**Contents**

**1 Критерии согласия 1**

**1 Критерий Пирсона 1**

**2 Критерий Колмогорова 4**

2.0.1 Алгоритм проверки гипотезы. 4

**2 Критерии однородности и независимости 5**

**1 Критерии однородности и независимости двух полных выборок 5**

1.1 Критерий Колмогорова-Смирнова 5

1.2 Критерий знаков 6

1.3 Критерии Вилкоксона и Манна-Уитни 8

1.4 Критерий серий 11

**2 Проверка однородности и независимости нескольких полных выборок 12**

2.1 Медианный критерий 12

2.2 Критерий Краскела - Уоллиса 14

2.2.1 Критерий 15

**3 Hезависимость сопряженных признаков 16**

3.1 Коэффициент ранговой корреляции Спирмэна 16

3.2 Коэффициент ранговой корреляции Кэндела 18

3.3 Критерий 19

# 1 Критерии согласия

Под критериями согласия понимается проверка гипотезы о согласии имеющихся статистических данных с выбранным теоретическим распределением. Многообразие приводимых критериев необходимо по существу дела. В качестве конкурирующей гипотезы могут выступать различные распределения. Чтобы избежать ошибок, рекомендуется проверку гипотезы о виде закона распределения, наблюдаемой случайной величины, проводить по нескольким критериям.

В настоящее время для проверки гипотез о согласии эмпирических распределений времени до отказа с теоретическими моделями по полным выборкам из непараметрических критериев чаще всего используют один из следующих трех критериев или их модификаций: Колмогорова; Пирсона или Крамера-Мизеса.

## 1 Критерий Пирсона

Пусть - выборка наблюдений случайной величины . Проверяется гипотеза утверждающая, что имеет закон распределения . Процедура применения критерия для проверки гипотезы состоит из следующих этапов.

1. По выборке наблюдений случайной величины найти оценки неизвестных параметров предполагаемого закона распределения Обозначим их число через Если распределение определено точно, то пункт 1 надо пропустить, при этом положим

2. Если -дискретная случайная величина, то определить частоты с которыми каждое значение или группа значений встречается в выборке. Если - непрерывная случайная величина, то разбить область ее значений на непересекающихся интервалов и определить число элементов выборки принадлежащие каждому интервалу. Очевидно, что в обоих случаях

3. В случае, если - дискретная случайная величина, используя преполагаемый закон распределения вычислить вероятности с которыми случайная величина принимает каждое значения, или вероятность появления группы значений. В случае, если - непрерывная случайная величина, следует определить вероятность попадания в каждый интервал

В обоих случаях

4. Вычислить выборочное значение статистики критерия

5. Принять статистическое решение: гипотеза не противоречит выборке наблюдений на заданном уровне значимости если ; если же то гипотеза отклоняется.

**Замечание.** Критерий использует тот факт, что случайная величина

имеет распределение близкое к нормальному Чтобы это утверждение было достаточно точным, необходимо, чтобы для всех интервалов выполнялось условие Интервалы в которых это условие не выполняется следует обьеденить с соседними. Это действие уменьшает опасность обесценивания -аппроксимации, но к сожелению приводят к снижению чувствительности критерия уменьшая количество эффективных интервалов. Следует избегать обьединение ячеек, кроме тех случаев, когда это действительно необходимо.

Известный статистик У.Кокрен предлагал: при унимодулярных распределения, когда ожидаемые частоты будут малы только на "хвостах", следует добиваться того, чтобы минимальная ожидаемая частота на каждом "хвосте" была не меньше 1.

**Пример 1.** Проверка гипотезы о распределении по закону Пуассона. В первых двух столбцах таблицы 1 приведены данные об отказах аппаратуры за часов работы. Общее число обследованных экземпляров аппаратуры при этом наблюдался отказ.

Проверить гипотезу о том, что число отказов имеет распределение Пуассона с надежностью 0.9.

Table 1:

|  |  |  |  |
| --- | --- | --- | --- |
| число | количество случаев, |  | ожидаемое число |
| отказов | в которых наблюда- |  | случаев с отказа- |
|  | лось отказов, |  | ми, |
| 0 | 427 | 0.54881 | 416 |
| 1 | 235 | 0.32929 | 249 |
| 2 | 72 | 0.09879 | 75 |
| 3 | 21 | 0.01976 | 15 |
| 4 | 1 | 0.00296 | 2 |
| 5 | 1 | 0.00036 | 0 |
|  | 0 | 0.00004 | 0 |
| Сумма | 757 | – | – |

Оценка параметра равна среднему числу отказов: По таблице распределения Пуассона с находим вероятности и ожидаемое число с отказами (третий и четвертый столбцы таблицы 1).

Для значения поэтому обьединим эти строки со сторокой для . В результате получим значения приведенные в таблице 2.

По выборке оценивался один параметр , то , число степеней свободы равно По таблице квартилей распределения находим следовательно, гипотеза о распределениии числа отказов по закону Пуассона не противоречит опытным данным.

Table 2:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
| 0 | 427 | 416 | 0.291 |
| 1 | 235 | 249 | 0.787 |
| 2 | 72 | 75 | 0.120 |
|  | 23 | 17 | 2.118 |
|  |  |  |  |
|  |  |  |  |
| – | – | – |  |
|  |  |  |  |

## 2 Критерий Колмогорова

Пусть наблюдения над случайной величиной, функцию распределения которой обозначим через При этом предположим, что непрерывная функция. Hа основании статистических данных построим эмпирическую функцию распределения , как описано в главе 2. Обозначим через

расстояние между теоретической и эмпирической функциями распределения. Статистику принято называть статистикой Колмогорова.

Теорема Колмогорова утверждает, что при введенных выше ограничениях величина при имеет предел, и он равен:

где распределение Колмогорова, значения которого протабулированы.

#### 2.0.1 Алгоритм проверки гипотезы.

1. Для вычисления статистики можно использовать формулу

2. Для уровня значимости из таблиц распределения Колмогорова берут такое значение при котором Если то говорят, что гипотеза не протеворечит опытным данным. Если то гипотезу надо отбросить.

**Замечание** Более подробно познакомится с возможностью применения критерия Колмогорова в зависимости от обьема выборки можно в работе [2].

**Пример 1.** При работе генератора псевдослучайных чисел была получена некоторая выборка случайных величин с равномерным на [0,1] распределением. Требуется оценить алгоритм генерации псевдослучайных чисел. Надежность принятия гипотезы принять 0.9.

Найдем максимальное значение между разностями

.

Полученная статистика обозначается . Далее ищется максимальное значение между разностями

.

Эта статистика обозначается . В конкретном примере

.

Максимальное из этих двух чисел и называется статистикой Колмогорова. В данном случае . Для уровня значисмости 0.1 по таблицам значения функции Колмогорова находится значение для оценки статистики .

.

Следовательно, данный датчик случайных чисел не оправдал себя. Гипотеза о равномерности распределения не принимается. (В примере был использован датчик, встроенный в MicroSoft Excell 97).

# 2 Критерии однородности и независимости

## 1 Критерии однородности и независимости двух полных выборок

Пусть имеются две серии результатов независимых наблюдений: - над случайной величиной с функцией распределения - над случайной величиной с функцией распределения . Требуется проверить гипотезу . В качестве альтернативных гипотез могут рассматриваться Следует отметить, что в приведенных критериях предпологается независимость выборок и поэтому получение отрицательного результата может означать, что нет независимости.

### 1.1 Критерий Колмогорова-Смирнова

В некоторых источниках он назван критерием Смирнова, который его разработал и предложил. Данный критерий является состоятельным, то есть такой, для которого вероятность принять ложную гипотезу стремиться к нулю с ростом числа испытаний до бесконечности.

Для первой выборки строим эмпирическую функцию распределения по формуле:

Аналогично для второй выборки строим функцию .

Рассмотрим величины

Как доказал H.В.Смирнов, при имеют место предельные соотношения

и

Через обозначают функцию Колмогорова. При существуют более простые предельные соотношения.

В общем случае, в качестве критической области размера можно взять

где – константа, которая выбирается так, чтобы . Точное выборочное распределение при выполнении неизвестно. Таблицы для при и и даны в [3].

Для больших можно использовать вышеописанные аппроксимации [2].

Для проверки гипотезы об однородности двух выборок на практике значения статистики часто применяют следующие формулы, эквивалентные предыдущим

Здесь и - это те же величины и , но расположенные в порядке возрастания их значений.

**Пример**  В результате проверки партий электрических лампочек, выпускаемых отечественным и импортным производителем, подсчитывалось количество сгоревших. Данные приведены в таблице.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |
| Отечественные | 4 | 9 | 2 | 10 | 12 | 9 | 11 | 0 | 4 | 1 |
| Импортные | 2 | 12 | 11 | 12 | 15 | 5 | 14 | 12 | 4 | 0 |

По приведенным выше формулам найходим .

По таблицам [?]. находим критическое значение статистики . На уровне значимости 0.05 . Следовательно, на данном уровне значимости можно предположить, что отечественные и импортные лампочки имеют одинаковую длительность горения.

### 1.2 Критерий знаков

Прост в применении, но использует информацию очень неэкономно. Полезен для предварительной прикидки. Строим разности

В силу независимости и случайная величина может принимать положительные и отрицательные значения с равной вероятностью. Вероятность того, что непрерывная случайная величина примет значение равное нулю равна нулю. Мы находимся в условиях схемы Бернули. Это означает, что число положительных разностей сравнительно немного отклоняется от . Если это не так, то гипотеза ошибочная.

вероятность k - успехов в l - опытах.

Пусть - уровень значимости. Отбросим , если число положительных разностей больше числа , где

-  **односторонний критерий знаков**. Аналогично, если число отрицательных разностей велико.

**Двухсторонний критерий знаков.** Если число положительных разностей больше или меньше , то отбрасываем со значимостью .

Для l от 5 до 100 составлены границы критической области в случае 1 %, 2 % и 5% - ых уровней значимости для двухсторонних критериев знаков. При можно использовать интегральную теорему Муавра - Лапласа. т.е.

- определена из таблиц нормального распределения (функция Лапласа)

**Замечание** Нулевые значения случайной величины следует исключить из рассмотрения.

**Пример** Предполагается, что один из двух приборов, определеяющих скорость автомобиля, имеет систематическую ошибку. Для проверки этого предположения определили скорость 10 автомобилей, причем скорость каждого фиксировалась двумя приборами. В результате получены следующие данные:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |
|  | 70 | 85 | 63 | 54 | 65 | 80 | 75 | 95 | 52 | 55 |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  | 72 | 86 | 62 | 55 | 63 | 80 | 78 | 90 | 53 | 57 |
|  |  |  |  |  |  |  |  |  |  |  |
| знак |  |  |  |  |  |  |  |  |  |  |
| разностей | – | – | + | – | + | 0 | – | + | – | – |
|  |  |  |  |  |  |  |  |  |  |  |

Позволяют ли эти результаты утверждать, что второй прибор действительно дает завышенные значения скорости? Принять

Составим последовательность знаков разностей, они приведены в таблице. Число ненулевых разностей l=9, число положительных разностей Критические значения равны 2 и 7. оснований отвергнуть гипотезу нет.

Часто удобнее проводить проверку гипотезы используя  **статистику Фишера:** гипотеза отклоняется, если при выполняется неравенство

(1)

где при выполняется неравенство

(2)

где при должно выполняться одно из неравенств (1),(2) с заменой на

**Пример** Рассмотрим данные из предыдущего примера и применим для проверки гипотезы при вычислении критической области статистику Фишера.

следовательно, гипотеза не противоречит результатам наблюдения

**Замечание** применяя критерий знаков, при подсчете числа знаков берутся неупорядоченные по величине результаты наблюдений. Если у нее есть две упорядоченные последовательности отказов, полученные например по плану , то можно их случайным образом перемешивать. При этом цензурирование наблюдения надо, как правило, исключить.

### 1.3 Критерии Вилкоксона и Манна-Уитни

Пусть величины и имеют непрерывные функции раcпределения и . Случайная величина меньше чем означает, что . Так как возрастает медленнее, это означает, что при каждом х

Перемешаем обе последовательности наблюдений и расположим их в порядке возрастания. Если верна, то можно ожидать хорошего перемешивания

Для оценки степени перемешивания в критерии Манна-Уитни подсчитывается число инверсий членов первой последовательности относительно второй. Если х предшествует k - значений y, то число инверсий x = k. Общее число инверсий равно сумме числа инверсий всех х.

Согласно одностороннему критерию Манна-Уитни гипотеза отбрасывается, если наблюдаемое число инверсий превосходит некоторую границу .

В двухстороннем критерии Манна-Уитни гипотеза отбрасывается когда число инверсий , так и в случае .

Определим эти границы. Пусть

т.к. - независимые и имеют одну функцию распределения ,

Статистикой Манна-Уитни назовем величину

Если , то

Приближение хорошее, если или При малых значениях m и n для критерия Манна-Уитни составлены таблицы критических значений с заданым уровнем значимости .

Значения для и приведены в [3].

С критерием Манна-Уитни тесно связана его друга форма называемая критерием Вилкоксона (Wilcoxon). В некоторых книгах на русском языке применияют транскрипцию Уилкоксон. Исторически это первый критерий основанный на понятии ранга элемента выборки. Он появился примерно 1945 году.

Каждому элементу упорядоченной объединенной выборки поставим ее номер в ряду - ранг. Если несколько элементов совпадают по величине, то каждому из них присваевается ранг равный среднему арифметическому их номеров.

Пусть - сумма рангов первой выборки. - сумма рангов второй выборки. Статистикой Вилкоксона называют Статистики Манна-Уитни и Вилкоксона связаны следующим образом:

**Правильность проверяется по формуле:**

Статистика Вилкоксона более удобна для практических расчетов без применения ЭВМ.

**Пример** Измерялось напряжение пробоя у диодов, отобранных случайным образом из двух партий. Результаты измерения (в вольтах) следующие:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1-я партия | 39 | 50 | 61 | 67 | 40 | 40 | 54 | – |
| 2-я партия | 60 | 53 | 42 | 41 | 40 | 54 | 63 | 69 |

Можно ли считать, что у диодов второй партии напряжение пробоя выше, чем у диодов первой партии? Принять

Составим вариационный ряд, отмечая принадлежность элемента к первой партии черточкой сверху. В результате получим ранжированную последовательность:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Элемент |  |  |  | 40 | 41 | 42 |  | 53 |  | 54 | 60 |  | 63 |  | 69 |
| ранг | 1 | 3 | 3 | 3 | 5 | 6 | 7 | 8 | 9.5 | 9.5 | 11 | 12 | 13 | 14 | 15 |

Сумма рангов первой выборки сумма рангов второй выборки Hаходим

Проверяем правильность вычислений:

Выборочное значение статистики равно меньшему из чисел 34.5 и 21.5, т.е.

по таблице из [3] при т.к. , то нет оснований считать, что напряжение пробоя у диодов второй партии больше, чем у диодов первой партии.

**Пример 2** В условиях преведущего примера получены результаты новой серии измерения пробоя у диодов (в вольтах):

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1-я партия | 50 | 41 | 48 | 60 | 46 | 60 | 51 | 42 | 62 | 54 | 42 | 46 |
| 2-я партия | 38 | 40 | 47 | 51 | 63 | 50 | 63 | 57 | 59 | 51 | – | – |

Имеется ли основания утверждать, что напряжение пробоя у диодов обеих партий различно?

Упорядочим результаты измерений и определим ранги каждого результата.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Элемент | 38 | 40 |  |  |  |  |  | 47 |  |  | 50 |
| ранг | 1 | 2 | 3 | 4.5 | 4.5 | 6.5 | 6.5 | 8 | 9 | 10.5 | 10.5 |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Элемент |  | 51 | 51 |  | 57 | 59 |  |  |  | 63 | 63 |
| ранг | 13 | 13 | 13 | 15 | 16 | 17 | 18.5 | 18.5 | 20 | 21.5 | 21.5 |

Hайдем суммы рангов

т.к. , то гипотезу следует отклонить в пользу гипотезы .

**Замечание** Критерии знаков и Вилкоксона-Манна-Уитни не могут служить для доказательства гипотезы Для обоснования этого факта рассмотрим пример.

**Пример** Рассмотрим две случайных величины. Первая имеет функцию распределения где

- неубывающие, непрерывные и возрастает от 0 до ; - от до 1 . Вторая имеет функцию распределения где

непрерывная функция, возрастающая от 0 до 1.

Число знаков всегда будет вблизи , где а число инверсий вблизи . Таким образом оба критерия приведут к ошибочному результату. И результат будет тем точнее, чем больше n, m.

### 1.4 Критерий серий

Нулевая гипотеза наблюдений. - медиана. Каждому элементу выборки поставлен в соответствие знак + или - в зависимости от, того больше от медианы или меньще. Пусть - число плюсов, а - число минусов. Серией называется последовательность из одинаковых знаков и ограниченная противоположными.

Статистикой критерия является число серий . Критическая область определяется неравенствами и которые определяются из таблиц при малых n. Если , то

критическая область определяется неравенством

**Замечание** Критерий применим когда для проверки случайности любой выборки элементам которой являются два символа, 0 и 1, А и В, + и - .

**Пример 1** Скорости автомобилей в некоторой точке трассы образовали следующий ряд (км/час):

Можно ли считать полученные значения случайными? Принять

Hайдем оценку медианы. Для этого составим вариационный ряд

Оценка медианы равна Исходному ряду наблюдений соответствует следующая последовательность знаков: - + + + - - - + - - + +, где По таблице Таким образом, гипотеза - принимается: полученные скорости можно считать случайными.

**Пример 2** Можно ли считать, что последовательность

получена из совокупности случайных последовательностей? Принять

В данной последовательности число нулей , а число единиц Число серий Так как то

- принимается

## 2 Проверка однородности и независимости нескольких полных выборок

Разовьем те же идеи для сравнения нескольких выборок. Пусть k совокупностей

Альтернатива может означат изменение медиан, а не формы распределения.

### 2.1 Медианный критерий

Медианный критерий является обобщением критерия серий. Строим объединенный вариационный ряд и определяем медиану выборки. В каждой выборке подсчитываем число наблюдений, которые больше или меньше этой выборочной медианы. Условимся игнорировать наблюдения, равные медиане.

При условии выполнения можно ожидать, что около половины каждой выборки из каждой совокупности будет меньше общей выборочной медианы и около половины будет больше. При условии, что число наблюдений в каждой выборке можно использовать критерий согласия

с k-1 степенью свободы

**Пример**

При выборочном контроле из трех партий были получены следующие данные наработки на отказ

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| -я партия | 21 | 60 | 6 | 69 | 42 | 34 | 26 | 57 | 14 | 31 |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| -я партия | 10 | 49 | 22 | 40 | 24 | 54 | 12 | 29 | 25 | 17 | 32 | 61 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| -я партия | 3 | 15 | 9 | 18 | 1 | 33 | 11 | 5 | 16 | 30 | 41 |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |

Можно ли утверждать с надежностью , что изделия во всех партиях однородные?

В нашем примере . Вторая выборка уменьшается на 1 значение. В результате получим

|  |  |  |  |
| --- | --- | --- | --- |
|  | больше | меньше | всего |
|  | медианы | медианы |  |
|  |  |  |  |
| -я партия | 7 | 3 | 10 |
|  |  |  |  |
| -я партия | 6 | 5 | 11 |
|  |  |  |  |
| -я партия | 3 | 8 | 11 |
|  |  |  |  |
| всего | 16 | 16 | 32 |
|  |  |  |  |

Ожидаемое распределение имеет вид

|  |  |  |  |
| --- | --- | --- | --- |
|  | больше | меньше | всего |
|  | медианы | медианы |  |
|  |  |  |  |
| -я партия | 5 | 5 | 10 |
|  |  |  |  |
| -я партия | 5.5 | 5.5 | 11 |
|  |  |  |  |
| -я партия | 5.5 | 5.5 | 11 |
|  |  |  |  |
| всего | 16 | 16 | 32 |
|  |  |  |  |

Выборочное значение равно

следовательно нет оснований отвергнуть на 5 % уровне значимости о равенстве распределений.

### 2.2 Критерий Краскела - Уоллиса

Данный критерий является аналогом критерия Вилкоксона. Он более полно использует информацию, чем предыдущий критерий.

Сделаем объединеную выборку. Для каждой из выборок вычислим суммы рангов. - cумма рангов элементов i -oй выборки.

Для каждого наблюдения в выборе можно указать средний ранг Если верна гипотеза , то можно ожидать, что все ранги примерно равны, и равны общему среднему

В качестве статистики критерия можно использовать меру, которая чувствительна к отклонениям от .

отвергают если , которая определяется из таблиц при небольших и k . Если , то

**Пример** (Данные взяты из примера предыдущего пункта)

Составим обьедененную выборку. Элементы первой выборки пометим чертой наверху , элементы второй через .

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| элементы | 1 | 3 | 5 |  | 9 |  | 11 |  |  | 15 | 16 |  |
| ранг | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| элементы | 18 |  |  |  |  |  |  | 30 |  |  | 33 |  |
| ранг | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| элементы |  | 41 |  |  |  |  |  |  |  |  |  |  |
| ранг | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 |  |  |  |

Сосчитаем ранги для каждой выборки

В нашем случае

Так как то гипотезу отвергаем на 5 % уровне.

#### 2.2.1 Критерий

Пусть ), независимые случайные выборки. Разобьем интервал , где , на интервалов. Составим таблицу сопряженности признаков.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| признаки | 1 | 2 |  | s | сумма |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| r |  |  |  |  |  |
| сумма |  |  |  |  |  |

Где

• — количество наблюдений в –ой выборке, попавших в –ый интервал;

• — общее количество наблюдений во всех выборках;

• ;

• .

Статистика строится следующим образом:

Число степеней свободы .

**Пример** Проверяли возраст сотрудников некоторых фирм. В результате были получены следующие значения.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 50 | 21 | 20 | 43 | 30 | 49 | 29 | 44 | 35 | 21 | 26 | 37 | 22 | 20 | 25 |
|  | 20 | 21 | 44 | 39 | 44 | 28 | 24 | 49 | 37 | 41 | 37 | 45 | 50 | 41 |
|  | 43 | 33 | 34 | 47 | 31 | 44 | 21 | 39 | 27 | 48 | 44 | 42 | 33 | 33 |

Проверить данные на независимость. Принять

Разобьем данные выборок на 3 интервала. По данных выборкам составим таблицу сопряженности признаков.

|  |  |  |  |
| --- | --- | --- | --- |
| 8 | 4 | 2 | 14 |
|  | 3 | 6 | 12 |
|  | 8 | 7 | 19 |
|  | 15 | 15 | 45 |

Найдем оценку статистики .

Число степеней свободы . Отсюда, . По таблицам находим точное значение статистики при k степенях свободы . Таким образом, выборки случайных величин принадлежат одной и той же генеральной совокупности на данном уровне значимости.

## 3 Hезависимость сопряженных признаков



Пусть мы имеем пар , которые составляют случайную выборку из некоторого двумерного непрерывного распределения.

Нулевая гипотеза предполагает независимость случайных величин и :

Опишем проверку гипотезы о независимости используя понятия ковариации и корреляции. Для этого нужно построить их статические оценки и если они не близки к нулю, то гипотеза о независимости отвергается. Можно построить аналогичные меры связи при непараметрическом походе, используя понятие рангов наблюдения при упорядочивании.

### 3.1 Коэффициент ранговой корреляции Спирмэна

Пусть "n" пар наблюдений . Составляем вариационный ряд и . Если две переменные сильно зависимы, то мы вправе ожидать, что ранги двух элементов примерно одинаковые.

где

Cвойства х и у сильно связаны.

При гипотезе существуют таблицы распределения .

Пусть то

отбрасываем, если Аналогично можно вывести односторонний критерий.

**Замечание** Если имеются связи в двух ранжируемых множествах, то необходима корекция , но эффективность корорекции мала, если доля связок невелика.

**Пример 1.** При тестировании двумя тестами испытуемые набрали баллы

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| индивидуум | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1-й тест | 31 | 82 | 25 | 26 | 53 | 30 | 29 |
| 2-й тест | 21 | 55 | 8 | 27 | 32 | 42 | 26 |

Проверить гипотезу о независимости двух тестов.

Ранги имеют вид

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| индивидуум | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1-й тест | 5 | 7 | 1 | 2 | 6 | 4 | 3 |
| 2-й тест | 2 | 7 | 1 | 4 | 5 | 6 | 3 |
| разность | 3 | 0 | 0 | -2 | 1 | -2 | 0 |

Оснований отвергнуть гипотезу нет.

**Пример 2.** У группы студентов измеряли вес и рост. Данные приведены в таблице

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| индивидуум | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| вес | 68.8 | 63.3 | 75.7 | 67.2 | 71.3 | 72.8 | 76.5 | 63.5 | 69.9 | 71.4 |
| рост | 167 | 113 | 159 | 153 | 150 | 181 | 173 | 115 | 125 | 166 |

Проверить независимость признаков.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| индивидуум | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| вес | 4 | 1 | 9 | 3 | 6 | 8 | 10 | 2 | 5 | 7 |
| рост | 8 | 1 | 6 | 5 | 4 | 10 | 9 | 2 | 3 | 7 |
| разность | -4 | 0 | 3 | -2 | 2 | -2 | 1 | 0 | 2 | 0 |

Ранговая корреляция значима

### 3.2 Коэффициент ранговой корреляции Кэндела

Упорядочим индивидумов по возрастанию значений х. На основании ранжирования поставим баллы. Ставим +1 столько раз, сколько индивидумов правее его получили больше баллов, чем он (ранги стоят в правильном порядке). Аналогично - 1 для каждого ранга справа, который меньше данного.

Коэффициент ранговой корреляции Кэндела

Знаменатель равен максимально возможному числу баллов, которое будет вычислено, когда имеется полное согласие между ранжировками . Свойства аналогичны .

Критическая область размера для проверки против имеет вид

Для определения таблицы. Для

При наличии связок в наших ранжировках мы должны пересмотреть .

где T число связок при ранжировании х, s число связок при ранжировании y, j - я связка состоит из наблюдений.

**Пример** (из предыдущего раздела рассмотрим пример 1)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| индивидуум | 3 | 4 | 7 | 6 | 1 | 5 | 2 |
| ранг 1-го теста | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| ранг 2-го теста | 1 | 4 | 3 | 6 | 2 | 5 | 7 |
| баллы | 6 | 1 | 2 | -1 | 2 | 1 | 0 |

Сосчитаем баллы

По таблице достигаем уровень значимости равен 0,14. Нет оснований отвергнуть гипотезу .

### 3.3 Критерий

Построим таблицу распределения двух исследуемых величин [?]. Для каждой случайной величины определим разряды попадания, и полученные две группы разрядов разместим в заголовках таблицы. В каждую клетку таблицы поместим частоты попаданий пар совместно наблюденных значений случайных величин в разряды, определяемые заголовками таблицы. Обозначим их через , , , — количество разрядов для случайных величин, соответственно и .

Построим выравнивающие частоты:

Статистика строится следующим образом:

Число степеней свободы .

По значениям и находится величина , которая будет выражать степень расхождения между наблюденными и выравнивающими частотами. Критическая область:

## Значение функции Лапласа и ее производной

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| 0.00 | 0.0000 | 0.3989 | 2.00 | 0.4772 | 0.540 |
| 0.10 | 0.03983 | 0.3970 | 2.10 | 0.4821 | 0.0440 |
| 0.20 | 0.079926 | 0.3910 | 2.20 | 0.4861 | 0.0355 |
| 0.30 | 0.1179 | 0.3814 | 2.30 | 0.4893 | 0.0283 |
| 0.40 | 0.1554 | 0.3683 | 2.40 | 0.4918 | 0.0224 |
| 0.50 | 0.1915 | 0.3581 | 2.50 | 0.4938 | 0.0175 |
| 0.60 | 0.2257 | 0.3332 | 2.60 | 0.4953 | 0.0136 |
| 0.70 | 0.2580 | 0.3123 | 2.70 | 0.4965 | 0.0104 |
| 0.80 | 0.2881 | 0.2897 | 2.80 | 0.4974 | 0.0079 |
| 0.90 | 0.3159 | 0.2661 | 2.90 | 0.4981 | 0.0060 |
| 1.00 | 0.3413 | 0.2420 | 3.00 | 0.49865 | 0.0044 |
| 1.10 | 0.3643 | 0.2179 | 3.20 | 0.49931 | 0.0024 |
| 1.20 | 0.3849 | 0.1942 | 3.40 | 0.49966 | 0.0012 |
| 1.30 | 0.4032 | 0.1714 | 3.60 | 0.499841 | 0.0006 |
| 1.40 | 0.4192 | 0.1497 | 3.80 | 0.499928 | 0.0003 |
| 1.50 | 0.4332 | 0.1295 | 4.00 | 0.499968 | 0.00001 |
| 1.60 | 0.4452 | 0.1109 | 4.50 | 0.499997 | 0.00000 |
| 1.70 | 0.4554 | 0.0940 | 5.00 | 0.49999997 | 0.00000 |
| 1.80 | 0.4641 | 0.0790 |  |  |  |
| 1.90 | 0.4713 | 0.0656 |  |  |  |
|  |  |  |  |  |  |

## Значение функции Пирсона

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
| n|x |  |  |  |  |  |  |  |
| 1 | 6.600 | 3.800 | 2.700 | 0.455 | 0.016 | 0.0039 | 0.00016 |
| 2 | 9.200 | 5.900 | 4.600 | 1.386 | 0.211 | 0.103 | 0.020 |
| 3 | 11.300 | 7.800 | 6.300 | 2.366 | 0.584 | 0.352 | 0.115 |
| 4 | 13.30 | 9.500 | 7.800 | 3.360 | 1.060 | 0.710 | 0.300 |
| 5 | 15.10 | 11.10 | 9.200 | 4.350 | 1.610 | 1.140 | 0.550 |
| 6 | 16.80 | 12.60 | 10.60 | 5.350 | 2.200 | 1.630 | 0.870 |
| 7 | 18.50 | 14.10 | 12.00 | 6.350 | 2.830 | 2.170 | 1.240 |
| 8 | 20.10 | 15.50 | 13.40 | 7.340 | 3.490 | 2.730 | 1.650 |
| 9 | 21.70 | 16.90 | 14.70 | 8.340 | 4.170 | 3.320 | 2.090 |
| 10 | 23.20 | 18.30 | 16.00 | 9.340 | 4.860 | 3.940 | 2.560 |
| 15 | 30.50 | 25.00 | 22.30 | 14.30 | 8.500 | 7.300 | 5.200 |
| 20 | 37.60 | 31.40 | 28.40 | 19.30 | 12.40 | 10.90 | 8.300 |
| 25 | 44.30 | 37.30 | 34.40 | 24.30 | 16.50 | 14.60 | 11.50 |
| 30 | 50.90 | 43.80 | 40.30 | 29.30 | 20.60 | 18.50 | 15.00 |
|  |  |  |  |  |  |  |  |

## Значение функции Стьюдента

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
| n|x |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| 1 | 0.063 | 0.121 | 0.172 | 0.215 | 0.250 | 0.303 | 0.352 |
| 2 | 0.070 | 0.136 | 0.195 | 0.246 | 0.289 | 0.352 | 0.408 |
| 3 | 0.073 | 0.142 | 0.205 | 0.259 | 0.304 | 0.372 | 0.438 |
| 4 | 0.074 | 0.145 | 0.210 | 0.266 | 0.313 | 0.383 | 0.442 |
| 5 | 0.075 | 0.147 | 0.213 | 0.270 | 0.318 | 0.390 | 0.449 |
| 6 | 0.076 | 0.148 | 0.215 | 0.273 | 0.322 | 0.394 | 0.454 |
| 7 | 0.076 | 0.150 | 0.216 | 0.275 | 0.325 | 0.398 | 0.457 |
| 8 | 0.077 | 0.150 | 0.217 | 0.277 | 0.327 | 0.400 | 0.460 |
| 9 | 0.077 | 0.151 | 0.218 | 0.278 | 0.328 | 0.402 | 0.462 |
|  | 0.077 | 0.151 | 0.219 | 0.279 | 0.330 | 0.404 | 0.463 |
|  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
| n|x |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  | 0.374 | 0.398 | 0.409 | 0.422 | 0.429 | 0.437 | 0.447 |
|  | 0.431 | 0.452 | 0.462 | 0.471 | 0.476 | 0.481 | 0.487 |
|  | 0.452 | 0.471 | 0.479 | 0.486 | 0.489 | 0.492 | 0.495 |
|  | 0.463 | 0.480 | 0.486 | 0.492 | 0.494 | 0.496 | 0.498 |
|  | 0.469 | 0.485 | 0.490 | 0.495 | 0.496 | 0.498 | 0.499 |
|  | 0.473 | 0.488 | 0.493 | 0.496 | 0.498 | 0.499 | 0.500 |
|  | 0.476 | 0.490 | 0.494 | 0.497 | 0.498 | 0.499 | 0.500 |
|  | 0.478 | 0.492 | 0.495 | 0.498 | 0.499 | 0.500 | 0.500 |
|  | 0.480 | 0.492 | 0.496 | 0.498 | 0.499 | 0.500 | 0.500 |
|  | 0.481 | 0.493 | 0.497 | 0.499 | 0.499 | 0.500 | 0.500 |
|  |  |  |  |  |  |  |  |