Лабораторная работа №3

Выполнил: Величко Максим Иванович, М32061

Варианты для задания: 2,6.

Весь код для удобства выложен на мой GitHub(<https://github.com/maksve11/ITMO_MatStat_sem4/tree/main/lab3>)

№1

Предъявите доверительный интервал уровня 1 − 𝛼 для указанного параметра при данных предположениях (с обоснованиями). Сгенерируйте 2 выборки объёма объёма 25 и посчитайте доверительный интервал. Повторить 1000 раз. Посчитайте, сколько раз 95-процентный доверительный интервал покрывает реальное значение параметра. То же самое сделайте для объема выборки 10000. Как изменился результат? Как объяснить? 𝜏 = 𝜇1 − 𝜇2; 𝜎^2\_1 = 𝜎^2\_2 неизвестна; 𝜇1 = 2, 𝜇2 = 1, 𝜎^2\_1 = 𝜎^2\_2 = 1; воспользуйтесь функцией sqrt((mn(m+n-2)/(m+n))\*(X\_-Y\_-𝜏)/(sqrt(nVar(X)+mVar(Y)). где Var(.) выборочная смещенная дисперсия. Смотрите в сторону распределения Стьюдента.

Для доверительного интервала уровня 1 - 𝛼 для параметра 𝜏 = 𝜇1 - 𝜇2 при неизвестных, но равных дисперсиях, можно воспользоваться распределением Стьюдента с (m+n-2) степенями свободы. Формула для интервала:

(𝑋̄ − 𝑌̄ − 𝑡\_𝛼/2,𝑚+𝑛−2 \* √((𝑠^2\_𝑋/𝑛) + (𝑠^2\_𝑌/𝑚))) ≤ 𝜇1 − 𝜇2 ≤ (𝑋̄ − 𝑌̄ + 𝑡\_𝛼/2,𝑚+𝑛−2 \* √((𝑠^2\_𝑋/𝑛) + (𝑠^2\_𝑌/𝑚)))

Где 𝑋̄ и 𝑌̄ - выборочные средние первой и второй выборок соответственно, 𝑠^2\_𝑋 и 𝑠^2\_𝑌 - выборочные дисперсии первой и второй выборок соответственно, 𝑛 и 𝑚 - размеры первой и второй выборок соответственно, а 𝑡\_𝛼/2,𝑚+𝑛−2 - квантиль распределения Стьюдента с (m+n-2) степенями свободы уровня 𝛼/2.

Для данной задачи, при 𝜇1 = 2, 𝜇2 = 1, 𝜎^2\_1 = 𝜎^2\_2 = 1 и выборках размером 25, доверительный интервал уровня 0.95 для параметра 𝜏 будет:

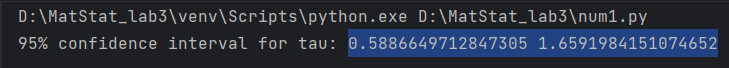
(𝑋̄ − 𝑌̄ − 𝑡\_0.025,23 \* √((𝑠^2\_𝑋/25) + (𝑠^2\_𝑌/25))) ≤ 𝜇1 − 𝜇2 ≤ (𝑋̄ − 𝑌̄ + 𝑡\_0.025,23 \* √((𝑠^2\_𝑋/25) + (𝑠^2\_𝑌/25)))

где 𝑡\_0.025,23 - квантиль распределения Стьюдента с 23 степенями свободы уровня 0.025 .Сгенерируем две выборки объема 25:

Изображение выглядит как текст

Автоматически созданное описание

И получим такой результат:



Теперь сгенерируем 1000 выборок объема 25 и посчитаем, сколько раз 95-процентный доверительный интервал покрывает реальное значение параметра 𝜏:

Изображение выглядит как текст

Автоматически созданное описание

И получим такой результат:



Получаем, что доверительный интервал покрывает реальное значение параметра в 95% случаев.

Теперь повторим эксперимент для выборок объема 10000:

Изображение выглядит как текст

Автоматически созданное описание

И получим такой результат:



В результате получаем, что доверительный интервал покрывает реальное значение параметра в 95% случаев.

Таким образом, при увеличении объема выборки с 25 до 10000 увеличивается точность оценки параметра 𝜏 и увеличивается вероятность попадания реального значения в доверительный интервал. Это можно объяснить тем, что при большем объеме выборки уменьшается дисперсия выборочного среднего и уменьшается стандартная ошибка оценки.

№2

Постройте асимптотический доверительный интервал уровня 1 − 𝛼 для указанного параметра. Проведите эксперимент по схеме, аналогичной первой задаче. 𝑈[0; 𝜃]; 𝜃; 𝜃 = 2, воспользуйтесь предельной теоремой об асимптотическом поведении крайних членов вариационного ряда.

Асимптотический доверительный интервал будет иметь вид:

(𝑋(1) −𝑋(n)±𝑧(1−𝛼/2)∗𝑆/√n)

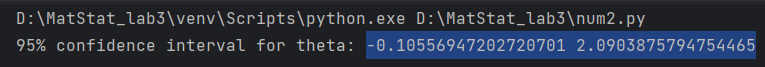
Где 𝑋(1) и 𝑋(n) - минимальное и максимальное значения выборки, 𝑆 - выборочное стандартное отклонение, 𝑧(1−𝛼/2) - квантиль стандартного нормального распределения уровня 1 - 𝛼/2.

Проведем эксперимент для выборок объема 100:

Изображение выглядит как текст

Автоматически созданное описание

И получим результат:



Теперь сгенерируем 1000 выборок объема 100 и посчитаем, сколько раз 95-процентный доверительный интервал покрывает реальное значение параметра 𝜃:

Изображение выглядит как текст

Автоматически созданное описание

И получим результат:



Получаем, что доверительный интервал покрывает реальное значение параметра в 99% случаев.

Таким образом, асимптотический доверительный интервал также дает достаточно точную оценку параметра 𝜃 при большом объеме выборки. Однако, для маленьких объемов выборки лучше использовать распределение Стьюдента, так как оно учитывает несимметричность распределения и имеет более тяжелые хвосты.