1 Статистика (Теория)

Проверка статистических гипотез

- X - полученная в эксперименте выборка
- $\mathbb{F}$ - множество априори допустимых распределений X
- $F \in \mathbb{F}$ - истинное распределение X X
- Нулевая гипотеза H : $F \in \mathbb{F} \subset \mathbb{F}$ 0X0
- АльтернативаH :$F \in \mathbb{F}$ =$\mathbb{F} \backslash \mathbb{F}$

Определение 1.
Простая гипотеза (или альтернатива) – такая, при которой множество $\mathbb{F}0$ (или $\mathbb{F}1$) содержит единственный элемент. Иначе гипотеза (или альтернатива) называется сложной.
Если мы предположим, что $\mathbb{F}$ – единственный элемент, то это означает, что F 1X мы сравниваем с единственным элементом. Соответствующая гипотеза полностью определяет распределения вероятностей, с которыми производится работа.

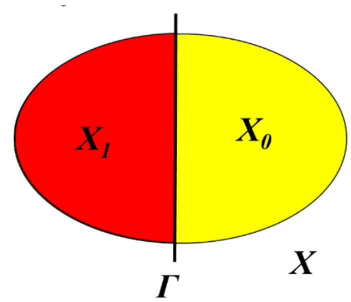
Определение 2.
Статистический критерий – правило, согласно которому, наблюдая X, можно принять решение об отклонении гипотезы H0.
Мы никогда не можем ничего подтвердить с помощью статистики, но некоторые факты можно опровергнуть. Можно сказать, что мы принимаем некоторую гипотезу H0 и оказывается, что данные, полученные в эксперименте, плохо соответствуют этой гипотезе.

Общий принцип - если при справедливости гипотезы H0 наблюдаемое событие (выборка X) маловероятно, то такую гипотезу следуетотклонить.

Определение 3.

Критическая область – множество выборок $\mathbb{X}1$, для которых гипотеза H0 отклоняется.

		Истина	
		Нет	Да
Отклонить?	Нет	Ошибка 2 рода β	Правильное не отклонение
	Да	Правильное отклонение	Ошибка 1 рода α



Ошибка 1 рода – гипотеза отклоняется, а на самом деле она была верна.

Ошибка 2 рода – гипотеза подтверждается, а на самом деле она была не верна.

Линию Г (гамма) необходимо провести таким образом, чтобы минимизировать ошибки α и β.

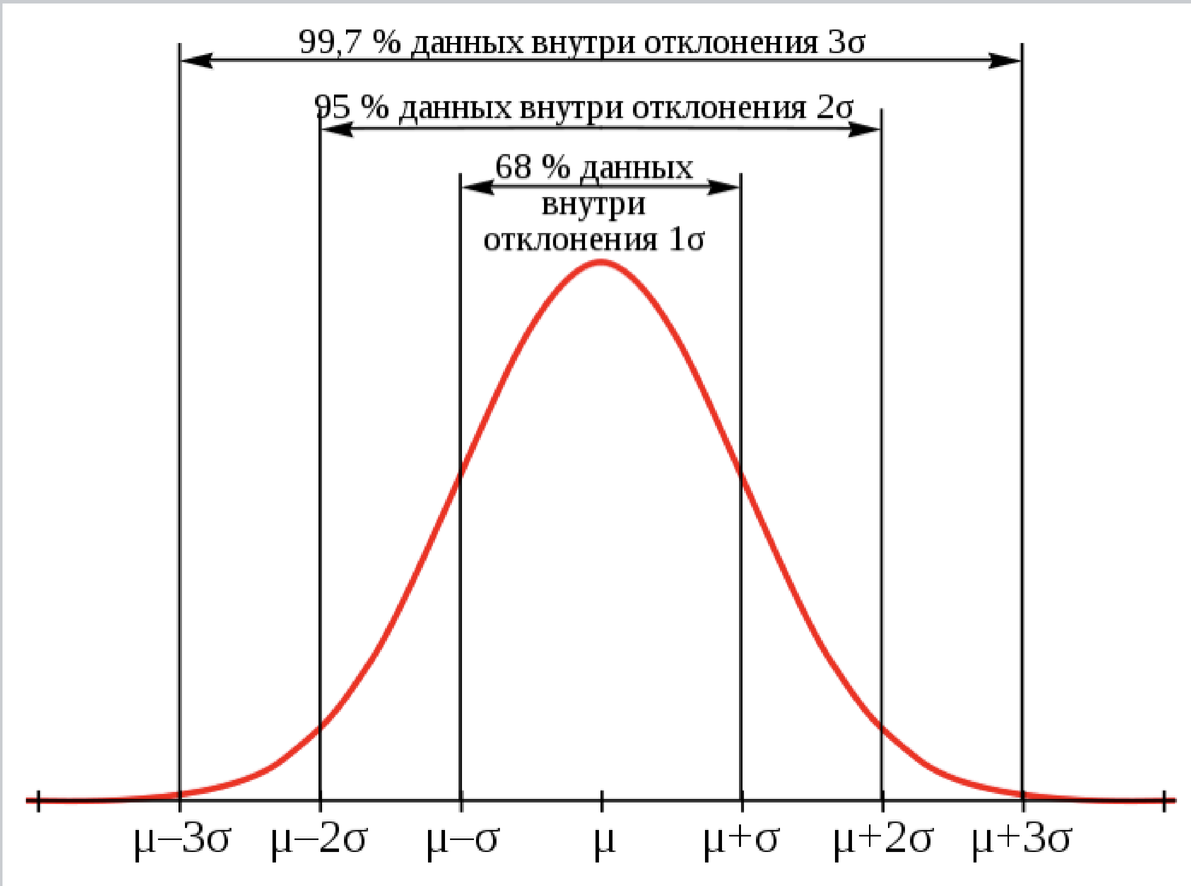
~~1. Если всегда отклонять $H0$, то получится β = 0; α > 0~~
~~2. Если никогда не отклонять $H0$, то получится α = 0; β > 0~~
3. Пусть ошибка 1 рода не превышает α, а ошибка 2 рода будет минимальная - критерий Неймана-Пирсона
4. Пусть "цена" ошибок 1 и 2 рода будет в среднем минимальная - байесовский критерий

- Если «цена» ошибок неизвестна, но заданы их вероятности – критерий Котельникова (апостериорной плотности распределения)
- Если неизвестны ни «цена», ни вероятность ошибок – критерий максимального правдоподобия

Про нормальное распределение

1. **Центральная предельная теорема (ЦПТ)**:

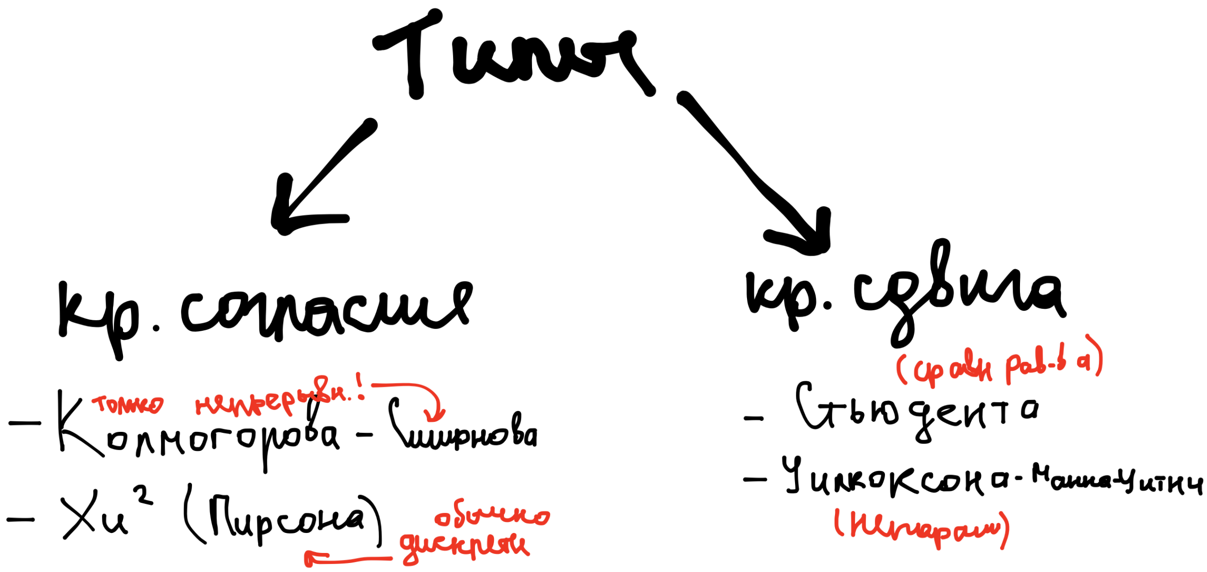
ЦПТ утверждает, что сумма достаточно большого количества слабо зависимых случайных величин, имеющих примерно одинаковые масштабы (ни одно из слагаемых не доминирует, не вносит в сумму определяющего вклада), имеет распределение, близкое к нормальному. Это объясняет, почему нормальное распределение так широко встречается в реальных данных.



Правило 68-95-99,7.
Для нормального распределения количество значений, отличающихся от среднего на число, меньшее чем одно стандартное отклонение, составляют 68,27 % выборок. В то же время количество значений, отличающиеся от среднего на два стандартных отклонения, составляют 95,45 %, а на три стандартных отклонения — 99,73 %.

Типы статистических критериев

В зависимости от проверяемой *нулевой гипотезы* статистические критерии делятся на группы.



Критерии согласия

Критерии согласия проверяют, согласуется ли заданная выборка с заданным фиксированным распределением, с заданным параметрическим семейством распределений, или с другой выборкой.

Критерии сдвига

Специальный случай двухвыборочных критериев согласия. Проверяется гипотеза сдвига, согласно которой распределения двух выборок имеют одинаковую форму и отличаются только сдвигом на константу.