 0.227711 |

**Никулин:** простая гипотеза, равновероятное, k = 5, 10, 15:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | | | | | | | | |
| k = 5 | | | k = 10 | | | k = 15 | | |
| n = 50 | n=100 | n=1000 | n=50 | n=100 | n=1000 | n = 50 | n=100 | n=1000 |
| 0.15 | 0.212651 | 0.24247 | 0.753976 | 0.199699 | 0.248494 | 0.90217 | 0.158675 | 0.205 | 0.89536 |
| 0.1 | 0.153675 | 0.177651 | 0.679819 | 0.139639 | 0.17988 | 0.85705 | 0.104217 | 0.14163 | 0.84843 |
| 0.05 | 0.0864458 | 0.106988 | 0.559759 | 0.0733133 | 0.101265 | 0.77271 | 0.051265 | 0.07488 | 0.75066 |
| 0.025 | 0.050241 | 0.0648193 | 0.446506 | 0.0372289 | 0.056988 | 0.68506 | 0.023855 | 0.03928 | 0.64181 |
| 0.01 | 0.0237349 | 0.0304819 | 0.323916 | 0.0155422 | 0.0249398 | 0.56066 | 0.0080723 | 0.01379 | 0.49458 |

**Пирсон:** cложная гипотеза, АОГ, k = 5, 10, 15:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | | | | | | | | |
| k = 5 | | | k = 10 | | | k = 15 | | |
| n = 50 | n=100 | n=1000 | n=50 | n=100 | n=1000 | n = 50 | n=100 | n=1000 |
| 0.15 | 0.186084 | 0.15006 | 0.60994 | 0.153373 | 0.207892 | 0.896024 | 0.130181 | 0.16127 | 0.867771 |
| 0.1 | 0.135422 | 0.1 | 0.516024 | 0.101988 | 0.141566 | 0.845361 | 0.083253 | 0.10494 | 0.803133 |
| 0.05 | 0.0766265 | 0.0500602 | 0.379578 | 0.0512048 | 0.0759036 | 0.743133 | 0.039458 | 0.054217 | 0.678193 |
| 0.025 | 0.0412048 | 0.0250602 | 0.277952 | 0.0246988 | 0.0392169 | 0.63241 | 0.020422 | 0.025422 | 0.553313 |
| 0.01 | 0.0171687 | 0.0100602 | 0.174277 | 0.0109036 | 0.0157831 | 0.476928 | 0.006868 | 0.01 | 0.375723 |

**Никулин:** сложная гипотеза, АОГ, k = 5, 10, 15:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | | | | | | | | |
| k = 5 | | | k = 10 | | | k = 15 | | |
| n = 50 | n=100 | n=1000 | n=50 | n=100 | n=1000 | n = 50 | n=100 | n=1000 |
| 0.15 | 0.184277 | 0.214759 | 0.70193 | 0.148373 | 0.190361 | 0.87512 | 0.12633 | 0.154096 | 0.84259 |
| 0.1 | 0.126325 | 0.153976 | 0.62048 | 0.096868 | 0.128494 | 0.81807 | 0.08024 | 0.101205 | 0.769699 |
| 0.05 | 0.0687952 | 0.089518 | 0.48735 | 0.045301 | 0.067289 | 0.70795 | 0.03898 | 0.050843 | 0.640542 |
| 0.025 | 0.035241 | 0.048434 | 0.36843 | 0.021386 | 0.032892 | 0.585 | 0.01771 | 0.023916 | 0.512651 |
| 0.01 | 0.0149398 | 0.021386 | 0.24355 | 0.007771 | 0.012169 | 0.43916 | 0.00608 | 0.008916 | 0.352048 |

**Пирсон:** сложная гипотеза, равновероятное, k = 5, 10, 15:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | | | | | | | | |
| k = 5 | | | k = 10 | | | k = 15 | | |
| n = 50 | n=100 | n=1000 | n=50 | n=100 | n=1000 | n = 50 | n=100 | n=1000 |
| 0.15 | 0.163313 | 0.185361 | 0.670301 | 0.16753 | 0.181325 | 0.470602 | 0.16416 | 0.18355 | 0.51512 |
| 0.1 | 0.103373 | 0.120904 | 0.54994 | 0.109639 | 0.123494 | 0.369277 | 0.11355 | 0.12608 | 0.41753 |
| 0.05 | 0.0513253 | 0.0600602 | 0.361566 | 0.0565663 | 0.063012 | 0.245482 | 0.05789 | 0.06277 | 0.281265 |
| 0.025 | 0.0263855 | 0.0292771 | 0.239578 | 0.0316265 | 0.0343976 | 0.158494 | 0.02928 | 0.031988 | 0.188012 |
| 0.01 | 0.0105422 | 0.0118072 | 0.127651 | 0.0128313 | 0.0139759 | 0.085904 | 0.01121 | 0.01392 | 0.109639 |

**Никулин:** сложная гипотеза, равновероятное, k = 5, 10, 15:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | | | | | | | | |
| k = 5 | | | k = 10 | | | k = 15 | | |
| n = 50 | n=100 | n=1000 | n=50 | n=100 | n=1000 | n = 50 | n=100 | n=1000 |
| 0.15 | 0.184277 | 0.214759 | 0.701867 | 0.146506 | 0.190301 | 0.87512 | 0.12627 | 0.154157 | 0.84259 |
| 0.1 | 0.126325 | 0.153976 | 0.620542 | 0.092469 | 0.128494 | 0.81771 | 0.08024 | 0.101205 | 0.76988 |
| 0.05 | 0.068795 | 0.089458 | 0.487289 | 0.043615 | 0.067289 | 0.70741 | 0.03898 | 0.050783 | 0.640422 |
| 0.025 | 0.035241 | 0.048374 | 0.368554 | 0.020783 | 0.032892 | 0.58506 | 0.01771 | 0.023916 | 0.51253 |
| 0.01 | 0.014939 | 0.021386 | 0.243494 | 0.007530 | 0.012169 | 0.43916 | 0.00608 | 0.008916 | 0.352108 |

**Вывод:**

Выяснилось, что критерий Никулина лучше, чем критерий Пирсона, для равновероятного группирования сложных гипотез, в случае маленьких выборок (n = 50, 100), для больших выборок (n = 1000) критерий Пирсона работает лучше. Что касается АОГ, здесь при увеличении выборки, а также k, мощность критерия, по сравнению с Пирсоном, возрастает.

Если рассматривать простые гипотезы с методом группирования АОГ, то здесь эти два критерия более-менее равны по своей эффективности, чего не скажешь о равновероятном группировании, в котором критерий Никулина при k = 10 и k =15 работает хуже для n = 100, 1000, для k = 15 при объёме выборки n =50 критерий Никулина сработал лучше конкурирующего. Для маленьких k = 5 Никулин работает хорошо при больших объёмах выборки, с маленькими выборками лучше работает Пирсон.

По итогу можно сказать, что применение этих двух критериев нужно рассматривать, отталкиваясь от поставленной задачи.