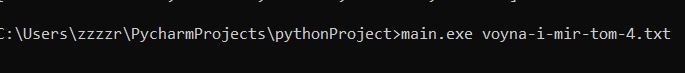
ЗАДАние на тему разработки программного средства поиска авторского инварианта

*Описание программного средства*

Программное средство представлено в виде .exe файла, принимающего на вход текстовый файл формата .txt. Попытка предоставить файл другого расширения, объект, файлом не являющийся или же вовсе отсутствие необходимого аргумента в виде названия файла приведут к досрочному завершению программы (в её коде они были обработаны)

Приложение запускается через интерфейс командной строки. Выглядит это следующим образом.



Далее программа проводит необходимые вычисления и выводит результат в файл result.txt, он создается автоматически. В файле находится сам авторский инвариант для всего текста и полученная выборка в ходе его расчета.

Общий инвариант считался как среднее между промежуточными в выборках. Инварианты выборок (то есть сами значения выборки) считались как процентное содержание служебных слов в выборке из 16 000 слов. Шагом до следующей выборки служит 5% от количества слов в тексте (такая вольная трактовка была принята в связи с указанием Фоменко и Фоменко о том, что шаг не столь важен, как размер выборки и не сильно влияет на итоговое значение).

Далее приведем код программы:

from typing import List, Any  
import re  
import statistics  
import argparse  
import sys  
  
  
def argument\_parsing(arg):  
 parser = argparse.ArgumentParser()  
 parser.add\_argument('name', help='Program name', type=str)  
 parser.add\_argument('file', help='Filename that we be used in calculating', type=str)  
  
 return parser.parse\_intermixed\_args(arg)  
  
  
def authors\_invariant(filename):  
 text = ""  
 try:  
 if '.' not in filename:  
 print("Not a filename!")  
 except TypeError:  
 print('Not a file name, terminating..')  
 return 3  
  
 filename\_dup = filename.split(".")  
  
 if filename\_dup[1] != 'txt':  
 print('Please, enter .txt file and nothing else!')  
 return 1  
  
 try:  
 with open(filename, encoding="windows-1251") as file:  
 text = file.read()  
 except:  
 print("Can not read this file or it's code not in windows-1251!\nPlease, use the following requirements.")  
 return 2  
  
 list\_of\_pieces = []  
  
 duty\_symbols = ['в', 'на','с','за','к','по','из','у','от','для',  
 'во','без','до','о','через','со','при','про','об','ко','над', 'из-за', 'из-под','под','и',  
 'что','но','а','да','хотя','когда','чтобы','если','тоже','или','то есть','зато','будто','не',  
 'как','же','даже','бы','ли','только','вот','то','ни','лишь','ведь','вон','нибудь','уже','либо']  
  
 text = re.split('[^а-яё-]+', text, flags=re.IGNORECASE)  
 text = [elem.lower() for elem in text if elem != '-']  
  
 step = 0  
 step\_t = 0.05 \* len(text)  
 count\_t = 0  
 temp\_list = []  
  
 for word in text:  
  
 if step != 0 and step < step\_t:  
 step += 1  
 continue  
  
 count\_t += 1  
 temp\_list.append(word)  
 if count\_t == 16000:  
 count\_t = 0  
 step = 1  
 list\_of\_pieces.append(temp\_list)  
 temp\_list = []  
  
 list\_of\_invariants = []  
  
 for piece in list\_of\_pieces:  
 list\_size = len(piece)  
 duty\_words\_amount = 0  
 for word in piece:  
 if word in duty\_symbols:  
 duty\_words\_amount += 1  
 list\_of\_invariants.append((duty\_words\_amount/list\_size)\*100)  
  
 print(list\_of\_invariants)  
 return list\_of\_invariants  
  
  
if \_\_name\_\_ == '\_\_main\_\_':  
 args = argument\_parsing(sys.argv)  
 with open('result.txt', 'w+', encoding='windows-1251') as file:  
 file.write("For this book " + args.file + " authors invariant is ")  
 values = authors\_invariant(args.file)  
 file.write(str(statistics.mean(values)))  
 file.write('\n')  
 file.write('And those values we got:')  
 file.write('\n')  
 for value in values:  
 file.write(str(value))  
 file.write('\n')

Нахождение инварианта и сравнение его для двух разных авторов

Далее приведем результаты тестирования нескольких различных авторов.

Для первого из них, И.С.Тургенева,были взяты произведения «Дворянское гнездо», «Отцы и дети», «Записки охотника» и «Вешние воды». Значения инвариантов для каждого из произведений соответственно получились следующими:

«Дворянское гнездо»: 23.003;

«Отцы и дети»: 23.16;

«Записки охотника» :22.9075;

«Вешние воды»: 23.103125;

**Среднее значение** =23,043, отклонение по формуле Фоменко и Фоменко = (Макс.-Мин.)/Среднее = 0.01096

Для второго, Л.Н.Толстого, были взяты 3 тома романа – эпопеи «Война и Мир», 2,3 и 4 тома, а также роман «Анна Каренина». Были получены следующие значения:

«Война и Мир» 2 том: 23.98875

«Война и Мир» 3 том: 24.02625

«Война и Мир» 4 том: 24.3075

«Анна Каренина»: 24.4111

**Среднее значение** = 24.1834, отклонение по формуле Фоменко и Фоменко = (Макс.-Мин.)/Среднее = 0.01752

Далее рассмотрим романы Федора Михайловича Достоевского «Бесы», «Братья Карамазовы», «Преступление и наказание», «Идиот». Для них получились следующие значения:

«Бесы»: 25.474107142857143

«Братья Карамазовы»: 26.045

«Преступление и наказание»: 25.9

«Идиот»: 26.075

**Среднее значение** = 25.8735, отклонение по формуле Фоменко и Фоменко = (Макс.-Мин.)/Среднее = 0.0232

Затем рассмотрим произведения Алексея Николаевича Толстого «Петр 1», «Хмурое утро», «Сестра», «Восемнадцатый год». Для них получились следующие значения:

«Петр 1»: 21.29609375

«Хмурое утро»: 21.85125

«Сестры»: 21.707812500000003

«Восемнадцатый год»: 21.203125

**Среднее значение** = 21.5145, отклонение по формуле Фоменко и Фоменко = (Макс.-Мин.)/Среднее = 0.03012

Следом проанализируем произведения Максима Горького «Детство», «Мать», «Сказки об Италии», «Дело Артамоновых». Для них получились следующие значения:

«Детство»: 20.954166666666666

«Мать»: 20.040625000000002

«Сказки об Италии»: 20.68125

«Дело Артамоновых»: 20.777083333333334

**Среднее значение** = 20.6132, отклонение по формуле Фоменко и Фоменко = (Макс.-Мин.)/Среднее = 0.0444

После этого проанализируем произведения Николая Семеновича Лескова «На ножах», «Воительница», «Соборяне». Для них получились следующие значения:

«На ножах»: 24.93125

«Воительница»: 24.9125

«Соборяне»: 25.071875

**Среднее значение** = 24.9718, отклонение по формуле Фоменко и Фоменко = (Макс.-Мин.)/Среднее = 0.0063

Следом проанализируем произведения Ивана Александровича Гончарова «Обломов», «Обыкновенная история», «Обрыв». Для них получились следующие значения:

«Обломов»: 24.977083333333333

«Обыкновенная история»: 24.778125

«Обрыв»: 25.015625

24.977083333333333 24.778125 25.015625

**Среднее значение** = 24.9236, отклонение по формуле Фоменко и Фоменко = (Макс.-Мин.)/Среднее = 0.0095

После чего рассмотрим произведения Ивана Сергеевича Шмелева «История любовная», «Солдаты». Для них получились следующие значения:

«История любовная»: 22.94375

«Солдаты»: 22.847916666666666

**Среднее значение** = 22.8960, отклонение по формуле Фоменко и Фоменко = (Макс.-Мин.)/Среднее = 0.0041

***Итоговая таблица*** выглядит следующим образом:

|  |  |  |
| --- | --- | --- |
| Автор | Авторский инвариант | Отклонение |
| И.С. Тургенев | 23,043 | 0.01096 |
| Л.Н. Толстой | 24.1834 | 0.01752 |
| А.Н. Толстой | 21.5145 | 0.03012 |
| И.А. Гончаров | 24.9236 | 0.0095 |
| Н.С. Лесков | 24.9718 | 0.0063 |
| И.С. Шмелев | 22.8960 | 0.0041 |
| М. Горький | 20.6132 | 0.0444 |
| Ф.М. Достоевский | 25.8735 | 0.0232 |

Таким образом, колебание данного параметра (содержание служебных слов) среди авторов существенно выше, чем отклонение в рамках одного автора. Именно за счет этого Фоменко и Фоменко сделали вывод о том, что данный параметр возможно принять за авторский инвариант. Мы также можем наблюдать тот факт, что авторские инварианты с учетом отклонения ни у одного автора не пересекаются с никаким другим.

Проверка статистических гипотез

Для непосредственной проверки статистических гипотез разработана ещё две программы - comparator.exe и hypothesis\_check.exe. Первая из них, получая на вход два файла, находит выборки для нахождения инвариантов для каждого из двух текстов, а затем по ним проверяет встроенными критериями равенства средних значений, дисперсий, равенство распределений и т.д. Однако выборки каждого конкретного файла могут быть очень малы (так у И.С.Тургенева это может быть всего 2 элемента в выборке, у Л.Н.Толстого – 5 элементов и т.д.).

На этом фоне некоторые критерии дают некорректные результаты. Поэтому была разработана ещё одна программа – hypothesis\_check.exe, которая проводит подсчеты для 4 файлов. Ей на вход подаются либо четыре файла одного автора, либо по два произведения различных авторов, но обязательно идущих подряд (например, два произведения И.С.Тургенева, а за ними – два произведения Л.Н.Толстого или наоборот). Далее все проверки будут осуществлены с помощью данного метода выбора двух произведений от каждого автора, а в случае одного автора – будет производится выбор четырех его произведений (где это будет возможным)

Прежде чем приступить к проверке гипотез и применению критериев, рассмотрим распределение статистик и коротко расскажем об основных применяемых далее критериев.

Краткие теоретические сведения об основных критериях и распределениях соответствующих им статистик

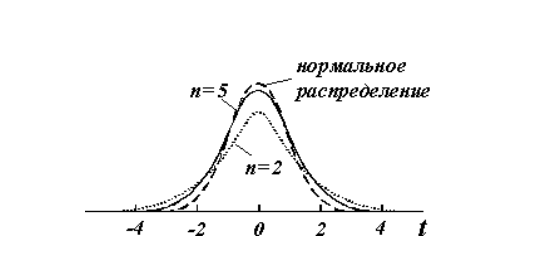
***Тест Стьюдента***

Для сравнения средних величин t-критерий Стьюдента рассчитывается по следующей формуле:

https://medstatistic.ru/pictures/formula_student.png

где **М1** - средняя арифметическая первой сравниваемой совокупности (группы), **М2** - средняя арифметическая второй сравниваемой совокупности (группы), **m1** - средняя ошибка первой средней арифметической, **m2** - средняя ошибка второй средней арифметической.

Таким образом, величина t измеряется отклонением выборочной средней от средней генеральной совокупности, выраженным в долях ошибки 5 выборки, принятой за единицу. Теоретическое распределение критерия Стьюдента дается в статистической таблице. Графически распределение Стьюдента при n=2 и n=5 в сравнении с нормальным распределением имеет следующий вид:



Максимум частоты нормального и t распределения совпадают, но форма кривой t распределения полностью зависит от числа наблюдений. При очень малом количестве наблюдений она принимает вид плосковершинной кривой, причем площадь, ограниченная кривой, больше, чем при нормальном распределении. При увеличении числа наблюдений распределение t приближается к нормальному и переходит в него при n ->∞.

*Нулевая гипотеза = <Средние значения равны>*

*Первая гипотеза (всеобщая альтернатива) = <Средние значения не равны>*

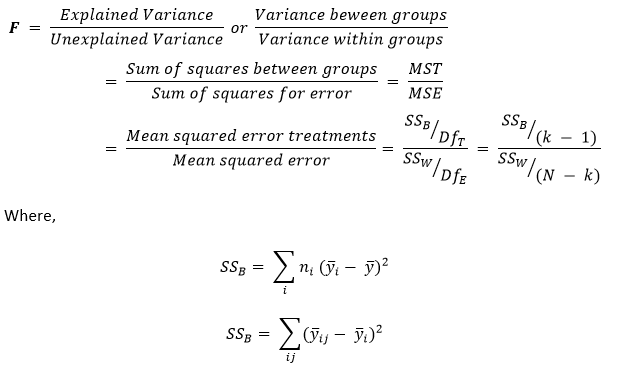
***Тест ANOVA на различие средних значений***

Дисперсионный анализ был введён Фишером - английским учёным, сделавшим огромный вклад в развитие науки

ANOVA — это акроним от ANalysis Of VAriance (дисперсионный анализ).

Анализ основан на расчете ***https://studfile.net/html/2706/304/html_63g4IzNBq4.hiYJ/img-yo0P71.png-статистики*** (статистика Фишера), которая представляет собой отношение двух *дисперсий*: межгрупповой и внутригрупповой. ***https://studfile.net/html/2706/304/html_63g4IzNBq4.hiYJ/img-wSo44Y.png-***тест в однофакторном дисперсионном анализе устанавливает, значимо ли отличаются средние нескольких независимых выборок.

Мы можем выполнить тест ANOVA, сравнив два типа вариаций. Первый вариант – между средними значениями выборки и вторым в каждой из выборок. Приведенная ниже формула описывает статистику одностороннего теста ANOVA.



Где:

yi – среднее значение выборки в i-й группе,

ni – Количество наблюдений в i-й группе,

y – Общее среднее значение данных,

k – Общее количество групп,

yij – j-е наблюдение в группе из k,

N – Общий размер выборки.

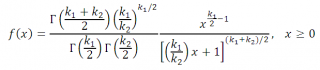
Распределение Фишера зависит только от числа степеней свободы 2 выборок, т.е. от объема выборок. Когда 2 сравниваемые выборки являются случайными независимыми из общей совокупности с генеральной средней μ , то фактическое значение критерия Фишера не выйдет за определенные пределы и не превысит критическое для данных значений степеней свободы теоретическое значение критерия Фишера, т.е. Fфакт< Fтеор. Если генеральные параметры сравниваемых групп различны, то Fфакт≥ Fтеор.

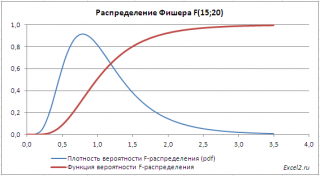
*Нулевая гипотеза:* заметных различий между группами нет.

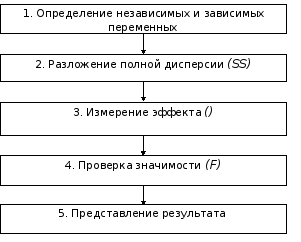
*Альтернативная гипотеза (всеобщая альтернатива):* между группами есть заметное различие.

Различие здесь подразумевает под собой различие средних значений.

Плотность F -распределения выражается формулой:







***Тест Манна – Уитни***

U-критерий Манна-Уитни является непараметрическим критерием, поэтому, в отличие от **t-критерия Стьюдента**, не требует наличия нормального распределения сравниваемых совокупностей.

U-критерий подходит для сравнения малых выборок: в каждой из выборок должно быть не менее 3 значений признака. Допускается, чтобы в одной выборке было 2 значения, но во второй тогда должно быть не менее пяти.

Условием для применения U-критерия Манна-Уитни является отсутствие в сравниваемых группах совпадающих значений признака (все числа – разные) или очень малое число таких совпадений.

Аналогом U-критерия Манна-Уитни для сравнения трех и более групп является **Критерий Краскела-Уоллиса**.

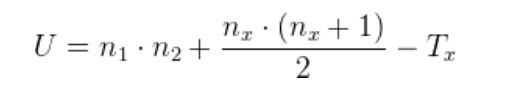
В составленном едином ранжированном ряду общее количество рангов получится равным:

N = n1 + n2

где n1 - количество элементов в первой выборке, а n2 - количество элементов во второй выборке.

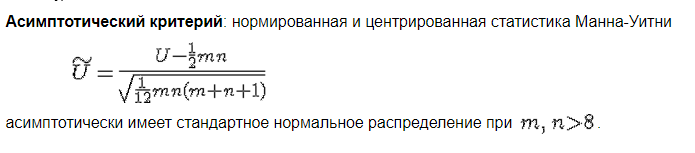
Далее вновь разделяем единый ранжированный ряд на два, состоящие соответственно из единиц первой и второй выборок, запоминая при этом значения рангов для каждой единицы. Подсчитываем отдельно сумму рангов, пришедшихся на долю элементов первой выборки, и отдельно - на долю элементов второй выборки. Определяем большую из двух ранговых сумм (Tx) соответствующую выборке с nx элементами.

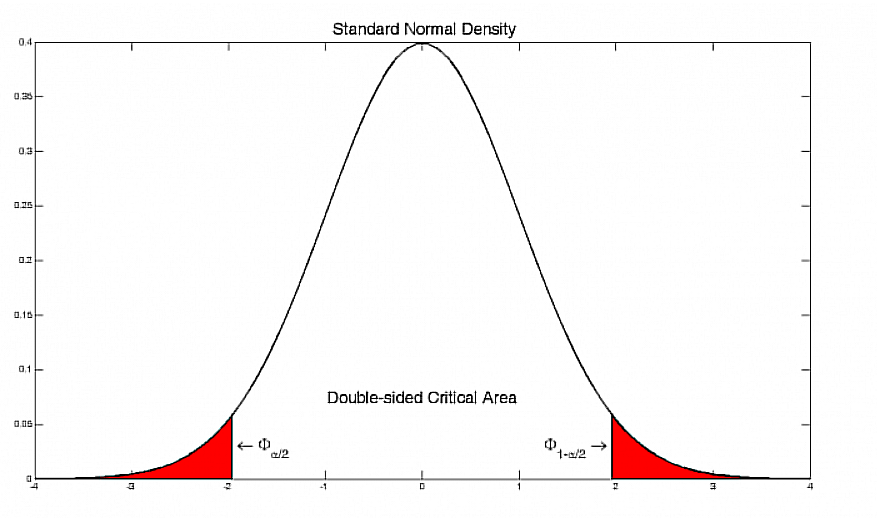
Наконец, находим значение U-критерия Манна-Уитни по формуле:



Полученное значение U-критерия сравниваем по таблице для избранного уровня статистической значимости с критическим значением U при заданной численности сопоставляемых выборок:

* Если полученное значение U **меньше**табличного или **равно**ему, то признается статистическая значимость различий между уровнями признака в рассматриваемых выборках (принимается альтернативная гипотеза). Достоверность различий тем выше, чем меньше значение U.
* Если же полученное значение U **больше**табличного, принимается нулевая гипотеза.





Критическая область асимптотического критерия Манна-Уитни.

Таким образом, Нулевая гипотеза = <*Принимаем гипотезу о равенстве распределений*>

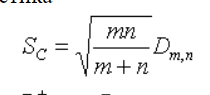
Альтернатива (всеобщая альтернатива) = <Отклоняем гипотезу о равенстве распределений>

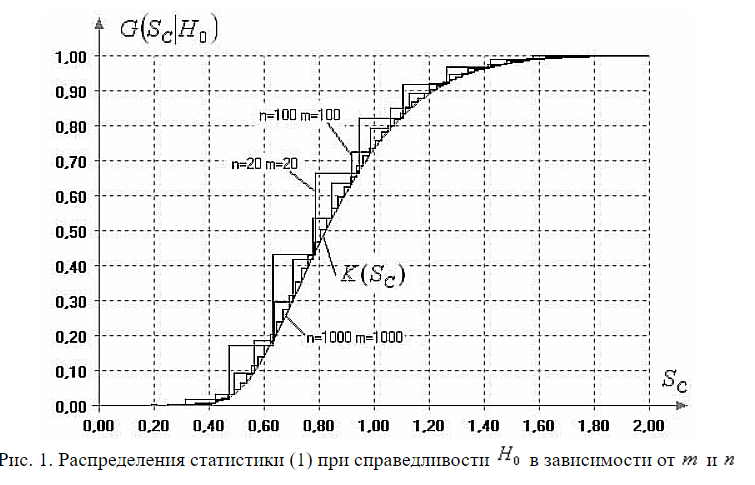
***Критерий Колмогорова – Смирнова***

Статистика критерия Смирнова измеряет различие ме­жду эмпирическими функциями распределения, построенными по выборкам.

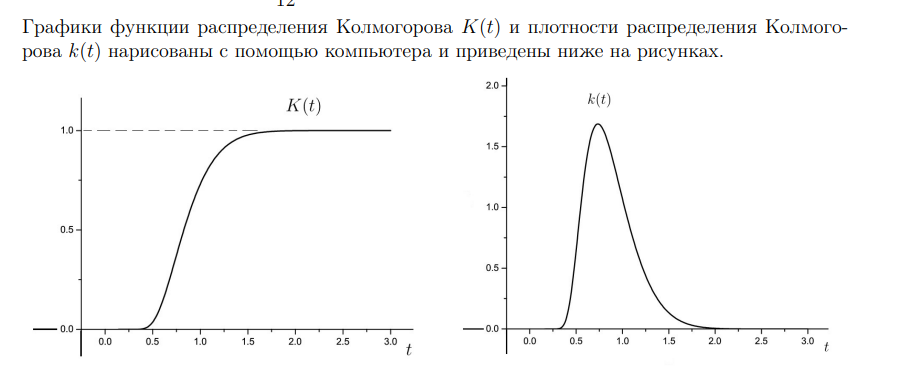
https://ami.nstu.ru/~headrd/seminar/publik_html/Izm_T_8.files/image011.gif

Если гипотеза https://ami.nstu.ru/~headrd/seminar/publik_html/Izm_T_8.files/image006.gif справедлива, то при неограниченном увеличении объемов выборок статистика

 в пределе подчиняется распределению Колмогорова.



На рис. 1 показаны условные распределения статистики (1) при спра­ведливости https://ami.nstu.ru/~headrd/seminar/publik_html/Izm_T_8.files/image006.gif в зависимости от https://ami.nstu.ru/~headrd/seminar/publik_html/Izm_T_8.files/image001.gif и https://ami.nstu.ru/~headrd/seminar/publik_html/Izm_T_8.files/image002.gif (при https://ami.nstu.ru/~headrd/seminar/publik_html/Izm_T_8.files/image001.gif=https://ami.nstu.ru/~headrd/seminar/publik_html/Izm_T_8.files/image002.gif). Как следует из получен­ной кар­тины, даже при https://ami.nstu.ru/~headrd/seminar/publik_html/Izm_T_8.files/image025.gif и https://ami.nstu.ru/~headrd/seminar/publik_html/Izm_T_8.files/image026.gif ступенчатость https://ami.nstu.ru/~headrd/seminar/publik_html/Izm_T_8.files/image023.gif со­храня­ется. Другим недостатком применения критерия со статистикой (1) является то (см. рис. 1), что рас­пределения https://ami.nstu.ru/~headrd/seminar/publik_html/Izm_T_8.files/image023.gif с ростом https://ami.nstu.ru/~headrd/seminar/publik_html/Izm_T_8.files/image001.gif и https://ami.nstu.ru/~headrd/seminar/publik_html/Izm_T_8.files/image002.gif приближаются к предельному распреде­лению https://ami.nstu.ru/~headrd/seminar/publik_html/Izm_T_8.files/image018.gif слева.



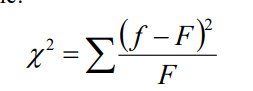
*Нулевая гипотеза = <Принимаем гипотезу о равенстве распределений>,*

*Всеобщая альтернатива = <Отклоняем гипотезу о равенстве распределений>*

***Критерий Хи – квадрат***

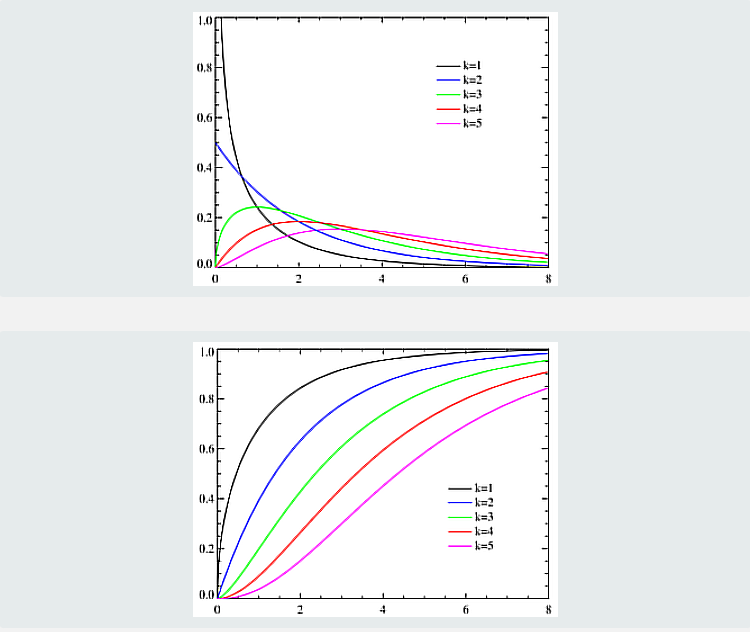
χ 2 – распределение Пирсона используется для сравнения фактических и теоретических частот при качественной изменчивости при решении селекционно-генетических задач.

Для сравнения частот рассчитывают χ 2 – критерий, или критерий согласия по следующей формуле:



где, f – частота фактическая, F – теоретическая частота. Фактическое значение критерия сравнивают с теоретическим, которое берут из статистических таблиц. Если χ 2 факт. ≥ χ 2 теор., различия между частотами существенны, если χ 2факт< χ 2 теор, различия между частотами несущественны.

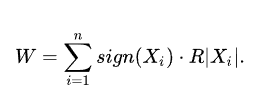
Кривая распределения Пирсона в значительной степени зависит от числа степеней свободы, т.е. как и распределение Фишера, от объема выборки. Для малого числа степеней свободы кривая асимметрична, но с увеличением числа степеней свободы при n ->∞ кривая становится нормальной.

Первая картинка — это плотность распределения (вероятность получить в выборке каждое из чисел на горизонтальной оси), вторая — интегральная функция распределения (вероятность получить значение меньше, чем на горизонтальной оси).

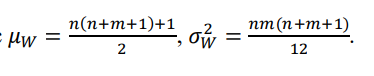
***Тест Уилкоксона***

Критерий Уилкоксона для связанных выборок (также используются названия Т-критерий Уилкоксона, критерий Вилкоксона, критерий знаковых рангов Уилкоксона, критерий суммы рангов Уилкоксона) – непараметрический статистический критерий, используемый для сравнения двух связанных (парных) выборок по уровню какого-либо количественного признака, измеренного в непрерывной или в порядковой шкале.

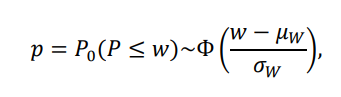
Более подробно формулу можно выразить следующей записью. Выбранная нами статистика теста Уилкоксона (signed-rank Wilcoxon test) хороша тем, что использует лишь ранги этих расстояний. Обозначим ранг , упорядоченных от меньшего значения к большему. Тогда статистика Уилкоксона имеет вид:



Известно, что статистика Уилкоксона асимптотически нормальна,



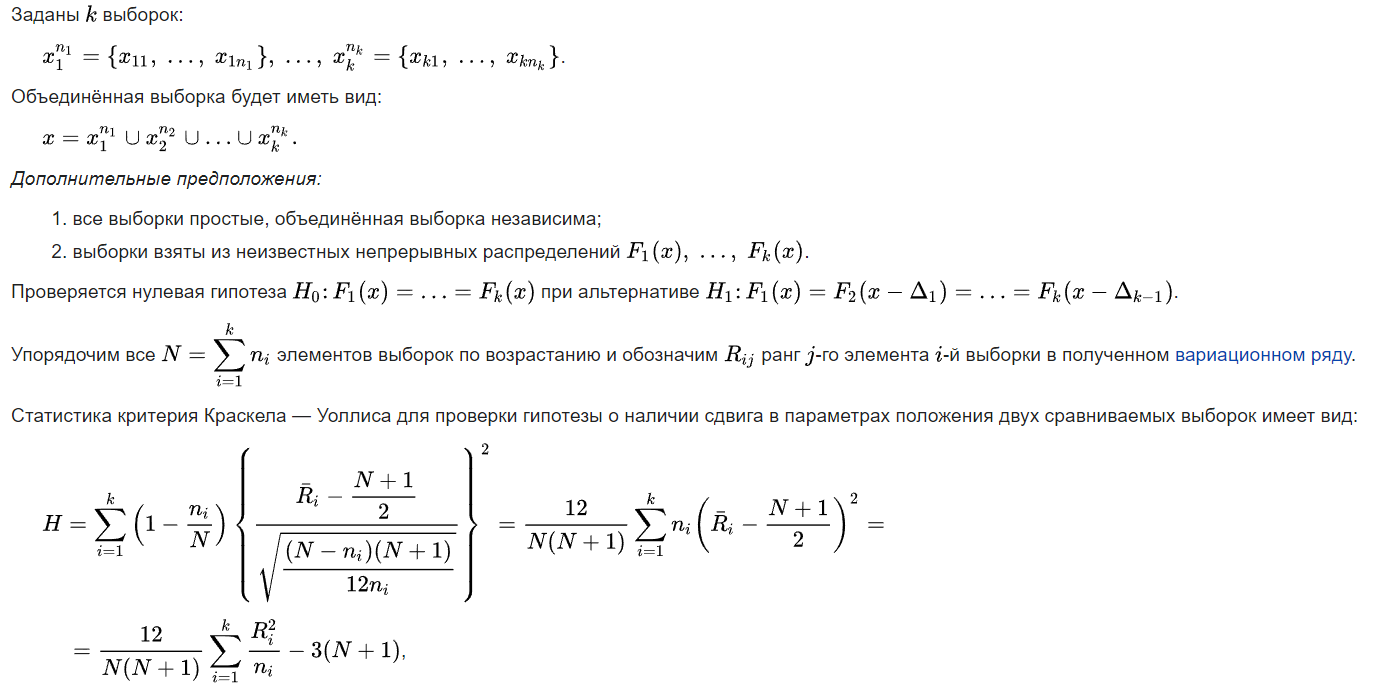
После вычисления статистики Вилкоксона для проверки гипотезы однородности определяется p-значение:

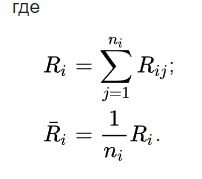


Таким образом, *Нулевая гипотеза = <Средние значения равны>, а её Альтернатива(всеобщая) = <Средние значения не равны>.*

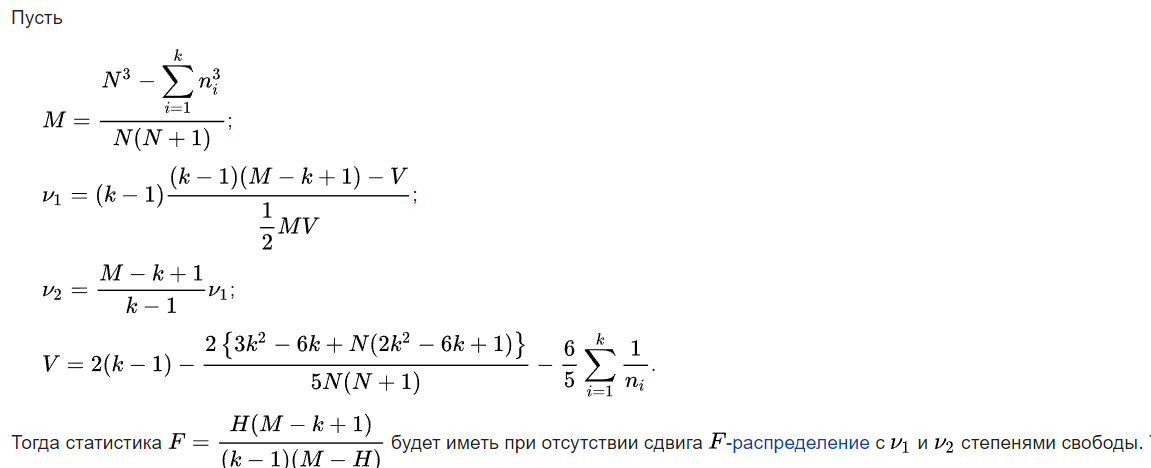
***Критерий Краскелла – Уоллиса***

Критерий Краскела — Уоллиса предназначен для проверки равенства медиан нескольких выборок. Данный критерий является многомерным обобщением критерия Уилкоксона — Манна — Уитни. Критерий Краскела — Уоллиса является [ранговым](https://ru.wikipedia.org/w/index.php?title=%D0%A0%D0%B0%D0%BD%D0%B3%D0%BE%D0%B2%D1%8B%D0%B9_%D0%BA%D1%80%D0%B8%D1%82%D0%B5%D1%80%D0%B8%D0%B9&action=edit&redlink=1), поэтому он инвариантен по отношению к любому [монотонному преобразованию](https://ru.wikipedia.org/w/index.php?title=%D0%9C%D0%BE%D0%BD%D0%BE%D1%82%D0%BE%D0%BD%D0%BD%D0%BE%D0%B5_%D0%BF%D1%80%D0%B5%D0%BE%D0%B1%D1%80%D0%B0%D0%B7%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5&action=edit&redlink=1) [шкалы](https://ru.wikipedia.org/wiki/%D0%A8%D0%BA%D0%B0%D0%BB%D0%B0) измерения.

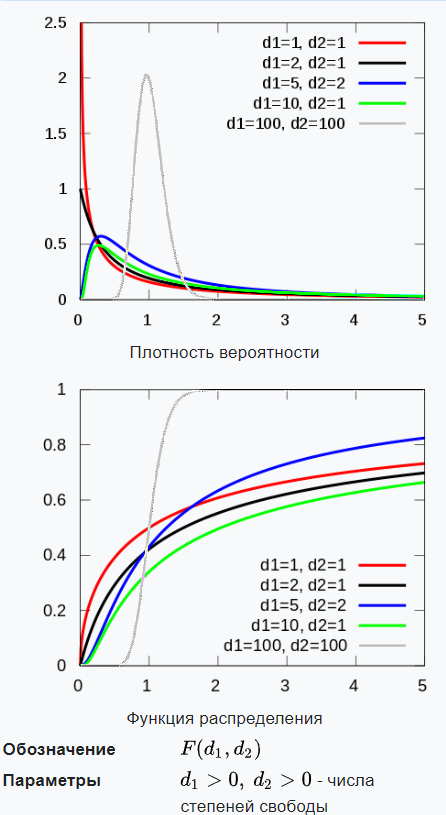




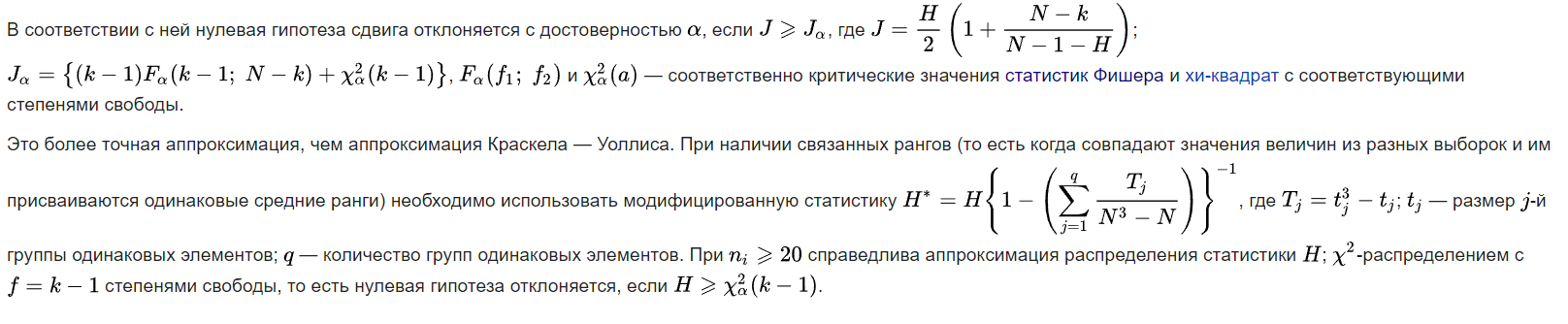
Аппроксимация Краскелла - Уоллиса



Где F-распределение – уже упомянутое выше распределение Фишера.



Аппроксимация Имана – Давенпорта

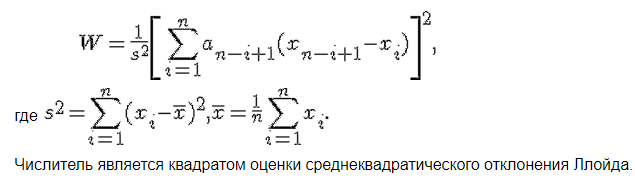


Таким образом, *Нулевая гипотеза = <Средние значения выборок равны>, её Всеобщая альтернатива = <Средние значения выборок различны>.* Вывод делается также на основе сравнения уровня значимости с p-value.

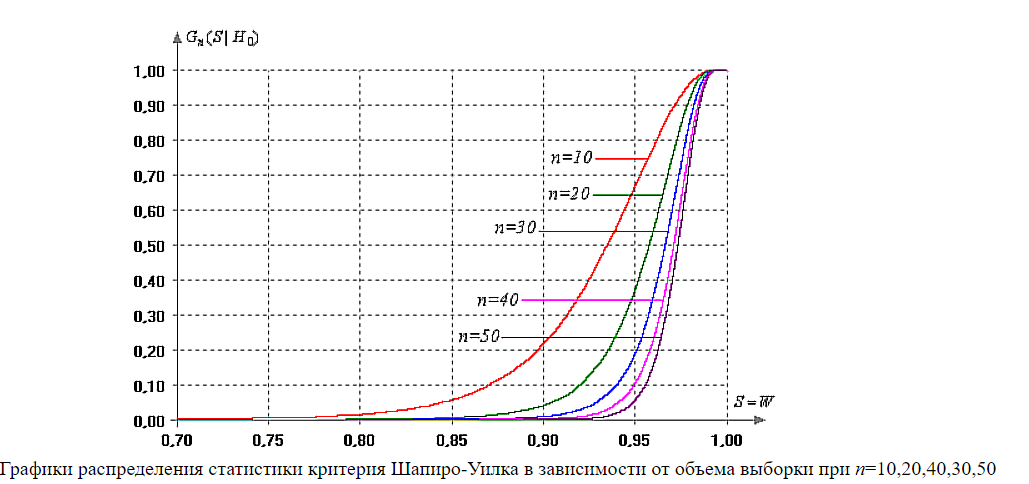
***Тест Шапиро – Уилка***

Критерий Шапиро-Уилка используется для проверки гипотезы H_0: «случайная величина X распределена нормально» и является одним наиболее эффективных критериев проверки нормальности.

Критерий Шапиро-Уилка основан на оптимальной линейной несмещённой оценке дисперсии к её обычной оценке методом максимального правдоподобия. Статистика критерия имеет вид:







Таким образом, *Нулевая гипотеза = <Случайная величина имеет нормальное распределение>, ее всеобщая альтернатива = <Случайная величина имеет распределение, отличное от нормального>.*

Проверка выборок указанными статистическими тестами/критериями

Убедимся в том, что полученные данные на предыдущем шаге пройдут проверку статистическими критериями. Для этого рассмотрим произведения перечисленных выше в таблице авторов. Далее представим в сравнении эти произведения при различных уровнях значимости. Для разных авторов выбрались следующие произведения:

Тургенев – «Отцы и дети», «Записки охотника»

Гончаров – «Обломов», «Обрыв»

Шмелев – «История любовная», «Солдаты»

А.Н. Толстой – «Петр 1», «Хмурое утро»

Л.Н. Толстой – «Анна Каренина», «Война и мир – 2 том»

Горький – «Мать», «Детство»

Лесков – «На ножах», «Соборяне»

Достоевский – «Бесы», «Идиот»

Для работ одного автора брались все его представленные выше произведения.

Данные о статистиках и гипотезах

|  |  |  |  |
| --- | --- | --- | --- |
| Критерий | Проверяемая  сущность | Статистика | Ограничение |
| Критерий Шапиро – Уилка | H0 = <Нормальность распределения>  H1 = <Распределение не является нормальным> | 2 , где an-i+1 , берутся из таблиц | -) Не подходит для выборок большого объёма (в этом случае использовать Андерсон – Дарлинг или Шапиро – Франчиа) |
| ANOVA | H0 = <Выборочные средние равны>  H1 = <Выборочные средние различны> | , где  = 2  = 2 | -) Три и более выборки (не имеет смысла) |
| Критерий Краскелла – Уоллиса | H0 = <Выборочные средние равны>  H1 = <Выборочные средние различны> |  | -) все k выборок простые, объединённая выборка независима  -) выборки взяты из неизвестных непрерывных распределений |
| T-критерий Стьюдента | H0 = <Выборочные средние равны>  H1 = <Выборочные средние различны> |  | -) Выборки простые\* и нормально распределены  -) Дисперсии неизвестны, но равны |
| U-критерий Манна – Уитни | H0 = <Распределения равны>  H1 = <Распределения различны> | Rx =  Ry =  U­x = mn + 1/2m(m+1) - Rx  U­y = mn + 1/2n(n+1) – Ry  U = min{Ux, Uy} | -) все выборки простые, объединённая выборка независима  -) выборки взяты из неизвестных непрерывных распределений |
| W-ранговый критерий Уилкоксона | H0 = <Выборочные средние равны>  H1 = <Выборочные средние различны> | j, где  r1, …, rn  - ранги всех значений выборки | -) Oбъём выборки до 25 элементов  -) Выборки одинакового объёма |
| Критерий Колмогорова - Смирнова | H0 = <Распределения равны>  H1 = <Распределения различны> |  | -) выборки взяты из неизвестных непрерывных распределений |

***Уровень значимости = 0.1***.

*Произведения одного автора:*

**Л.Н. Толстой**

**Выборка:**

[23.99375, 23.21875, 23.53125, 24.39375, 24.80625]

[24.075, 23.7375, 24.19375, 23.35625, 24.76875]

[24.456249999999997, 24.31875, 23.70625, 24.631249999999998, 24.425]

[23.73125, 24.0125, 25.025, 24.64375, 24.025, 25.0125, 24.28125, 24.625, 24.34375]

Статистика P-value

|  |  |  |  |
| --- | --- | --- | --- |
| Тест Шапиро – первая и вторая выборки | 0.9531287550926208 | 0.7055749297142029 | Принимаем гипотезу о нормальности распределения |
| Тест Шапиро – третья и четвертая выборки | 0.9526628255844116 | 0.6027610301971436 | Принимаем гипотезу о нормальности распределения |
| ANOVA | 1.1227716916854662 | 0.36350711249099216 | Принимаем гипотезу о равенстве выборочных средних |
| Краскелла – Уоллиса | 2.5688888888888926 | 0.4629698495183314 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Стьюдента | 0.445019610874379 | 0.6606533322957311 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Уилкоксона | 16.0 | 0.275390625 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Манна – Уитни | 43.0 | 0.12073854334140409 | Принимаем гипотезу о равенстве распределений |
| Критерий Колмогорова – Смирнова | 0.4142857142857143 | 0.20557234751608155 | Принимаем гипотезу о равенстве распределений |
| Двухвыборочный критерий Колмогорова – Смирнова | 0.4142857142857143 | 0.20557234751608155 | Принимаем гипотезу о равенстве распределений |

**И.С.Тургенев**

[23.1875, 22.88125, 21.96875, 23.15625, 23.34375]

[22.7375, 23.46875]

[22.825, 23.443749999999998, 23.2375]

[22.93125, 23.075000000000003]

Статистика P-value

|  |  |  |  |
| --- | --- | --- | --- |
| Тест Шапиро – первая и вторая выборки | 0.8792604207992554 | 0.22318527102470398 | Принимаем гипотезу о нормальности распределения |
| Тест Шапиро – третья и четвертая выборки | 0.9735967516899109 | 0.8977965712547302 | Отклонить гипотезу о нормальности распределения |
| ANOVA | 0.22743237873766656 | 0.8747554679766818 | Принимаем гипотезу о равенстве выборочных средних |
| Краскелла – Уоллиса | 0.6025641025641093 | 0.8958451305579371 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Стьюдента | -0.563677839021931 | 0.5853911903445019 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Уилкоксона | 5.0 | 0.625 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Манна – Уитни | 16.0 | 0.8762626262626263 | Принимаем гипотезу о равенстве распределений |
| Критерий Колмогорова – Смирнова | 0.2857142857142857 | 0.9090909090909092 | Принимаем гипотезу о равенстве распределений |
| Двухвыборочный критерий Колмогорова – Смирнова | 0.2857142857142857 | 0.9090909090909092 | Принимаем гипотезу о равенстве распределений |

**Ф.М.Достоевский**

[26.400000000000002, 25.3125, 25.34375, 24.73125, 25.1, 26.08125, 25.35]

[27.3375, 26.34375, 24.5125, 25.724999999999998, 26.30625, 25.48125, 25.59375]

[27.06875, 25.056250000000002, 26.03125, 25.15, 26.58125, 25.35, 25.98125, 27.38125]

[27.200000000000003, 26.656249999999996, 24.89375, 26.94375, 26.637499999999996, 24.55, 26.150000000000002, 25.31875, 25.937500000000004, 26.1625]

Статистика P-value

|  |  |  |  |
| --- | --- | --- | --- |
| Тест Шапиро – первая и вторая выборки | 0.961934208869931 | 0.7547894716262817 | Принимаем гипотезу о нормальности распределения |
| Тест Шапиро – третья и четвертая выборки | 0.9573202133178711 | 0.5508359670639038 | Принимаем гипотезу о нормальности распределения |
| ANOVA | 0.847071738535625 | 0.4798407730711082 | Принимаем гипотезу о равенстве выборочных средних |
| Краскелла – Уоллиса | 2.4870934267382383 | 0.4776281064692328 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Стьюдента | -1.287739851517813 | 0.2076819810451186 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Уилкоксона | 26.0 | 0.10400390625 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Манна – Уитни | 95.5 | 0.254407050027496 | Принимаем гипотезу о равенстве распределений |
| Критерий Колмогорова – Смирнова | 0.31746031746031744 | 0.33130906533151083 | Принимаем гипотезу о равенстве распределений |
| Двухвыборочный критерий Колмогорова – Смирнова | 0.31746031746031744 | 0.33130906533151083 | Принимаем гипотезу о равенстве распределений |

**А.Н.Толстой**

[20.96875, 21.35625, 22.075, 21.9, 20.66875, 20.225, 21.64375, 21.53125]

[21.5375, 22.125, 21.21875, 21.95625, 22.418750000000003]

[21.31875, 21.6, 21.41875, 20.474999999999998]

[22.31875, 22.44375, 20.94375, 21.125]

|  |  |  |  |
| --- | --- | --- | --- |
| Тест Шапиро – первая и вторая выборки | 0.9641405344009399 | 0.8159711360931396 | Принимаем гипотезу о нормальности распределения |
| Тест Шапиро – третья и четвертая выборки | 0.9501833915710449 | 0.7130816578865051 | Принимаем гипотезу о нормальности распределения |
| ANOVA | 1.326027339823393 | 0.2987284545590459 | Принимаем гипотезу о равенстве выборочных средних |
| Краскелла – Уоллиса | 3.2512987012987082 | 0.3544786796248943 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Стьюдента | 0.1886801512445907 | 0.8523442596368178 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Уилкоксона | 14.0 | 0.640625 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Манна – Уитни | 57.0 | 0.7501007420512065 | Принимаем гипотезу о равенстве распределений |
| Критерий Колмогорова – Смирнова | 0.2403846153846154 | 0.8667256376234705 | Принимаем гипотезу о равенстве распределений |
| Двухвыборочный критерий Колмогорова – Смирнова | 0.2403846153846154 | 0.8667256376234705 | Принимаем гипотезу о равенстве распределений |

*Произведения разных авторов*

**Л.Н.Толстой и М.Горький**

[23.99375, 23.21875, 23.53125, 24.39375, 24.80625]

[23.73125, 24.0125, 25.025, 24.64375, 24.025, 25.0125, 24.28125, 24.625, 24.34375]

[20.79375, 21.1375, 20.931250000000002]

[19.975, 20.556250000000002, 19.90625, 19.725]

Статистика P-value

|  |  |  |  |
| --- | --- | --- | --- |
| Тест Шапиро – первая и вторая выборки | 0.966253399848938 | 0.8229710459709167 | Принимаем гипотезу о нормальности распределения |
| Тест Шапиро – третья и четвертая выборки | 0.9084219336509705 | 0.38505473732948303 | Принимаем гипотезу о нормальности распределения |
| ANOVA | 107.45121923531421 | 2.9779348212704665e-11 | Отклоняем гипотезу о равенстве выборочных средних |
| Краскелла – Уоллиса | 14.638383838383831 | 0.002153263166318698 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Стьюдента | 13.057989127706307 | 1.3510089372555462e-09 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Уилкоксона | 0.0 | 0.015625 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Манна – Уитни | 70.0 | 0.00010283833813245579 | Отклоняем гипотезу о равенстве распределений |
| Критерий Колмогорова – Смирнова | 1.0 | 0.00010283833813245579 | Отклоняем гипотезу о равенстве распределений |
| Двухвыборочный критерий Колмогорова – Смирнова | 1.0 | 0.00010283833813245579 | Отклоняем гипотезу о равенстве распределений |

**И.С.Тургенев и И.А.Гончаров**

[22.825, 23.443749999999998, 23.2375]

[23.1875, 22.88125, 21.96875, 23.15625, 23.34375]

[24.825, 25.59375, 25.25, 24.56875, 24.525, 25.1]

[25.78125, 24.349999999999998, 25.33125, 25.724999999999998, 26.28125, 24.2, 24.2875, 24.16875]

Статистика P-value

|  |  |  |  |
| --- | --- | --- | --- |
| Тест Шапиро – первая и вторая выборки | 0.816076397895813 | 0.042428158223629 | Отклонить гипотезу о нормальности |
| Тест Шапиро – третья и четвертая выборки | 0.9319276213645935 | 0.3246665596961975 | Принимаем гипотезу о нормальности распределения |
| ANOVA | 16.465912344598067 | 2.122864935345004e-05 | Отклоняем гипотезу о равенстве выборочных средних |
| Краскелла – Уоллиса | 14.694664031620547 | 0.002097074807763693 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Стьюдента | -7.3189729735610 | 4.479526793943783e-07 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Уилкоксона | 0.0 | 0.0078125 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Манна – Уитни | 0.0 | 6.254495418582106e-06 | Отклоняем гипотезу о равенстве распределений |
| Критерий Колмогорова – Смирнова | 1.0 | 6.254495418582106e-06 | Отклоняем гипотезу о равенстве распределений |
| Двухвыборочный критерий Колмогорова – Смирнова | 1.0 | 6.254495418582106e-06 | Отклоняем гипотезу о равенстве распределений |

**Н.СЛесков и Ф.М.Достоевский**

[24.79375, 25.2625, 25.24375, 24.9875]

[24.825, 25.056250000000002, 25.118750000000002, 24.5375, 24.85625, 25.906249999999996, 25.156250000000004, 23.99375]

[26.400000000000002, 25.3125, 25.34375, 24.73125, 25.1, 26.08125, 25.35]

[27.06875, 25.056250000000002, 26.03125, 25.15, 26.58125, 25.35, 25.98125, 27.38125]

Статистика P-value

|  |  |  |  |
| --- | --- | --- | --- |
| Тест Шапиро – первая и вторая выборки | 0.9404017925262451 | 0.5032945275306702 | Принимаем гипотезу о нормальности распределения |
| Тест Шапиро – третья и четвертая выборки | 0.917828381061554 | 0.17847254872322083 | Принимаем гипотезу о нормальности распределения |
| ANOVA | 4.6631434644944925 | 0.010930897137135973 | Отклоняем гипотезу о равенстве выборочных средних |
| Краскелла – Уоллиса | 10.1143806178549 | 0.017618466755198722 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Стьюдента | -3.17660076113350 | 0.003935043204045219 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Уилкоксона | 3.0 | 0.06297905121445523 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Манна – Уитни | 28.5 | 0.002907004642823949 | Отклоняем гипотезу о равенстве распределений |
| Критерий Колмогорова – Смирнова | 0.65 | 0.003504169959951357 | Отклоняем гипотезу о равенстве распределений |
| Двухвыборочный критерий Колмогорова – Смирнова | 0.65 | 0.003504169959951357 | Отклоняем гипотезу о равенстве распределений |

**А.Н.Толстой и И.С.Шмелев**

[24.025, 22.4875, 22.03125]

[23.76875, 22.34375, 22.71875]

[20.96875, 21.35625, 22.075, 21.9, 20.66875, 20.225, 21.64375, 21.53125]

[21.5375, 22.125, 21.21875, 21.95625, 22.418750000000003]

Статистика P-value

|  |  |  |  |
| --- | --- | --- | --- |
| Тест Шапиро – первая и вторая выборки | 0.8815450072288513 | 0.27622759342193604 | Принимаем гипотезу о нормальности распределения |
| Тест Шапиро – третья и четвертая выборки | 0.9641405344009399 | 0.8159711360931396 | Принимаем гипотезу о нормальности распределения |
| ANOVA | 6.340743403392538 | 0.005444572371527678 | Отклоняем гипотезу о равенстве выборочных средних |
| Краскелла – Уоллиса | 11.06421052631579 | 0.011383683941335116 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Стьюдента | 4.109408419407771 | 0.0007316785091069349 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Уилкоксона | 1.0 | 0.0625 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Манна – Уитни | 74.0 | 0.0008845643520566121 | Отклоняем гипотезу о равенстве распределений |
| Критерий Колмогорова – Смирнова | 0.7692307692307693 | 0.0069290874244434635 | Отклоняем гипотезу о равенстве распределений |
| Двухвыборочный критерий Колмогорова – Смирнова | 0.7692307692307693 | 0.0069290874244434635 | Отклоняем гипотезу о равенстве распределений |

**У*ровень значимости = 0.05:***

*Произведения одного автора:*

**Л.Н. Толстой**

[23.99375, 23.21875, 23.53125, 24.39375, 24.80625]

[24.075, 23.7375, 24.19375, 23.35625, 24.76875]

[24.456249999999997, 24.31875, 23.70625, 24.631249999999998, 24.425]

[23.73125, 24.0125, 25.025, 24.64375, 24.025, 25.0125, 24.28125, 24.625, 24.34375]

Статистика P-value

|  |  |  |  |
| --- | --- | --- | --- |
| Тест Шапиро – первая и вторая выборки | 0.9531287550926208 | 0.7055749297142029 | Принимаем гипотезу о нормальности распределения |
| Тест Шапиро – третья и четвертая выборки | 0.9526628255844116 | 0.6027610301971436 | Принимаем гипотезу о нормальности распределения |
| ANOVA | 1.1227716916854662 | 0.36350711249099216 | Принимаем гипотезу о равенстве выборочных средних |
| Краскелла – Уоллиса | 2.5688888888888926 | 0.4629698495183314 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Стьюдента | 0.44501961087437913 | 0.6606533322957311 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Уилкоксона | 16.0 | 0.275390625 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Манна – Уитни | 43.0 | 0.12073854334140409 | Принимаем гипотезу о равенстве распределений |
| Критерий Колмогорова – Смирнова | 0.4142857142857143 | 0.20557234751608155 | Принимаем гипотезу о равенстве распределений |
| Двухвыборочный критерий Колмогорова – Смирнова | 0.4142857142857143 | 0.20557234751608155 | Принимаем гипотезу о равенстве распределений |

**И.С.Тургенев**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Тест Шапиро – первая и вторая выборки | 0.8792604207992554 | 0.22318527102470398 | | Принимаем гипотезу о нормальности распределения |
| Тест Шапиро – третья и четвертая выборки | 0.9735967516899109 | | 0.8977965712547302 | Отклонить гипотезу о нормальности распределения |
| ANOVA | 0.22743237873766656 | | 0.8747554679766818 | Принимаем гипотезу о равенстве выборочных средних |
| Краскелла – Уоллиса | 0.6025641025641093 | | 0.8958451305579371 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Стьюдента | -0.563677839021931 | | 0.5853911903445019 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Уилкоксона | 5.0 | | 0.625 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Манна – Уитни | 16.0 | | 0.8762626262626263 | Принимаем гипотезу о равенстве распределений |
| Критерий Колмогорова – Смирнова | 0.2857142857142857 | | 0.9090909090909092 | Принимаем гипотезу о равенстве распределений |
| Двухвыборочный критерий Колмогорова – Смирнова | 0.2857142857142857 | | 0.9090909090909092 | Принимаем гипотезу о равенстве распределений |

**Ф.М.Достоевский**

|  |  |  |  |
| --- | --- | --- | --- |
| Тест Шапиро – первая и вторая выборки | 0.961934208869931 | 0.7547894716262817 | Принимаем гипотезу о нормальности распределения |
| Тест Шапиро – третья и четвертая выборки | 0.9573202133178711 | 0.5508359670639038 | Принимаем гипотезу о нормальности распределения |
| ANOVA | 0.847071738535625 | 0.4798407730711082 | Принимаем гипотезу о равенстве выборочных средних |
| Краскелла – Уоллиса | 2.4870934267382383 | 0.4776281064692328 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Стьюдента | -1.287739851517813 | 0.2076819810451186 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Уилкоксона | 26.0 | 0.10400390625 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Манна – Уитни | 95.5 | 0.254407050027496 | Принимаем гипотезу о равенстве распределений |
| Критерий Колмогорова – Смирнова | 0.31746031746031744 | 0.33130906533151083 | Принимаем гипотезу о равенстве распределений |
| Двухвыборочный критерий Колмогорова – Смирнова | 0.31746031746031744 | 0.33130906533151083 | Принимаем гипотезу о равенстве распределений |

**А.Н.Толстой**

|  |  |  |  |
| --- | --- | --- | --- |
| Тест Шапиро – первая и вторая выборки | 0.9641405344009399 | 0.8159711360931396 | Принимаем гипотезу о нормальности распределения |
| Тест Шапиро – третья и четвертая выборки | 0.9501833915710449 | 0.7130816578865051 | Принимаем гипотезу о нормальности распределения |
| ANOVA | 1.326027339823393 | 0.2987284545590459 | Принимаем гипотезу о равенстве выборочных средних |
| Краскелла – Уоллиса | 3.2512987012987082 | 0.3544786796248943 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Стьюдента | 0.1886801512445907 | 0.8523442596368178 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Уилкоксона | 14.0 | 0.640625 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Манна – Уитни | 57.0 | 0.7501007420512065 | Принимаем гипотезу о равенстве распределений |
| Критерий Колмогорова – Смирнова | 0.2403846153846154 | 0.8667256376234705 | Принимаем гипотезу о равенстве распределений |
| Двухвыборочный критерий Колмогорова – Смирнова | 0.2403846153846154 | 0.8667256376234705 | Принимаем гипотезу о равенстве распределений |

*Произведения разных авторов*

**Л.Н.Толстой и М.Горький**

|  |  |  |  |
| --- | --- | --- | --- |
| Тест Шапиро – первая и вторая выборки | 0.966253399848938 | 0.8229710459709167 | Принимаем гипотезу о нормальности распределения |
| Тест Шапиро – третья и четвертая выборки | 0.9084219336509705 | 0.38505473732948303 | Принимаем гипотезу о нормальности распределения |
| ANOVA | 107.45121923531421 | 2.9779348212704665e-11 | Отклоняем гипотезу о равенстве выборочных средних |
| Краскелла – Уоллиса | 14.638383838383831 | 0.002153263166318698 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Стьюдента | 13.057989127706307 | 1.3510089372555462e-09 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Уилкоксона | 0.0 | 0.015625 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Манна – Уитни | 70.0 | 0.00010283833813245579 | Отклоняем гипотезу о равенстве распределений |
| Критерий Колмогорова – Смирнова | 1.0 | 0.00010283833813245579 | Отклоняем гипотезу о равенстве распределений |
| Двухвыборочный критерий Колмогорова – Смирнова | 1.0 | 0.00010283833813245579 | Отклоняем гипотезу о равенстве распределений |

**И.С.Тургенев и И.А.Гончаров**

|  |  |  |  |
| --- | --- | --- | --- |
| Тест Шапиро – первая и вторая выборки | 0.816076397895813 | 0.042428158223629 | Отклонить гипотезу о нормальности распределения |
| Тест Шапиро – третья и четвертая выборки | 0.9319276213645935 | 0.3246665596961975 | Принимаем гипотезу о нормальности распределения |
| ANOVA | 16.465912344598067 | 2.122864935345004e-05 | Отклоняем гипотезу о равенстве выборочных средних |
| Краскелла – Уоллиса | 14.694664031620547 | 0.002097074807763693 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Стьюдента | -7.3189729735610 | 4.479526793943783e-07 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Уилкоксона | 0.0 | 0.0078125 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Манна – Уитни | 0.0 | 6.254495418582106e-06 | Отклоняем гипотезу о равенстве распределений |
| Критерий Колмогорова – Смирнова | 1.0 | 6.254495418582106e-06 | Отклоняем гипотезу о равенстве распределений |
| Двухвыборочный критерий Колмогорова – Смирнова | 1.0 | 6.254495418582106e-06 | Отклоняем гипотезу о равенстве распределений |

**Н.СЛесков и Ф.М.Достоевский**

|  |  |  |  |
| --- | --- | --- | --- |
| Тест Шапиро – первая и вторая выборки | 0.9404017925262451 | 0.5032945275306702 | Принимаем гипотезу о нормальности распределения |
| Тест Шапиро – третья и четвертая выборки | 0.917828381061554 | 0.17847254872322083 | Принимаем гипотезу о нормальности распределения |
| ANOVA | 4.6631434644944925 | 0.010930897137135973 | Отклоняем гипотезу о равенстве выборочных средних |
| Краскелла – Уоллиса | 10.1143806178549 | 0.017618466755198722 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Стьюдента | -3.17660076113350 | 0.003935043204045219 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Уилкоксона | 3.0 | 0.06297905121445523 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Манна – Уитни | 28.5 | 0.002907004642823949 | Отклоняем гипотезу о равенстве распределений |
| Критерий Колмогорова – Смирнова | 0.65 | 0.003504169959951357 | Отклоняем гипотезу о равенстве распределений |
| Двухвыборочный критерий Колмогорова – Смирнова | 0.65 | 0.003504169959951357 | Отклоняем гипотезу о равенстве распределений |

**А.Н.Толстой и И.С.Шмелев**

|  |  |  |  |
| --- | --- | --- | --- |
| Тест Шапиро – первая и вторая выборки | 0.8815450072288513 | 0.27622759342193604 | Принимаем гипотезу о нормальности распределения |
| Тест Шапиро – третья и четвертая выборки | 0.9641405344009399 | 0.8159711360931396 | Принимаем гипотезу о нормальности распределения |
| ANOVA | 6.340743403392538 | 0.005444572371527678 | Отклоняем гипотезу о равенстве выборочных средних |
| Краскелла – Уоллиса | 11.06421052631579 | 0.011383683941335116 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Стьюдента | 4.109408419407771 | 0.0007316785091069349 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Уилкоксона | 1.0 | 0.0625 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Манна – Уитни | 74.0 | 0.0008845643520566121 | Отклоняем гипотезу о равенстве распределений |
| Критерий Колмогорова – Смирнова | 0.7692307692307693 | 0.0069290874244434635 | Отклоняем гипотезу о равенстве распределений |
| Двухвыборочный критерий Колмогорова – Смирнова | 0.7692307692307693 | 0.0069290874244434635 | Отклоняем гипотезу о равенстве распределений |

***Уровень значимости = 0.01***

*Произведения одного автора:*

**Л.Н. Толстой**

[23.99375, 23.21875, 23.53125, 24.39375, 24.80625]

[24.075, 23.7375, 24.19375, 23.35625, 24.76875]

[24.456249999999997, 24.31875, 23.70625, 24.631249999999998, 24.425]

[23.73125, 24.0125, 25.025, 24.64375, 24.025, 25.0125, 24.28125, 24.625, 24.34375]

Статистика P-value

|  |  |  |  |
| --- | --- | --- | --- |
| Тест Шапиро – первая и вторая выборки | 0.9531287550926208 | 0.7055749297142029 | Принимаем гипотезу о нормальности распределения |
| Тест Шапиро – третья и четвертая выборки | 0.9526628255844116 | 0.6027610301971436 | Принимаем гипотезу о нормальности распределения |
| ANOVA | 1.1227716916854662 | 0.36350711249099216 | Принимаем гипотезу о равенстве выборочных средних |
| Краскелла – Уоллиса | 2.5688888888888926 | 0.4629698495183314 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Стьюдента | 0.44501961087437913 | 0.6606533322957311 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Уилкоксона | 16.0 | 0.275390625 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Манна – Уитни | 43.0 | 0.12073854334140409 | Принимаем гипотезу о равенстве распределений |
| Критерий Колмогорова – Смирнова | 0.4142857142857143 | 0.20557234751608155 | Принимаем гипотезу о равенстве распределений |
| Двухвыборочный критерий Колмогорова – Смирнова | 0.4142857142857143 | 0.20557234751608155 | Принимаем гипотезу о равенстве распределений |

**И.С.Тургенев**

|  |  |  |  |
| --- | --- | --- | --- |
| Тест Шапиро – первая и вторая выборки | 0.8792604207992554 | 0.22318527102470398 | Принимаем гипотезу о нормальности распределения |
| Тест Шапиро – третья и четвертая выборки | 0.9735967516899109 | 0.8977965712547302 | Принимаем гипотезу о нормальности распределения |
| ANOVA | 0.22743237873766656 | 0.8747554679766818 | Принимаем гипотезу о равенстве выборочных средних |
| Краскелла – Уоллиса | 0.6025641025641093 | 0.8958451305579371 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Стьюдента | -0.563677839021931 | 0.5853911903445019 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Уилкоксона | 5.0 | 0.625 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Манна – Уитни | 16.0 | 0.8762626262626263 | Принимаем гипотезу о равенстве распределений |
| Критерий Колмогорова – Смирнова | 0.2857142857142857 | 0.9090909090909092 | Принимаем гипотезу о равенстве распределений |
| Двухвыборочный критерий Колмогорова – Смирнова | 0.2857142857142857 | 0.9090909090909092 | Принимаем гипотезу о равенстве распределений |

**Ф.М.Достоевский**

|  |  |  |  |
| --- | --- | --- | --- |
| Тест Шапиро – первая и вторая выборки | 0.961934208869931 | 0.7547894716262817 | Принимаем гипотезу о нормальности распределения |
| Тест Шапиро – третья и четвертая выборки | 0.9573202133178711 | 0.5508359670639038 | Принимаем гипотезу о нормальности распределения |
| ANOVA | 0.847071738535625 | 0.4798407730711082 | Принимаем гипотезу о равенстве выборочных средних |
| Краскелла – Уоллиса | 2.4870934267382383 | 0.4776281064692328 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Стьюдента | -1.287739851517813 | 0.2076819810451186 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Уилкоксона | 26.0 | 0.10400390625 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Манна – Уитни | 95.5 | 0.254407050027496 | Принимаем гипотезу о равенстве распределений |
| Критерий Колмогорова – Смирнова | 0.31746031746031744 | 0.33130906533151083 | Принимаем гипотезу о равенстве распределений |
| Двухвыборочный критерий Колмогорова – Смирнова | 0.31746031746031744 | 0.33130906533151083 | Принимаем гипотезу о равенстве распределений |

**А.Н.Толстой**

|  |  |  |  |
| --- | --- | --- | --- |
| Тест Шапиро – первая и вторая выборки | 0.9641405344009399 | 0.8159711360931396 | Принимаем гипотезу о нормальности распределения |
| Тест Шапиро – третья и четвертая выборки | 0.9501833915710449 | 0.7130816578865051 | Принимаем гипотезу о нормальности распределения |
| ANOVA | 1.326027339823393 | 0.2987284545590459 | Принимаем гипотезу о равенстве выборочных средних |
| Краскелла – Уоллиса | 3.2512987012987082 | 0.3544786796248943 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Стьюдента | 0.1886801512445907 | 0.8523442596368178 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Уилкоксона | 14.0 | 0.640625 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Манна – Уитни | 57.0 | 0.7501007420512065 | Принимаем гипотезу о равенстве распределений |
| Критерий Колмогорова – Смирнова | 0.2403846153846154 | 0.8667256376234705 | Принимаем гипотезу о равенстве распределений |
| Двухвыборочный критерий Колмогорова – Смирнова | 0.2403846153846154 | 0.8667256376234705 | Принимаем гипотезу о равенстве распределений |

*Произведения разных авторов*

**Л.Н.Толстой и М.Горький**

|  |  |  |  |
| --- | --- | --- | --- |
| Тест Шапиро – первая и вторая выборки | 0.966253399848938 | 0.8229710459709167 | Принимаем гипотезу о нормальности распределения |
| Тест Шапиро – третья и четвертая выборки | 0.9084219336509705 | 0.38505473732948303 | Принимаем гипотезу о нормальности распределения |
| ANOVA | 107.45121923531421 | 2.9779348212704665e-11 | Отклоняем гипотезу о равенстве выборочных средних |
| Краскелла – Уоллиса | 14.638383838383831 | 0.002153263166318698 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Стьюдента | 13.057989127706307 | 1.3510089372555462e-09 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Уилкоксона | 0.0 | 0.015625 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Манна – Уитни | 70.0 | 0.00010283833813245579 | Отклоняем гипотезу о равенстве распределений |
| Критерий Колмогорова – Смирнова | 1.0 | 0.00010283833813245579 | Отклоняем гипотезу о равенстве распределений |
| Двухвыборочный критерий Колмогорова – Смирнова | 1.0 | 0.00010283833813245579 | Отклоняем гипотезу о равенстве распределений |

**И.С.Тургенев и И.А.Гончаров**

|  |  |  |  |
| --- | --- | --- | --- |
| Тест Шапиро – первая и вторая выборки | 0.816076397895813 | 0.042428158223629 | Принимаем гипотезу о нормальности распределения |
| Тест Шапиро – третья и четвертая выборки | 0.9319276213645935 | 0.3246665596961975 | Принимаем гипотезу о нормальности распределения |
| ANOVA | 16.465912344598067 | 2.122864935345004e-05 | Отклоняем гипотезу о равенстве выборочных средних |
| Краскелла – Уоллиса | 14.694664031620547 | 0.002097074807763693 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Стьюдента | -7.3189729735610 | 4.479526793943783e-07 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Уилкоксона | 0.0 | 0.0078125 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Манна – Уитни | 0.0 | 6.254495418582106e-06 | Отклоняем гипотезу о равенстве распределений |
| Критерий Колмогорова – Смирнова | 1.0 | 6.254495418582106e-06 | Отклоняем гипотезу о равенстве распределений |
| Двухвыборочный критерий Колмогорова – Смирнова | 1.0 | 6.254495418582106e-06 | Отклоняем гипотезу о равенстве распределений |

**Н.СЛесков и Ф.М.Достоевский**

|  |  |  |  |
| --- | --- | --- | --- |
| Тест Шапиро – первая и вторая выборки | 0.9404017925262451 | 0.5032945275306702 | Принимаем гипотезу о нормальности распределения |
| Тест Шапиро – третья и четвертая выборки | 0.917828381061554 | 0.17847254872322083 | Принимаем гипотезу о нормальности распределения |
| ANOVA | 4.6631434644944925 | 0.010930897137135973 | Принимаем гипотезу о равенстве выборочных средних |
| Краскелла – Уоллиса | 10.1143806178549 | 0.017618466755198722 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Стьюдента | -3.17660076113350 | 0.003935043204045219 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Уилкоксона | 3.0 | 0.06297905121445523 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Манна – Уитни | 28.5 | 0.002907004642823949 | Отклоняем гипотезу о равенстве распределений |
| Критерий Колмогорова – Смирнова | 0.65 | 0.003504169959951357 | Отклоняем гипотезу о равенстве распределений |
| Двухвыборочный критерий Колмогорова – Смирнова | 0.65 | 0.003504169959951357 | Отклоняем гипотезу о равенстве распределений |

**А.Н.Толстой и И.С.Шмелев**

|  |  |  |  |
| --- | --- | --- | --- |
| Тест Шапиро – первая и вторая выборки | 0.8815450072288513 | 0.27622759342193604 | Принимаем гипотезу о нормальности распределения |
| Тест Шапиро – третья и четвертая выборки | 0.9641405344009399 | 0.8159711360931396 | Принимаем гипотезу о нормальности распределения |
| ANOVA | 6.340743403392538 | 0.005444572371527678 | Отклоняем гипотезу о равенстве выборочных средних |
| Краскелла – Уоллиса | 11.06421052631579 | 0.011383683941335116 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Стьюдента | 4.109408419407771 | 0.0007316785091069349 | Отклоняем гипотезу о равенстве выборочных средних |
| Тест Уилкоксона | 1.0 | 0.0625 | Принимаем гипотезу о равенстве выборочных средних |
| Тест Манна – Уитни | 74.0 | 0.0008845643520566121 | Отклоняем гипотезу о равенстве распределений |
| Критерий Колмогорова – Смирнова | 0.7692307692307693 | 0.0069290874244434635 | Отклоняем гипотезу о равенстве распределений |
| Двухвыборочный критерий Колмогорова – Смирнова | 0.7692307692307693 | 0.0069290874244434635 | Отклоняем гипотезу о равенстве распределений |

Вывод

Было написано два программных средства, для нахождения авторского инварианта и проверки статистических гипотез. Согласно критерию, озвученному в статье Фоменко и Фоменко, колебания авторского инварианта внутри произведений одного писателя должны быть много меньше колебаний между произведениями разных авторов, что мы и видим в данном случае. Также была проведена проверка статистических гипотез при различных уровнях значимости, которые (гипотезы) в большинстве своем подтвердили корректность нахождения инварианта и его достаточную степень различимости для произведений разных авторов. Можно сделать вывод о том, что при уровне значимости = **0.5**, средние значения и соответствующие распределения у одинаковых авторов совпадают, а у различных авторов – отличаются, причем это выполняется во всех случаях. При уровне же значимости в **0.01**, для произведений одного автора все по – прежнему стабильно, а вот для разных авторов в некоторых случаях наблюдается расхождения в том, что указывается равенство их средних значений. Однако же во всем остальном даже и этот уровень значимости позволяет сформулировать статистический вывод о том, что соответствующие гипотезы подтверждаются, а значит, согласно таблице инвариантов авторов и проверке гипотез, работа проделана не бессмысленно.