**Практична робота № 6**

**(4 години)**

***Тема:*** Квантитативні дослідження мови та мовлення на прикладі україномовних текстів засобами мови Python.

***Мета:*** Отримати практичні навички квантитативного дослідження текстової інформації на основі отримання кількісних і статистичних даних.

**Зміст роботи:**

У звіті відобразити завдання, код, результат.

Результати роботи кожного завдання мають виводитись на екран з текстовими поясненнями.

На перевірку надіслати звіт, текстовий файл і файл .py

**Завдання 1.**

Використовуючи результати практичної роботи № 5 отримати квантитатичні показники і представити їх у вигляді таблиць. (приклад Таблиця 1,2). Статистичні розрахунки виконати засобами мови програмування Python.

Для виконання завдання необхідно:

1. Отримати варіаційні ряди даних
* З тексту видалити всі розділові знаки.
* Створити список слів.
* Підрахувати кількість букв у кожному слові.
* Створити множину можливих варіантів кількості букв.
* Розрахувати скільки слів в тексті з такою кількістю букв (відповідно до множини варіантів).
* Розрахувати відносне значення кількості букв у тексті.
1. На основі отриманих даних побудувати таблицю за зразком Таблиці 1. (таблицю можна створювати в текстовому редакторі).
2. На основі отриманих варіаційних рядів провести розрахунки за формулами. Результати оформити у вигляді Таблиці 2.

***Розрахунки провести за формулами:***

Середнє значення обчислюється за формулою:

$$\overbar{Х}=\frac{\sum\_{}^{}х\_{і}ni\_{і} }{\sum\_{}^{}n\_{і}}$$

де хі – значення варіанта, ni – кількість появ варіанта у досліді, і – номер варіанта.

Середнє квадратичне відхилення розраховано за формулою:

$$σ=\sqrt{\frac{\sum\_{}^{}(х\_{і}-\overbar{х)}^{2}n\_{i}}{\sum\_{}^{}n\_{i}}}$$

де xi – варіанта, ni – кількість появ варіант у дослідах, i – номер варіанти $\overbar{х}$– середнє значення.

Міру коливання середньої частоти визначали за формулою:

$$σ\_{\overbar{x}}=\frac{σ}{\sqrt{\sum\_{}^{}n\_{i}}}$$

де $σ$ – середнє квадратичне відхилення, ni – кількість появ варіант у дослідах, i – номер варіанти, $\overbar{х}$– середнє значення.

 Стандартну та відносну похибку, відповідно, за формулами:

$$S\_{\overbar{x}}=\frac{σ}{\sqrt{\sum\_{}^{}n\_{i}-1}} ϵ=\frac{1.96σ\_{\overbar{x}}}{\overbar{x}}$$

де $σ$ – середнє квадратичне відхилення, ni – кількість появ варіант у дослідах, i – номер варіанти, $\overbar{х}$– середнє значення, $σ\_{\overbar{х}}$ – міра коливання середньої частоти.

**Розподіл довжини слів твору** ……

Таблиця 1.

|  |  |  |  |
| --- | --- | --- | --- |
| Довжина слова | Кількість | Довжина слова | Кількість |
| абсолютна | відносна (%) | абсолютна | відносна (%) |
|  | 8993 | 8,468 |  | 482 | 0,458 |
|  | 15038 | 14,303 |  | 133 | 0,126 |
|  | 8454 | 8,041 |  | 55 | 0,052 |
|  | 13088 | 12,448 |  | 28 | 0,027 |
|  | 12743 | 12,12 |  | 30 | 0,029 |
|  | 11918 | 11,335 |  | 22 | 0,21 |
|  | 10867 | 10,336 |  | 6 | 0,006 |
|  | 8737 | 8,31 |  | 10 | 0,01 |
|  | 6114 | 5,815 |  | 5 | 0,005 |
|  | 3815 | 3,628 |  | 2 | 0,002 |
|  | 2515 | 2,392 |  | 2 | 0,002 |
|  | 1350 | 1,284 |  | 1 | 0,001 |
|  | 823 | 0,783 |  | 1 | 0,001 |

**Порівняльна таблиця довжини слова у творі** ……

Таблиця 2.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Автор  | Середнє значення | Середнє квадратичне відхилення | Міра коливання середньої частоти | Стандартна похибка | Відносна похибка |
| А. Головко | 4,74 | 0,10 |  |  | 0,34 |

**Контрольні запитання.**

1. Як створити текстову змінну вPython на основі текстового файлу?
2. Поясніть в чому полягає відмінність рядків та списків.
3. Як здійснити доступ до елементів рядка, списку?
4. Назвіть та поясніть методи роботи зі списками.
5. Назвіть та поясніть методи роботи з рядками.
6. Які операції можна виконувати над рядками ?