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 №3**

по дисциплине: Компьютерные технологии моделирования и анализа данных

на тему: Экспериментальное исследование свойств критерия согласия Пирсона

Вариант №2

Факультет: ФПМИ

Группа: ПММ-21

Выполнил: Сухих А.С., Черненко Д.А.

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

Дата выполнения: 24.11.22

Отметка о защите:

Новосибирск 2022

**Цель работы.** Исследование влияния способов группирования на предельные распределения статистики критерия согласия хи-квадрат Пирсона при простых и сложных гипотезах (при использовании для вычисления оценок по негруппированным данным метода максимального правдоподобия).

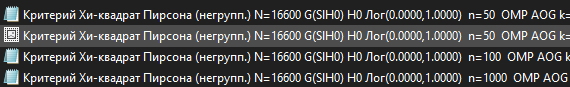
|  |  |  |
| --- | --- | --- |
| № п/п | Гипотеза | Гипотеза |
| 2 | Логистическое | Нормальное |

Исследовать распределения статистики критерия для простых и сложных гипотез (при использовании оценок максимального правдоподобия по негруппированным данным) при справедливой нулевой и при справедливой конкурирующей гипотез.

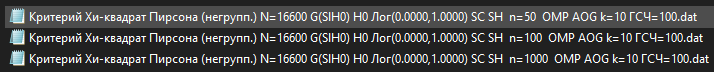
1. В соответствии с заданным наблюдаемым распределением (гипотеза ) смоделировать эмпирические распределения статистики критерия при простой гипотезе (а) (по выборке не оцениваются параметры), для сложной гипотезы (б) (по выборке оцениваются все параметры).

Эмпирические распределения были смоделированы:

Простые гипотезы:



Сложные гипотезы:



2. Идентифицировать построенные законы распределения (найти аналитические модели, наиболее хорошо описывающие эмпирические распределения).

**Простые гипотезы:**

1. n = 50, подошло распределение SL - Джонсона. (согласуется)

Вероятность согласия: P=1-G(S|H0)= 0.1281101920643576

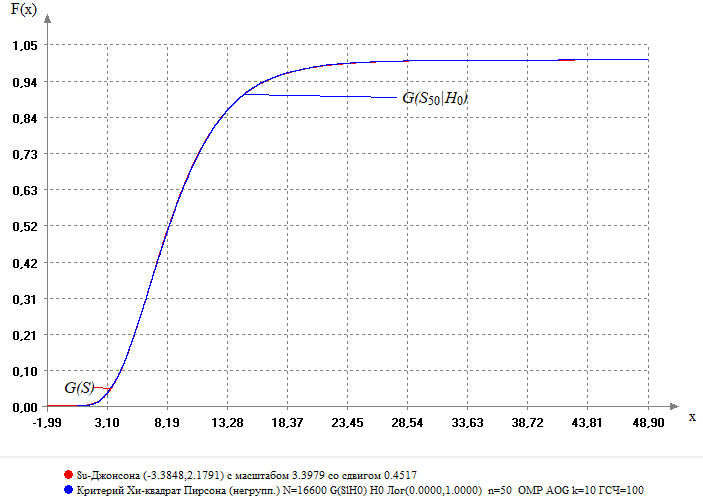
\* Хи-квадрат Пирсона S=2.868674698922462 P=0.7202237004477087

**Лучше данного распределения подходит**

n = 50, подошло распределение Su - Джонсона.

Вероятность согласия: P=1-G(S|H0)= 0.1293948831991909

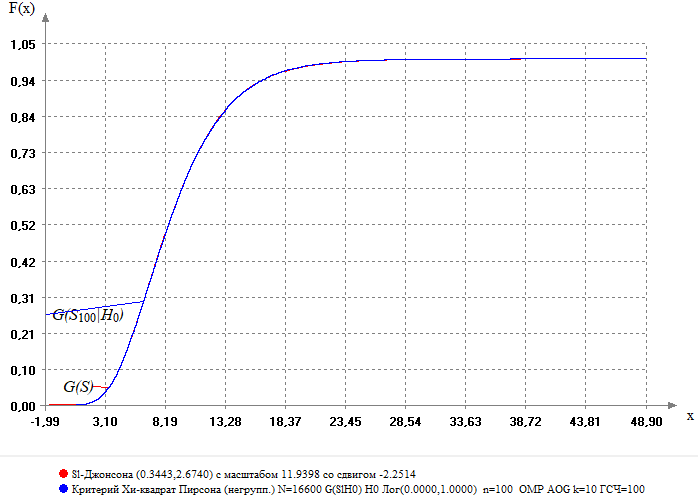
\* Хи-квадрат Пирсона S=8.556626502927484 P=0.1281101920643576



1. n = 100, подошло распределение SL - Джонсона. (согласуется)

Вероятность согласия: P=1-G(S|H0)= 0.3732418418487965

\* Хи-квадрат Пирсона S=5.362650590667649 P=0.3732418418487965

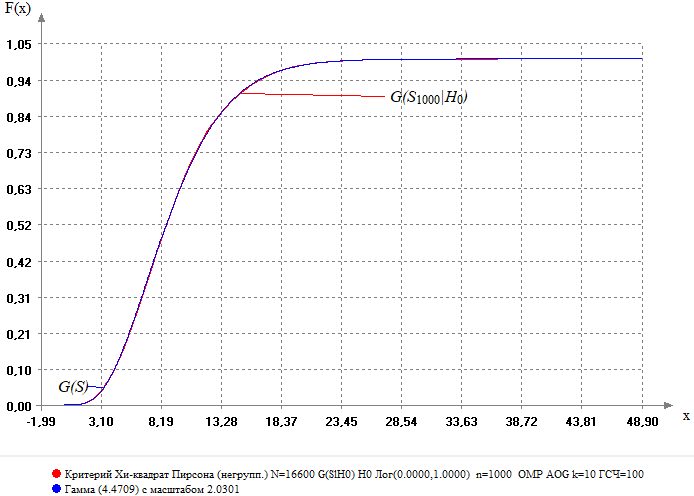


Распределение Su - Джонсона при данном количестве элементов n отвергается.

1. n = 1000, подошло лишь распределение Гамма (4.4709,2.0301) (согласуется)

Вероятность согласия: P=1-G(S|H0) = 0.1876196165838482

\* Хи-квадрат Пирсона S=10.01686747145985 P=0.1876196165838482

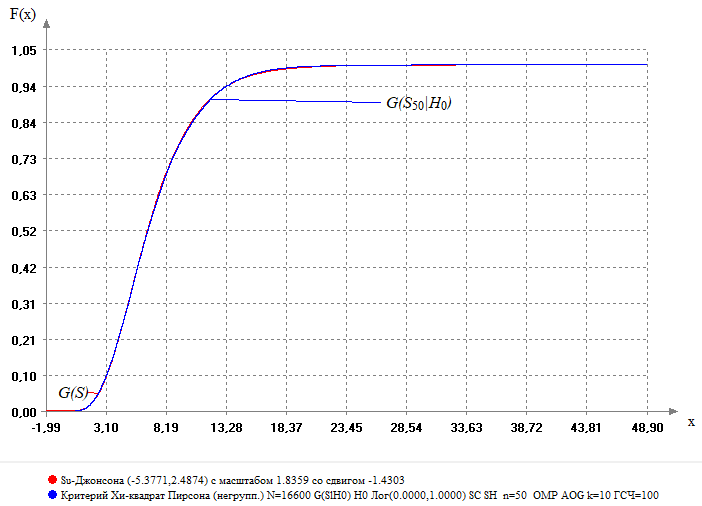


**Сложные гипотезы:**

1. n = 50, лучшим распределением оказалась Su - Джонсона (отвергается)

Вероятность согласия: P=1-G(S|H0) = 0.01708899423210665

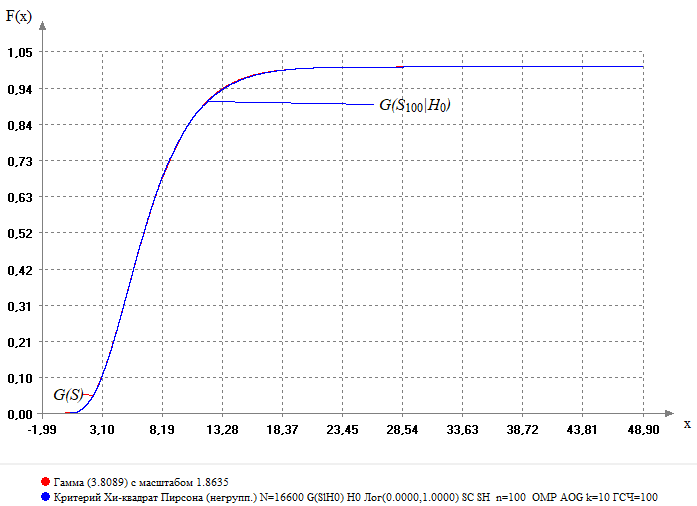
\* Хи-квадрат Пирсона S=13.77710843957602 P=0.01708899423210665



1. n = 100, лучшим распределением оказалась Гамма (3.8089,1.8635) (согласуется)

Вероятность согласия: P=1-G(S|H0) = 0.4292562857794421

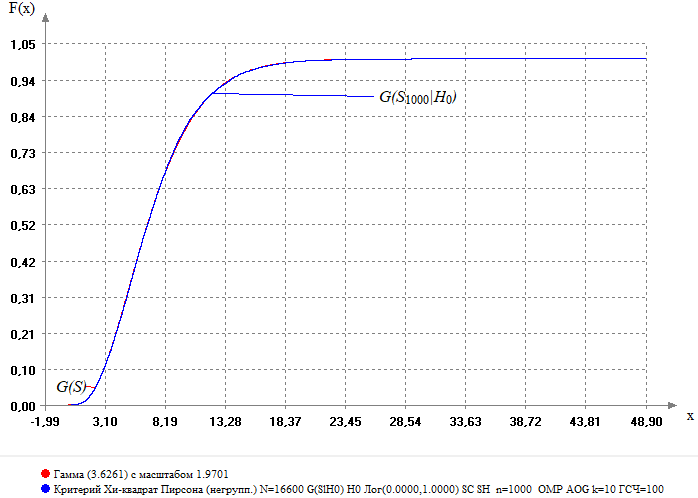
\* Хи-квадрат Пирсона S=6.996385594379026 P=0.4292562857794421



1. n = 1000, лучшим распределением оказалась Гамма (3.6261,1.9701)

Вероятность согласия: P=1-G(S|H0) = 0.2486160361780861

\* Хи-квадрат Пирсона S=9.056626448149206 P=0.2486160361780861

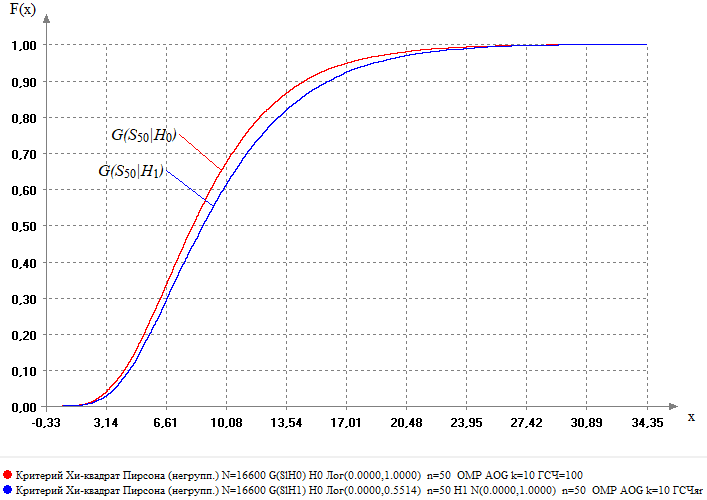


3. Повторить п.1, моделируя выборку по закону, соответствующему гипотезе , а оценивая по этой выборке параметры закона, соответствующего гипотезе .

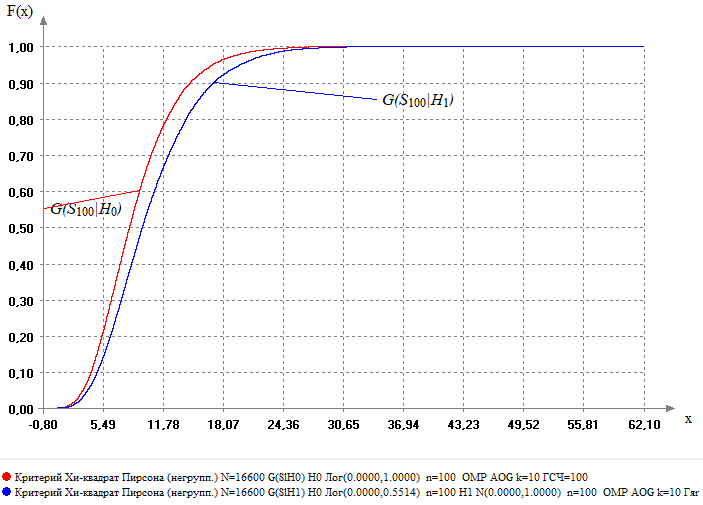
Наибольший интерес представляет способность критериев согласия различать близкие конкурирующие законы, то есть мощность критериев относительно близких конкурирующих гипотез.

**Простая гипотеза:**

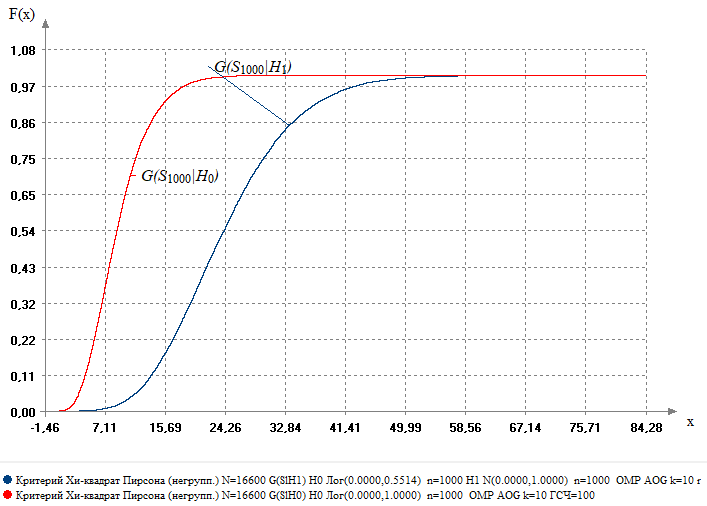
АОГ, k = 10, n = 50



АОГ, k = 10, n = 100:

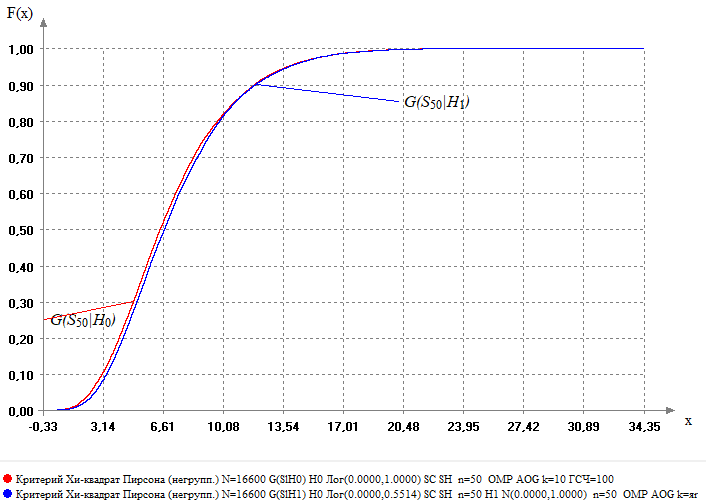


АОГ, k = 10, n = 1000:

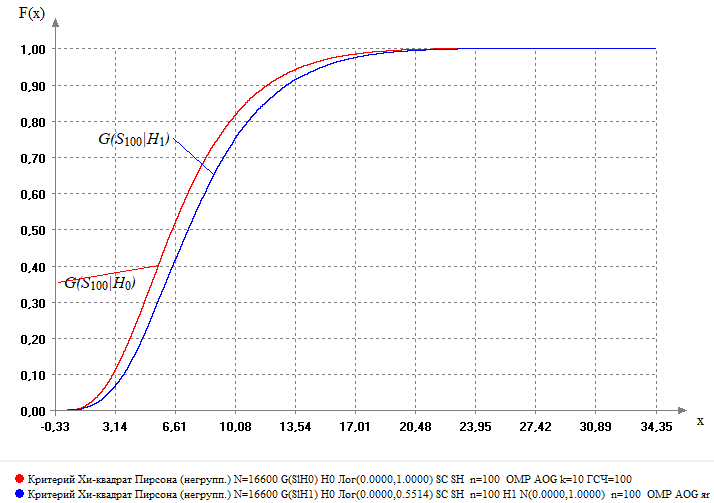


**Сложная гипотеза:**

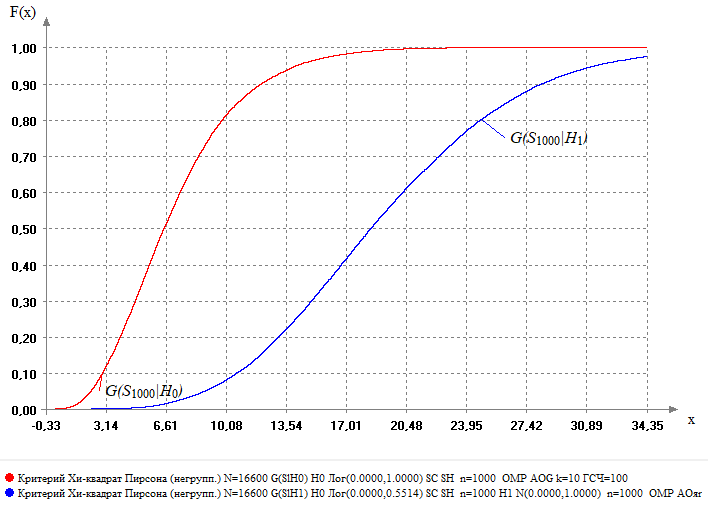
АОГ, k = 10, n = 50:



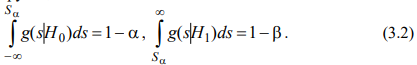
АОГ, k = 10, n = 100:



АОГ, k = 10, n = 1000:



4. В результате такого моделирования будут получены пары эмпирических распределений и для простой и сложной гипотез, по которым, задаваясь значением , можно вычислить мощность критерия относительно конкурирующей гипотезы:



Опираясь на эти распределения, построить оперативные характеристики критерия для проверки простой и сложной гипотез как функции вида .

**Простая гипотеза, АОГ, k = 10:**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | | |
|  |  |  |
| 0.15 | 0.199639 | 0.248614 | 0.902229 |
| 0.1 | 0.139699 | 0.17994 | 0.857169 |
| 0.05 | 0.073253 | 0.101386 | 0.772831 |
| 0.025 | 0.0372892 | 0.0570482 | 0.685241 |
| 0.01 | 0.0155422 | 0.025 | 0.560964 |

**Сложная гипотеза, АОГ, k = 10:**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | | |
|  |  |  |
| 0.15 | 0,153373 | 0,207892 | 0,896024 |
| 0.1 | 0,101988 | 0,141566 | 0,845361 |
| 0.05 | 0,0512048 | 0,0759036 | 0,743133 |
| 0.025 | 0,0246988 | 0,0392169 | 0,63241 |
| 0.01 | 0,0109036 | 0,0157831 | 0,476928 |

5. Повторяя пункты 3-4, исследовать влияние количества интервалов, способа группирования (равновероятное и асимптотически оптимальное) на мощность критерия.

**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 |

**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 |

**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 | 0.186084 | 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 |

**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 |

**Вывод:**

Наиболее подходящими распределениями для простых и сложных гипотез оказались распределения Гамма(), Su – Джонсона, Sl – Джонсона.

Выяснилось, что при небольших объёмах выборок, распределения для верных гипотез и достаточно близки друг к другу, так как их мощности минимальны и входят в пределы 0,15 — 0,2. Если начать увеличивать выборку, распределение, соответствующее верной гипотезе, будет постепенно смещаться вправо, а мощность начнёт увеличиваться. Данная ситуация относится как к сложной, так и к простой гипотезе.

При рассмотрении разных методов группирования выборок выяснилось, что преимушество в мощности АОГ проявляется только относительно БЛИЗКИХ конкурирующих гипотез (законов).