# [7]. Статистика Манна-Уитни. Математическое ожидание, дисперсия и распределение при гипотезе однородности.

Имеем:

Статистика Манна-Уитни:

где и – ранги для объектов, которые подверглись и не подверглись обработке, соответственно.

В отличие от простой суммы рангов, такая статистика имеет несколько преимуществ:

1. Распределение статистики (при гипотезе однородности) при и будет совпадать:

Просто сумма рангов (при гипотезе однородности) имеет следующее распределение:

где – количество различных комбинаций рангов, образующих сумму .

Получается, что, если поменять группы местами, то нужно заново находить распределение.

В случае статистики Манна-Уитни при любых распределение статистики будет одинаковым, если и в обратном случае. Поэтому можно составить 1 таблицу распределения статистики для пары .

Пример:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Сумма рангов () | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Сумма рангов ( | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Статистика Манна-Уитни | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| Количество комбинаций | 1 | 1 | 2 | 2 | 2 | 1 | 1 |

1. Существует альтернативный способ подсчета:

Значит можно считать статистику сравнивая объекты между собой:

где – индикатор.

Распределение статистики при гипотезе однородности:

где – число различных комбинаций чисел из отрезка , образующих сумму .

При и больших 10 распределение стримится к нормальному (по Предельной Центральной Теореме).

Вероятность индикатора:

* из равновероятностного распределения на отрезке
* Всего комбинаций:
* Комбинаций, в которых :

Математическое ожидание (при гипотезе однородности):

Дисперсия индикатора:

Дисперсия (при гипотезе однородности). Для подсчета нужно рассмотреть разные ситуации:

# [8]. Тест суммы рангов Вилкоксона. Равные и неравные наблюдения.

При использовании теста суммы рангов Вилкоксона необходимо различать различные ситуации:

1. Неравные наблюдения

Все наблюдения различны, поэтому вероятность появления конкретных рангов у объектов, которые подверглись обработке, при гипотезе однородности:

В данном случае можно использовать заранее построенные таблицы распределения статистики.

1. Равные наблюдения

При ситуации, когда характеристики объектов могут повторяться есть 2 метода использования теста суммы рангов Вилкоксона:

* Искусственный порядок

Искусственно определяем порядок и используем метод с неравными наблюдениями.

* Мидранги

Для объектов с равными характеристиками поставим мидранг, равный среднему рангов, которые относятся к данным объектам.

В данном случае комбинации рангов перестают быть равновероятными, поэтому для каждого случая придется считать распределение.

Пример использования мидрангов:

Наблюдения: }

Ранги:

Вероятности комбинаций рангов объектов, которые подверглись обработке, при гипотезе однородности:

Получившееся распределение статистики:

|  |  |  |  |
| --- | --- | --- | --- |
|  | 3.5 | 5 | 6.5 |
|  | 2/6 | 2/6 | 2/6 |

# [9]. Знаково-ранговый тест Вилкоксона, распределение, пример применения.

Знаково-ранговый тест Вилкоксона – это тест для набора парных наблюдений, который учитывает знак и значение разности между характеристиками объектов в паре.

Алгоритм подсчета:

1. Для каждой пары наблюдений находим разницу между характеристиками
2. Ранжируем модули разниц в порядке возрастания
3. Каждому рангу приписываем знак (+ или -) соответствующей разницы
4. Считаем статистику : сумму положительных рангов

Тест:

Имеем пары наблюдений

обработка не имеет влияния, распределения и совпадают

обработка имеет влияние, после нее характеристики больше

При гипотезе однородности знак каждого ранга равен + или – равновероятно (с вероятностью ) и все знаков независимы. Тогда всего имеется возможных комбинаций знаков.

Вероятность конкретной комбинации при гипотезе однородности:

где - количество положительных рангов.

Так как все комбинации равновероятны, то распределение статистики при гипотезе однородности:

где – количество комбинаций, сумма рангов которых равна .

Аналогично можно ввести статистику : сумма рангов со знаком – (сумма самих рангов, т.е. натуральных чисел), т.к. сумма этих статистик:

Решающее правило для такой статистики:

Данные статистики эквивалентны. На практике бывает легче считать какую-то статистику, т.к. она состоит из меньшего числа слагаемых.

Пример использования:

Для проверки нового удобрения 3 поля с клубникой поделили пополам и случайно распределили удобрение (новое и старое) для каждой половины. Взвесив урожай с каждой половины поля, с помощью знаково-рангового теста Вилкоксона можно понять имеет ли новое удобрение лучший эффект.