**Lucrare de laborator nr.1**

**la disciplina**

**Analiza statistică și vizualizarea datelor**

**Tema:** Analiza datelor statistice și vizualizarea tendințelor

În tabelul ce urmează este prezentat un set de date despre vârsta media a bărbaților atât în mediul rural cât și în cel urban, în perioada anilor 2019-2023.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| Nr. | **Anii** | **Mediul** | **Speranța de viață** |
| 1. | 2019 | Urban | 68,7 |
| 2. | 2019 | Rural | 65,7 |
| 3. | 2020 | Urban | 66,9 |
| 4. | 2020 | Rural | 65,3 |
| 5. | 2021 | Urban | 65,7 |
| 6. | 2021 | Rural | 64,7 |
| 7. | 2022 | Urban | 68,2 |
| 8. | 2022 | Rural | 66,2 |
| 9. | 2023 | Urban | 69, X |
| 10. | 2023 | Rural | 66,4 |

**X reprezintă prima cifră conform registrului grupei**

**Sarcina:**

**În limbajul R:**

1. Să se calculeze următoarele măsuri descriptive: media, mediana, abaterea standard, valoarea maximă și valoarea minimă.
2. Să se construiască o histogramă pentru distribuția vârstei în mediul urban și rural.
3. Identificați valorile extreme utilizând diagrama boxplot.
4. Să se verifice dacă datele sunt distribuite normal , folosind testul Shapiro-Wilk.
5. Folosind testul t-Student, să se compare media speranței de viață între mediul urban și rural.
6. Calculați corelația Pearson pentru ani și speranța de viață pentru a examina relația dintre ani și evoluția speranței de viață.
7. Organizați datele în intervale și calculați frecvențele.
8. Construiți un grafic de linii care să arate evoluția speranței de viață pe parcursul anilor.
9. Concluzii