

Індивідуальне завдання №2

Постановка задачі

- Зчитати дані з текстового файлу, побудувати полігон або гістограму частот;
- на основі графічного представлення сформулювати гіпотезу про закон розподілу досліджуваної ознаки генеральної сукупності.
- передбачити можливість користувачу задати параметри розподілу вручну або оцінити на основі даних вибірки;
- для заданого користувачем рівня значущості перевірити сформульовану гіпотезу за критерієм χ^2 .

Короткі теоретичні відомості

Однією з найбільш важливих задач математичної статистики є задача про визначення закону розподілу ймовірностей випадкової величини (ознаки генеральної сукупності) за даними вибірки.

Якщо закон розподілу випадкової величини невідомий, то формують нульову гіпотезу про вигляд густини розподілу. Наприклад: „випадкова величина має густину нормального розподілу ймовірностей”.

Для перевірки таких гіпотез часто застосовують критерій „хі-квадрат” Пірсона, який ґрунтується на визначенні відхилення емпіричних характеристик від гіпотетичних характеристик.

Для їх знаходження потрібно знайти χ^2 критичне, яке вираховується за наступною формулами:

$$K = \sum_{i=1}^m \frac{(n_i - np_i)^2}{np_i}$$

А також χ^2 емпіричне, значення якого можна знайти з таблиці, перед тим знайшовши альфа (задається користувачем) та ступінь вільності, який вираховується за формулою

$$d.f = r - s, \text{ де } (r+1) - \text{к-сть інтервалів, а } s - \text{к-сть невідомих параметрів}$$

Кожен з розподілів має свої невідомі параметри, а саме:

- 1) Біномний розподіл – p
- 2) Розподіл Пуассона – λ
- 3) Рівномірний розподіл – a, b, σ

4) Показниковий розподіл - λ

5) Нормальний розподіл – μ, σ

Також варто зауважити, що після розрахунку $P(i)$ та $n \cdot P(i)$ потрібно перевірити вибірку на дві умови:

1) $n \geq 5$

2) $n \cdot P(i) > 10$

Якщо одна з умов не виконується, то відбувається з'єднання стовпців за наступним алгоритмом: два стовпці, де один з яких не відповідає умовам з'єднується з будь яким іншим, при тому усі їхні параметри додаються. Після цього ксі емпіричне вираховується вже по новій таблиці.

Програмна реалізація

Програма реалізована мовою Python, крім того я використав бібліотеку tkinter для створення графічного інтерфейсу. Програма зчитує вхідні дані з файлу формату txt, для зчитування використовуються дві функції, а сама для зчитування дискретних та неперервних даних. Далі формується полігон частот для дискретних або гістограма для неперервних даних відповідно. Далі програма перевіряє чи виконується умова, що $N_i \geq 5$, якщо ні, то проводиться необхідна кількість злиттів необхідних стовпців. Далі програма має розгалуження відповідно до задачі, а саме: у першій задачі я перевіряю на нормальний закон розподілу, отже спочатку програма обраховує середнє, девіацію, та знаходить параметри розподілу (хочу зауважити, що це обраховується лише у випадку, коли користувач не задав наперед параметри розподілу вручну). У другій задачі на закон розподілу Пуассона, де аналогічно обраховується параметр розподілу лямбда. Після знаходження параметрів розподілу знаходиться теоретичні ймовірності P_i та $N \cdot P_i$, проводиться перевірка чи $N \cdot P_i > 10$, якщо ні то проводиться злиття необхідних стовпців. І в останню чергу обраховується χ^2 критичне та χ^2 емпіричне

Висновки

Я реалізував програму, що зчитує дані з текстового файлу (як дискретні так і неперервні), перевіряє задану гіпотезу про закон розподілу досліджуваної ознаки генеральної сукупності, а саме нормальний закон у першій та закон розподілу Пуассона 2 другій задачі. Крім того, у програмі я передбачив можливість користувачу задати параметри розподілу вручну або оцінити на основі даних вибірки. для заданого користувачем рівня значущості перевіряю сформульовану гіпотезу за критерієм χ^2 .

Під час опрацювання теоретичного матеріалу для індивідуального завдання та під час реалізації програми я навчився перевіряти гіпотези про закону розподілу ймовірностей випадкової величини (ознаки генеральної сукупності)

за даними вибірки за допомогою критерія Пірсона. Крім того перевіряв гіпотези про такі закони розподілу: біномний закон розподілу, закон розподілу Пуассона, рівномірний закон розподілу, показниковий закон розподілу, нормальний закон розподілу.

Моя програма може знайти практичне застосування у навчальних цілях та у приватному секторі, оскільки user interface робить її використання простим, а мова python дану програму дозволяє використовувати на будь-яких платформах.