Пусть X1, … , Xn – выборка из распределения F(theta), theta in THETA

(THETA – множество возможных значений параметра),

T1(X1, … , Xn), T2(X1, … , Xn) – статистики, такие, что:

P(T1(X1, … , Xn) < theta < T2(X1, … , Xn))=1-alfa

Тогда (T1(X1, … , Xn), T2(X1, … , Xn)) – доверительный интервал

Надежность: (1-alfa)

Точность: (Т2(X1, … , Xn) – Т1(X1, … , Xn))

1)

2)

3) – независимы

Следствие:

На этой неделе мы с вами познакомимся с понятием Доверительных интервалов, напишем программу для их вычисления на языке программирования Питон и решим несколько задачек.

Пусть нам дана выборка из некоего распределения F, которое зависит от неизвестного нам параметра theta, истинное значение которого накрывается интервалом от статистики T1 до статистики T2 с некоторой вероятностью (1-alfa). Тогда такой интервал (T1, T2) называется доверительным, а (1-alfa) есть уровень его доверия (выбирается исследователем), статистика T1 и T2 – нижняя и верхняя доверительные границы соответственно.

Надежность интервала характеризуется уровнем доверия.

Точность – длинной самого интервала (мы заинтересованы в том, чтобы он был как можно уже).

Точность и надежность имеют обратную зависимость, то есть чем большое первая, тем меньше вторая и наоборот.

Следующие утверждения будем использовать без оказательства, они следуют из леммы Фишера.

Пусть X1, … , Xn – выборка из нормального распределения N(mu, sigma\*\*2)

1. ДИ для матожидания при известной дисперсии

sigma\*\*2 – известна

P(T1< mu < T2) = 1-alfa

Изображение выглядит как текст, карта

Автоматически созданное описание

Границы интервала – **случайные величины.**

Подставив числа, получим **реализацию** ДИ.

Для начала рассмотрим ДИ для параметров нормального распределения.

Пусть нам дана выборка из нормального распределения с параметрами mu и sigma \*\*2.

Если нам известна дисперсия, но неизвестно матожидание, то нужно найти такие статистики Т1 и T2, которые накроют значение mu с некоторой вероятностью (1-alfa).

Для этого возьмем известную нам оценку mu – выборочное среднее. Из леммы Фишера имеем, что стандартизированное среднее имеет стандартное нормальное распределение.

Это значит, что если мы фиксируем уровень доверия (1-alfa), то можно изобразить график распределения, где вероятность попасть в центральную область равна (1- alfa), а значения +/– Z(alfa/2) (которые можно найти в таблице стандартного нормального распределения) показывают значения, вероятности попасть левее или правее которых равны по (alfa/2) соответственно.

Из общей формулы для ДИ имеем вероятность попадания стандартизированного среднего.

Преобразуем выражение и получим ДИ для mu.

Получившиеся границы нашего интервала – **случайные величины.**

Когда у нас есть конкретная реализация выборки и мы подставляем значения в формулу, то получаем **реализацию этого интервала**, где границы уже не показывают вероятность, а параметр может как находиться, так и не находиться внутри конкретной реализации ДИ.

---- Тестовое задание (1) -------

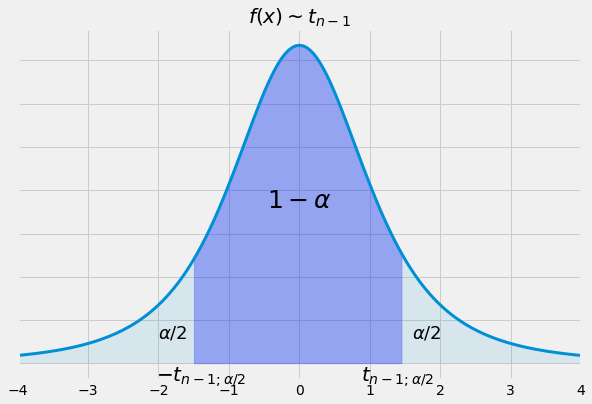
1. ДИ для матожидания при неизвестной дисперсии

sigma\*\*2 – неизвестна

P(T1< mu < T2) = 1-alfa

,

где



– квантиль распределения Стьюдента уровня (1-alfa/2) с (n-1) степенью свободы

Если нам неизвестны матожидание и дисперсия, то нужно найти такие статистики Т1 и T2, которые накроют значение mu с некоторой вероятностью (1-alfa).

Для этого воспользуемся следствием из леммы Фишера. Заметим, что все рассуждения аналогичны предыдущим, меняется только распределение и оценка, для которой мы будем искать ДИ.

Возьмем известную оценку sigma\*\*2 – выборочную дисперсию, которая находится по формуле.

Если зафиксировать уровень доверия (1-alfa), то аналогично предыдущему примеру значения +/–t(n-1)(a/2) такие, что вероятности попасть левее или правее них равны по (alfa/2) соответственно.

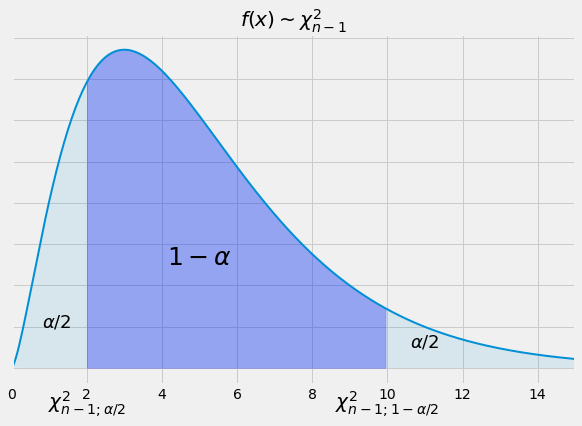
Из общей формулы для ДИ имеем вероятность попадания величины с распределением Стьюдента.

Преобразуем выражение и получим ДИ для mu.

----- Тестовое задание (2) ------

1. ДИ для дисперсии

P(T1 < sigma\*\*2 < T2) = 1-alfa



– квантиль распределения Хи-квадрат уровня (alfa/2) с (n-1) степенью свободы

– квантиль распределения Хи-квадрат уровня (1-alfa/2) с (n-1) степенью свободы

Если нам известно матожидание, но неизвестна дисперсия, то нужно найти такие статистики Т1 и T2, которые накроют значение sigma\*\*2 с некоторой вероятностью (1-alfa).

Для этого воспользуемся следствием из леммы Фишера.

Заметим, что распределение Хи-квадрат имеет несимметричный график, значит нам нужно брать отдельно значения для левой и нижней границы (в таблице). Соответственно Хи-квадрат для alfa/2 и для 1- alfa/2.

Из общей формулы для ДИ имеем вероятность попадания величины с распределением Стьюдента.

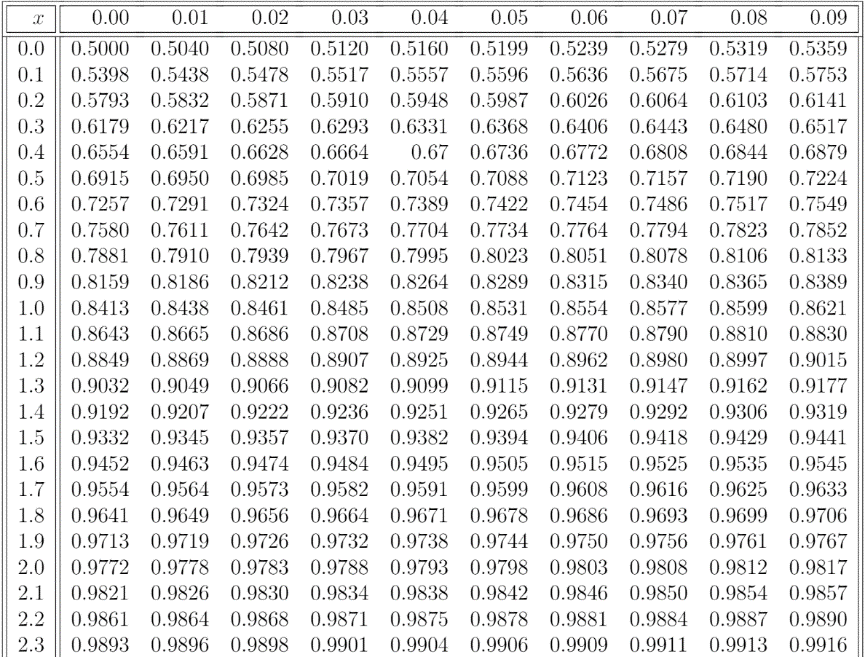
Преобразуем выражение и получим ДИ для sigma\*\*2.

---- Тестовое задание (3) -------

Тестовые задания:

1. Для уровня доверия равного 0.95 чему будет равно значение Z(a/2)? Где Z – стандартная нормальная величина.

(0.8289, 0.835, 1.64, 1.96)



1. Пусть дана выборка из 10 значений некоторой величины X, дисперсия неизвестна. Какую величину нужно использовать для нахождения 95% ДИ для матожидания X (mu)?

(N(0,1)(0.975), t(9, 0.975), t(10, 0.95), t(9, 0.05))

1. Пусть дана выборка из 10 значений некоторой величины X, дисперсия неизвестна. Какую(ие) величину(ы) нужно использовать для нахождения 95% ДИ для дисперсии X (sigma\*\*2)?

(t(9, 0.975), chi(9, 0.95) & -chi(9, 0.95), chi(9, 0.975) & chi(9, 0.025), chi(10, 0.95) & chi(10, 0.5))