Статистика 2

**Эмпирическое распределение** – те значения, которые мы получили в нашей выборке

# Расстояние Пирсона / Хи-квадрат:

Расстояние Пирсона показывает на сколько различается эмпирическое распределение от теоретического, насколько наши значения отклоняются от ожидаемых, это и есть **Хи-квадрат Пирсона,** используется для того, что бы проверить, что некоторая случайная величина подчиняется тому или иному закону распределения

O - наше значение

E - ожидаемое значение



X Пирсона



Наблюдаемые частоты номинативного признака

# Распределение Хи-квадрат:

Изображение выглядит как текст

Автоматически созданное описание

В сумме квадратов накапливается маленький остаток со временем и из-за этого центр смещается вправо

Непараметрические критерии

# **Параметрическим критериям проверки гипотез –** критерии, которые предполагают, что выборки, с которыми они имеют дело, взяты из каких-то конкретных распределений и проверяют гипотезы, связанные со значениями параметров этих распределений.

**Непараметрические** – выборка без четкого распределения.

**Нулевое распределение** - распределение при условии справедливости нулевой гипотезы.

# Критерий знаков:

Мы сводим наше распределение к бинумиальному с вероятностью 0.5, путем построения гипотезы.

Для примера можно рассмотреть данные о времени ремонта интернет-оборудования клиентов провайдера Verizon. Выборка состоит из 23 наблюдений, распределение признака не похоже на нормальное. Хочется понять, позволяют ли собранные данные утверждать, что среднее время ремонта составляет больше восьми часов

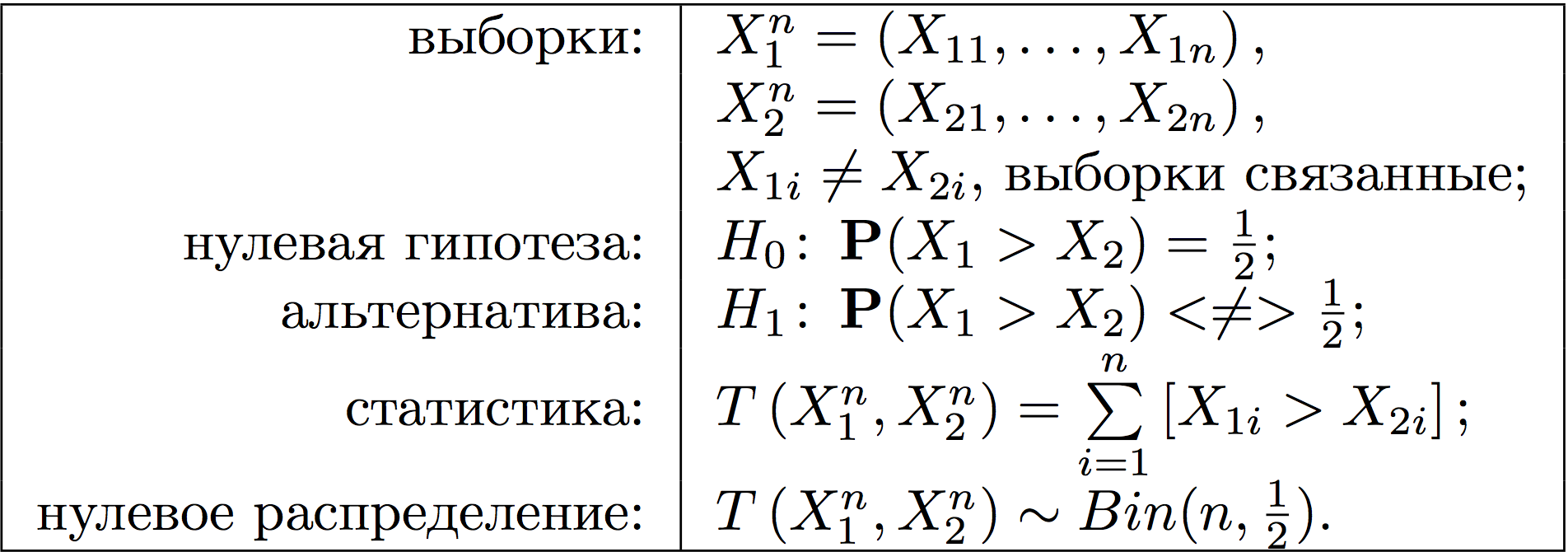
Итак, в задаче о времени ремонта интернет-оборудования проверяется нулевая гипотеза о том, что медиана времени ремонта составляет 8 часов.

Односторонняя альтернатива утверждает, что ремонт в среднем длится дольше 8 часов

В имеющейся выборке ремонт занял больше 8 часов в 15 случаях из 23. Критерий знаков утверждает, что это недостаточно много. Его достигаемый уровень значимости (вероятность получить 15 из 23 в усло- виях справедливости нулевой гипотезы) p = 0.105.

(Мы построили биномиальное распределение с n = 23 и вероятность 0.5. и посмотрели, какой p будет у значения 15. От куда взялось число 15? мы строим гипотезу о том, что медиана = 8 и считаем в скольких случаях, это будет не так и проверяем на сколько это критично)

# Критерий знаков для связанных выборок:



H\_0 говорит о том, что *качество* выборок одинаковое (выборки равны), при условии, что нет советующих одинаковых элементов и если это так, будет верно в 50% случаях, а следовательно его можно проверить биномиальным распределением



# Ранговые критерии:

Для проверки гипотез о средних критерии знаков выбрасывают большую часть информации, содержащуюся в выборке. Вместо исходных значений признака используется бинарный̆ вектор. Ранговые критерии позволяют сохранить больше информации.

Упорядочиваем нашу выборку по возрастанию. Если при этом есть какие-то части вариационного ряда, в которых элементы полностью совпадают, эти части называются «связками».

Рангом наблюдения Xi называется его позиция в вариационном ряду. Если Xi оказывается в связке X(k1), . . . , X(k2), то rank(Xi)= (k1+k2) / 2. то есть в связке все объекты получают одинаковый средний ранг.

# Критерий знаковых рангов Уилкоксона:

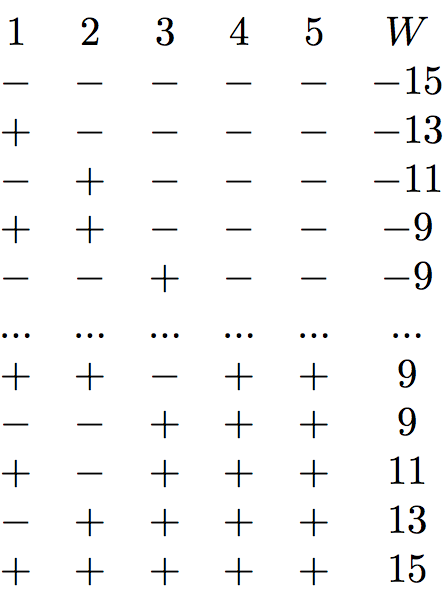
Изображение выглядит как знак, дисплей

Автоматически созданное описание

Использовать ранги можно для решения следующей задачи: имеются 24 шайбы, произведённые на одном и том же конвейере, для которых измерены диаметры. По этой выборке требуется понять, соответствует ли диаметр шайбы стандартному размеру в m0 = 10 мм.

Критерий знаковых рангов даёт достигаемый уровень значимости p = 0.0673, нулевая гипотеза не от- вергается. Выборочная медиана диаметра составляет 10.5 мм, 95% доверительный интервал: [9.95, 11.15] мм. Доверительной интервал содержит целевое значение m0 = 10. Так и должно быть, когда достигаемый уровень значимости выше порога.



Как мы это получили? Если верна нулевая гипотеза, то при сумме всех рангов, мы должны получить 0! (ибо каждое значение мы вычитаем из медианы. мы строим гипотезу о том что 10мм это медиана и вместо нуля получили какое-то число. А дальше смотрим на сколько вероятно получить такое число по нормальному распределению.



# Двуxвыборочная задача со связанными выборками:

Как и до этого в курсе, двухвыборочная задача со связанными выборками решается с использованием того же самого критерия, что и одновыборочная.

# 

# Двухвыборочная задача с независимыми выборками:

В этой задаче измеряемый признак — это респираторный обмен, соотношение числа молекул углекислого газа и кислорода в выдыхаемом воздухе. Респираторный обмен является косвенным признаком того, из чего в данный момент мышцы вырабатывают энергию, из жиров или углеводов. В эксперименте измеряется респираторный обмен у 18 испытуемых в процессе физических упражнений. За час до этого 9 из них получили таблетку кофеина, а оставшиеся 9 — таблетку плацебо. Хочется понять, повлиял ли кофеин на среднее значение показателей респираторного обмена.

Изображение выглядит как снимок экрана

Автоматически созданное описание

Смотри в конспект, там понятно