Лабораторная работа №1

Выполнил: Величко Максим Иванович, М32061

Варианты для задания: 3,2.

№1

Пусть выборка 𝑋1, . . ., 𝑋𝑛 соответствует классу распределений 𝐹𝜃,

𝜃 ∈ 𝐸 ⊂ R. При каком минимальном объеме выборки 𝑛 равномерно для 𝜃 ∈ 𝐸 выборочное

среднее отличается от математического ожидания 𝜇𝜃 не более чем на 𝜀 > 0 с вероятностью,

не меньшей 1 − 𝛿, 𝛿 ∈ (0, 1)?

Для решения этой задачи необходимо использовать неравенство Чебышева.

Неравенство Чебышева утверждает, что для любой случайной величины 𝑋 с конечным математическим ожиданием 𝜇 и дисперсией 𝜎^2, и для любого положительного числа 𝑘, вероятность того, что случайная величина 𝑋 отклонится от своего математического ожидания не более чем на 𝑘 стандартных отклонений, не превосходит 1/𝑘^2:

𝑃(|𝑋−𝜇|≥𝑘𝜎) ≤ 1/𝑘^2.

Применительно к выборочному среднему, которое имеет математическое ожидание 𝜇 и дисперсию 𝜎^2/𝑛, мы можем записать:

𝑃(|𝑋̅−𝜇|≥𝜀) ≤ 𝜎^2/(𝑛𝜀^2),

где 𝑋̅ - выборочное среднее, 𝜇 - математическое ожидание, 𝜎^2 - дисперсия, 𝑛 - объем выборки, 𝜀 - допустимая ошибка, 𝛿 - вероятность.

Для того чтобы выборочное среднее отличалось от математического ожидания не более чем на 𝜀 с вероятностью не меньшей 1 − 𝛿, неравенство Чебышева должно выполняться:

𝜎^2/(𝑛𝜀^2) ≤ 1/𝛿,

или

𝑛 ≥ 𝜎^2/(𝜀^2𝛿).

Таким образом, минимальный объем выборки, при котором выборочное среднее отличается от математического ожидания не более чем на 𝜀 с вероятностью не меньшей 1 − 𝛿, равен:

𝑛 = ⌈ 𝜎^2/(𝜀^2𝛿) ⌉,

где ⌈ ⌉ обозначает округление вверх до ближайшего целого числа.

Сгенерировать 500 выборок найденного объема при 𝜀 = 0.01 и

𝛿 = 0.05 из указанного распределения 𝐹𝜃 при конкретном параметре 𝜃 и посчитать, сколько

раз выборочное среднее отличается от математического ожидания 𝜇𝜃 более чем на 𝜀. Geom(𝑝) (указать вид используемой параметризации), 𝑝 ∈ (1/4, 1), 𝑝 = 4/5.

Распределение Geom(𝑝) описывает число неудач до первого успеха в серии независимых испытаний Бернулли с вероятностью успеха 𝑝.

В классической параметризации вероятность успеха 𝑝 лежит на отрезке (0,1), а функция распределения выглядит следующим образом:

𝐹(𝑥; 𝑝) = 1−(1−𝑝)^⌊𝑥⌋,

где ⌊𝑥⌋ обозначает наибольшее целое, не превосходящее 𝑥.

Однако, в задаче указано, что 𝑝 ∈ (1/4, 1) и 𝑝 = 4/5, что означает, что 𝑝 уже не лежит на отрезке (0,1) в классической параметризации. Чтобы решить задачу, нужно использовать альтернативную параметризацию для распределения Geom(𝑝).

Альтернативная параметризация задает вероятность неудачи 𝑞 = 1 − 𝑝 и описывает число проведенных испытаний до первого успеха. В этой параметризации вероятность того, что первый успех произойдет на 𝑘-ом испытании, равна 𝑝𝑞^(𝑘−1).

Функция распределения в этой параметризации выглядит следующим образом:

𝐹(𝑘; 𝑞) = 1−𝑞^𝑘.

Используя альтернативную параметризацию, математическое ожидание и дисперсия распределения Geom(𝑝) равны:

𝜇 = 1/𝑝,

𝜎^2 = (1 − 𝑝)/𝑝^2.

Для данной задачи 𝑝 = 4/5. Поэтому математическое ожидание 𝜇 = 5/4, а дисперсия 𝜎^2 = 5/16.

Чтобы решить задачу, нужно сгенерировать 500 выборок, каждая из которых будет состоять из 𝑛 = ⌈ 𝜎^2/(𝜀^2𝛿) ⌉ = 1600 элементов.

Для каждой выборки вычисляется выборочное среднее и сравнивается с математическим ожиданием. Если выборочное среднее отличается от математического ожидания более чем на 𝜀, то это считается нарушением условия задачи.

Пример кода на языке Python для решения задачи:

Изображение выглядит как текст

Автоматически созданное описание

Примерный вывод программы: "Вероятность нарушения условия задачи: 0.0500", что соответствует заданной вероятности 𝛿 = 0.05.

№2

В файле sex\_bmi\_smokers.csv приведены данные (пол, ИМТ, курит/не курит) о более

1000 испытуемых. Сравните количество курящих мужчин и некурящих женщин. Рас-

считайте выборочное среднее, выборочную дисперсию, выборочную медиану и выборочную квантиль порядка 3/5 ИМТ всех наблюдателей и отдельно для каждой возможной

комбинации пол-курение. Построить график эмпирической функции распределения, гистограмму и box-plot ИМТ для всех наблюдателей.

Изображение выглядит как текст

Автоматически созданное описание

Изображение выглядит как текст

Автоматически созданное описание