**МИНОБРНАУКИ РОССИИ**

**Санкт-Петербургский государственный**

**электротехнический университет**

**«ЛЭТИ» им. В.И. Ульянова (Ленина)**

**Кафедра математического обеспечения и применения ЭВМ**

отчет

**по лабораторной работе №1**

**по дисциплине «Статистические методы обработки экспериментальных данных»**

Тема: **Формирование и первичная обработка выборки. Ранжированный и интервальный ряды.**

|  |  |  |
| --- | --- | --- |
| Студент гр. 8383 |  | Киреев К.А. |
| Студент гр. 8383 |  | Муковский Д.В. |
| Преподаватель |  | Середа А.-В.И. |

Санкт-Петербург

2022

**Цель работы**

Ознакомление с основными правилами формирования выборки и подготовки выборочных данных к статистическому анализу.

**Основные теоретические положения**

Ранжированный ряд– это распределение отдельных единиц совокупности в порядке возрастания или убывания исследуемого признака. Ранжирование позволяет легко разделить количественные данные по группам, сразу обнаружить наименьшее и наибольшее значения признака, выделить значения, которые чаще всего повторяются. Вариационный ряд– последовательность значений заданной выборки , расположенных в порядке неубывания:

Интервальный ряд распределения – это таблица, состоящая из двух столбцов (строк) – интервалов варьирующего признака и числа единиц совокупности, попадающих в данный интервал (частот - ), или долей этого числа в общей численности совокупностей (частостей - ). Полигоном частот называют ломанную, отрезки которой соединяют точки . Для построения полигона частот на оси абсцисс откладывают варианты , а на оси ординат – соответствующие им частоты . Точки соединяют отрезками прямых и получают полигон частот. Гистограммой частот (частостей) называется ступенчатая фигура, состоящая из прямоугольников с основаниями, равными интервалам значений и высотами, равными отношению частот (или частостей) к шагу. Эмпирической функцией распределения, построенной по выборке  объема , называется случайная функция , равная

Значения эмпирической функции распределения принадлежат отрезку [0,1].

**Постановка задачи**

Осуществить формирование репрезентативной выборки заданного объема из имеющейся генеральной совокупности экспериментальных данных. Осуществить последовательное преобразование полученной выборки в ранжированный, вариационный и интервальный ряды. Применительно к интервальному ряду построить и отобразить графически полигон, гистограмму и эмпирическую функцию распределения для абсолютных и относительных частот. Полученные результаты содержательно проинтерпретировать.

**Выполнение работы**

Выборка состоит из данных наблюдений относительно объемного веса при влажности и модуля упругости при сжатии вдоль волокон древесины резонансной ели.

Формирование репрезентативной выборки заданного объема из имеющейся генеральной совокупности экспериментальных данных представлено в таблице 1. Объём выборки: 104.

*Таблица 1*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *1* | 460 | 124.5 | *25* | 394 | 112.1 | *49* | 411 | 112.9 | *73* | 428 | 131.6 | *97* | 378 | 103.8 |
| *2* | 525 | 148.3 | *26* | 434 | 118.6 | *50* | 451 | 124.3 | *74* | 510 | 140.6 | *98* | 576 | 170.1 |
| *3* | 503 | 146.6 | *27* | 518 | 151.3 | *51* | 466 | 130.3 | *75* | 478 | 126.6 | *99* | 452 | 116.1 |
| *4* | 482 | 148.2 | *28* | 522 | 143.8 | *52* | 433 | 130.0 | *76* | 421 | 115.1 | *100* | 543 | 155.4 |
| *5* | 470 | 124.0 | *29* | 511 | 149.5 | *53* | 492 | 137.5 | *77* | 510 | 153.9 | *101* | 538 | 165.0 |
| *6* | 400 | 114.6 | *30* | 437 | 124.3 | *54* | 503 | 148.5 | *78* | 351 | 102.9 | *102* | 523 | 172.8 |
| *7* | 398 | 109.0 | *31* | 352 | 87.7 | *55* | 451 | 128.6 | *79* | 493 | 149.7 | *103* | 434 | 108.7 |
| *8* | 514 | 174.6 | *32* | 406 | 112.4 | *56* | 415 | 107.1 | *80* | 411 | 115.2 | *104* | 458 | 128.0 |
| *9* | 518 | 154.0 | *33* | 448 | 125.9 | *57* | 459 | 145.4 | *81* | 422 | 108.6 |  |  |  |
| *10* | 383 | 109.7 | *34* | 493 | 129.7 | *58* | 442 | 123.4 | *82* | 402 | 120.8 |  |  |  |
| *11* | 412 | 117.9 | *35* | 468 | 128.9 | *59* | 424 | 117.1 | *83* | 438 | 126.7 |  |  |  |
| *12* | 320 | 64.5 | *36* | 345 | 95.9 | *60* | 397 | 108.6 | *84* | 485 | 138.6 |  |  |  |
| *13* | 473 | 137.9 | *37* | 523 | 152.6 | *61* | 414 | 113.5 | *85* | 496 | 155.3 |  |  |  |
| *14* | 438 | 134.1 | *38* | 498 | 144.3 | *62* | 437 | 129.2 | *86* | 453 | 126.4 |  |  |  |
| *15* | 359 | 71.9 | *39* | 482 | 139.9 | *63* | 512 | 169.9 | *87* | 377 | 96.1 |  |  |  |
| *16* | 569 | 157.4 | *40* | 487 | 146.0 | *64* | 525 | 165.9 | *88* | 540 | 156.7 |  |  |  |
| *17* | 423 | 115.9 | *41* | 331 | 84.6 | *65* | 546 | 177.0 | *89* | 502 | 137.2 |  |  |  |
| *18* | 460 | 140.7 | *42* | 416 | 120.5 | *66* | 422 | 122.9 | *90* | 408 | 110.0 |  |  |  |
| *19* | 372 | 81.7 | *43* | 358 | 98.3 | *67* | 495 | 150.9 | *91* | 417 | 124.3 |  |  |  |
| *20* | 383 | 107.4 | *44* | 463 | 144.9 | *68* | 452 | 131.0 | *92* | 474 | 132.5 |  |  |  |
| *21* | 409 | 116.7 | *45* | 462 | 138.8 | *69* | 465 | 140.7 | *93* | 480 | 153.9 |  |  |  |
| *22* | 444 | 130.0 | *46* | 413 | 110.8 | *70* | 391 | 107.5 | *94* | 483 | 130.3 |  |  |  |
| *23* | 463 | 136.7 | *47* | 506 | 153.5 | *71* | 426 | 128.2 | *95* | 472 | 135.6 |  |  |  |
| *24* | 482 | 150.1 | *48* | 465 | 140.9 | *72* | 482 | 136.4 | *96* | 477 | 146.0 |  |  |  |

В таблице 2 представлена выборка только для .

*Таблица 2*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***i*** |  | ***i*** |  | ***i*** |  | ***i*** |  | ***i*** |  |
| *1* | 460 | *25* | 394 | *49* | 411 | *73* | 428 | *97* | 378 |
| *2* | 525 | *26* | 434 | *50* | 451 | *74* | 510 | *98* | 576 |
| *3* | 503 | *27* | 518 | *51* | 466 | *75* | 478 | *99* | 452 |
| *4* | 482 | *28* | 522 | *52* | 433 | *76* | 421 | *100* | 543 |
| *5* | 470 | *29* | 511 | *53* | 492 | *77* | 510 | *101* | 538 |
| *6* | 400 | *30* | 437 | *54* | 503 | *78* | 351 | *102* | 523 |
| *7* | 398 | *31* | 352 | *55* | 451 | *79* | 493 | *103* | 434 |
| *8* | 514 | *32* | 406 | *56* | 415 | *80* | 411 | *104* | 458 |
| *9* | 518 | *33* | 448 | *57* | 459 | *81* | 422 |  |  |
| *10* | 383 | *34* | 493 | *58* | 442 | *82* | 402 |  |  |
| *11* | 412 | *35* | 468 | *59* | 424 | *83* | 438 |  |  |
| *12* | 320 | *36* | 345 | *60* | 397 | *84* | 485 |  |  |
| *13* | 473 | *37* | 523 | *61* | 414 | *85* | 496 |  |  |
| *14* | 438 | *38* | 498 | *62* | 437 | *86* | 453 |  |  |
| *15* | 359 | *39* | 482 | *63* | 512 | *87* | 377 |  |  |
| *16* | 569 | *40* | 487 | *64* | 525 | *88* | 540 |  |  |
| *17* | 423 | *41* | 331 | *65* | 546 | *89* | 502 |  |  |
| *18* | 460 | *42* | 416 | *66* | 422 | *90* | 408 |  |  |
| *19* | 372 | *43* | 358 | *67* | 495 | *91* | 417 |  |  |
| *20* | 383 | *44* | 463 | *68* | 452 | *92* | 474 |  |  |
| *21* | 409 | *45* | 462 | *69* | 465 | *93* | 480 |  |  |

*Продолжение таблицы 2*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *22* | 444 | *46* | 413 | *70* | 391 | *94* | 483 |  |  |
| *23* | 463 | *47* | 506 | *71* | 426 | *95* | 472 |  |  |
| *24* | 482 | *48* | 465 | *72* | 482 | *96* | 477 |  |  |

* Ранжированный ряд

В таблице 3 представлено преобразование выборки в ранжированный ряд.

*Таблица 3*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***i*** |  | ***i*** |  | ***i*** |  | ***i*** |  | ***i*** |  |
| *1* | 320 | *25* | 413 | *49* | 452 | *73* | 482 | *97* | 525 |
| *2* | 331 | *26* | 414 | *50* | 452 | *74* | 483 | *98* | 525 |
| *3* | 345 | *27* | 415 | *51* | 453 | *75* | 485 | *99* | 538 |
| *4* | 351 | *28* | 416 | *52* | 458 | *76* | 487 | *100* | 540 |
| *5* | 352 | *29* | 417 | *53* | 459 | *77* | 492 | *101* | 543 |
| *6* | 358 | *30* | 421 | *54* | 460 | *78* | 493 | *102* | 546 |
| *7* | 359 | *31* | 422 | *55* | 460 | *79* | 493 | *103* | 569 |
| *8* | 372 | *32* | 422 | *56* | 462 | *80* | 495 | *104* | 576 |
| *9* | 377 | *33* | 423 | *57* | 463 | *81* | 496 |  |  |
| *10* | 378 | *34* | 424 | *58* | 463 | *82* | 498 |  |  |
| *11* | 383 | *35* | 426 | *59* | 465 | *83* | 502 |  |  |
| *12* | 383 | *36* | 428 | *60* | 465 | *84* | 503 |  |  |
| *13* | 391 | *37* | 433 | *61* | 466 | *85* | 503 |  |  |
| *14* | 394 | *38* | 434 | *62* | 468 | *86* | 506 |  |  |
| *15* | 397 | *39* | 434 | *63* | 470 | *87* | 510 |  |  |
| *16* | 398 | *40* | 437 | *64* | 472 | *88* | 510 |  |  |
| *17* | 400 | *41* | 437 | *65* | 473 | *89* | 511 |  |  |
| *18* | 402 | *42* | 438 | *66* | 474 | *90* | 512 |  |  |
| *19* | 406 | *43* | 438 | *67* | 477 | *91* | 514 |  |  |
| *20* | 408 | *44* | 442 | *68* | 478 | *92* | 518 |  |  |
| *21* | 409 | *45* | 444 | *69* | 480 | *93* | 518 |  |  |
| *22* | 411 | *46* | 448 | *70* | 482 | *94* | 522 |  |  |
| *23* | 411 | *47* | 451 | *71* | 482 | *95* | 523 |  |  |
| *24* | 412 | *48* | 451 | *72* | 482 | *96* | 523 |  |  |

В таблице 3 можно заметить, что наименьшее значение в выборке , а наибольшее значение .

* Вариационный ряд

В таблицах 4 и 5 представлено преобразование полученной выборки в вариационный ряд с абсолютными и относительными частотами соответственно.

*Таблица 4*

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***i*** |  |  | ***i*** |  |  | ***i*** |  |  |  |  |  |
| *1* | 320 | 1 | *22* | 412 | 1 | *43* | 453 | 1 | *64* | 493 | 2 |
| *2* | 331 | 1 | *23* | 413 | 1 | *44* | 458 | 1 | *65* | 495 | 1 |
| *3* | 345 | 1 | *24* | 414 | 1 | *45* | 459 | 1 | *66* | 496 | 1 |
| *4* | 351 | 1 | *25* | 415 | 1 | *46* | 460 | 2 | *67* | 498 | 1 |
| *5* | 352 | 1 | *26* | 416 | 1 | *47* | 462 | 1 | *68* | 502 | 1 |
| *6* | 358 | 1 | *27* | 417 | 1 | *48* | 463 | 2 | *69* | 503 | 2 |
| *7* | 359 | 1 | *28* | 421 | 1 | *49* | 465 | 2 | *70* | 506 | 1 |
| *8* | 372 | 1 | *29* | 422 | 2 | *50* | 466 | 1 | *71* | 510 | 2 |
| *9* | 377 | 1 | *30* | 423 | 1 | *51* | 468 | 1 | *72* | 511 | 1 |
| *10* | 378 | 1 | *31* | 424 | 1 | *52* | 470 | 1 | *73* | 512 | 1 |
| *11* | 383 | 2 | *32* | 426 | 1 | *53* | 472 | 1 | *74* | 514 | 1 |
| *12* | 391 | 1 | *33* | 428 | 1 | *54* | 473 | 1 | *75* | 518 | 2 |
| *13* | 394 | 1 | *34* | 433 | 1 | *55* | 474 | 1 | *76* | 522 | 1 |
| *14* | 397 | 1 | *35* | 434 | 2 | *56* | 477 | 1 | *77* | 523 | 2 |
| *15* | 398 | 1 | *36* | 437 | 2 | *57* | 478 | 1 | *78* | 525 | 2 |
| *16* | 400 | 1 | *37* | 438 | 2 | *58* | 480 | 1 | *79* | 538 | 1 |
| *17* | 402 | 1 | *38* | 442 | 1 | *59* | 482 | 4 | *80* | 540 | 1 |
| *18* | 406 | 1 | *39* | 444 | 1 | *60* | 483 | 1 | *81* | 543 | 1 |
| *19* | 408 | 1 | *40* | 448 | 1 | *61* | 485 | 1 | *82* | 546 | 1 |
| *20* | 409 | 1 | *41* | 451 | 2 | *62* | 487 | 1 | *83* | 569 | 1 |
| *21* | 411 | 2 | *42* | 452 | 2 | *63* | 492 | 1 | *84* | 576 | 1 |

*Таблица 5*

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***i*** |  |  | ***i*** |  |  | ***i*** |  |  |  |  |  |
| *1* | 320 | 0.0096 | *22* | 412 | 0.0096 | *43* | 453 | 0.0096 | *64* | 493 | 0.0192 |
| *2* | 331 | 0.0096 | *23* | 413 | 0.0096 | *44* | 458 | 0.0096 | *65* | 495 | 0.0096 |
| *3* | 345 | 0.0096 | *24* | 414 | 0.0096 | *45* | 459 | 0.0096 | *66* | 496 | 0.0096 |
| *4* | 351 | 0.0096 | *25* | 415 | 0.0096 | *46* | 460 | 0.0192 | *67* | 498 | 0.0096 |
| *5* | 352 | 0.0096 | *26* | 416 | 0.0096 | *47* | 462 | 0.0096 | *68* | 502 | 0.0096 |

*Продолжение таблицы 5*

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *6* | 358 | 0.0096 | *27* | 417 | 0.0096 | *48* | 463 | 0.0192 | *69* | 503 | 0.0192 |
| *7* | 359 | 0.0096 | *28* | 421 | 0.0096 | *49* | 465 | 0.0192 | *70* | 506 | 0.0096 |
| *8* | 372 | 0.0096 | *29* | 422 | 0.0192 | *50* | 466 | 0.0096 | *71* | 510 | 0.0192 |
| *9* | 377 | 0.0096 | *30* | 423 | 0.0096 | *51* | 468 | 0.0096 | *72* | 511 | 0.0096 |
| *10* | 378 | 0.0096 | *31* | 424 | 0.0096 | *52* | 470 | 0.0096 | *73* | 512 | 0.0096 |
| *11* | 383 | 0.0192 | *32* | 426 | 0.0096 | *53* | 472 | 0.0096 | *74* | 514 | 0.0096 |
| *12* | 391 | 0.0096 | *33* | 428 | 0.0096 | *54* | 473 | 0.0096 | *75* | 518 | 0.0192 |
| *13* | 394 | 0.0096 | *34* | 433 | 0.0096 | *55* | 474 | 0.0096 | *76* | 522 | 0.0096 |
| *14* | 397 | 0.0096 | *35* | 434 | 0.0192 | *56* | 477 | 0.0096 | *77* | 523 | 0.0192 |
| *15* | 398 | 0.0096 | *36* | 437 | 0.0192 | *57* | 478 | 0.0096 | *78* | 525 | 0.0192 |
| *16* | 400 | 0.0096 | *37* | 438 | 0.0192 | *58* | 480 | 0.0096 | *79* | 538 | 0.0096 |
| *17* | 402 | 0.0096 | *38* | 442 | 0.0096 | *59* | 482 | 0.0385 | *80* | 540 | 0.0096 |
| *18* | 406 | 0.0096 | *39* | 444 | 0.0096 | *60* | 483 | 0.0096 | *81* | 543 | 0.0096 |
| *19* | 408 | 0.0096 | *40* | 448 | 0.0096 | *61* | 485 | 0.0096 | *82* | 546 | 0.0096 |
| *20* | 409 | 0.0096 | *41* | 451 | 0.0192 | *62* | 487 | 0.0096 | *83* | 569 | 0.0096 |
| *21* | 411 | 0.0192 | *42* | 452 | 0.0192 | *63* | 492 | 0.0096 | *84* | 576 | 0.0096 |

* Интервальный ряд

С помощью формулы Стерджесса было вычислено количество интервалов:

Получено нечетное количество интервалов.

Ширина интервала была вычислена по формуле:

В таблице 6 представлен полученный интервальный ряд.

*Таблица 6*

|  |  |  |  |
| --- | --- | --- | --- |
| ***Границы***  ***интервалов*** | ***Середины***  ***интервалов*** | ***Абсолютная***  ***частота*** | ***Относительная***  ***частота*** |
| [320, 357) | 338.5 | 5 | 0.048 |
| [357, 394) | 375.5 | 8 | 0.077 |
| [394, 431) | 412.5 | 23 | 0.221 |

*Продолжение таблицы 6*

|  |  |  |  |
| --- | --- | --- | --- |
| [431, 468) | 449.5 | 25 | 0.240 |
| [468, 505) | 486.5 | 24 | 0.231 |
| [505, 542) | 523.5 | 15 | 0.144 |
| [542, 576) | 559 | 4 | 0.038 |

* Графики для интервального ряда абсолютных частот

Полигон представлен на рис. 1.

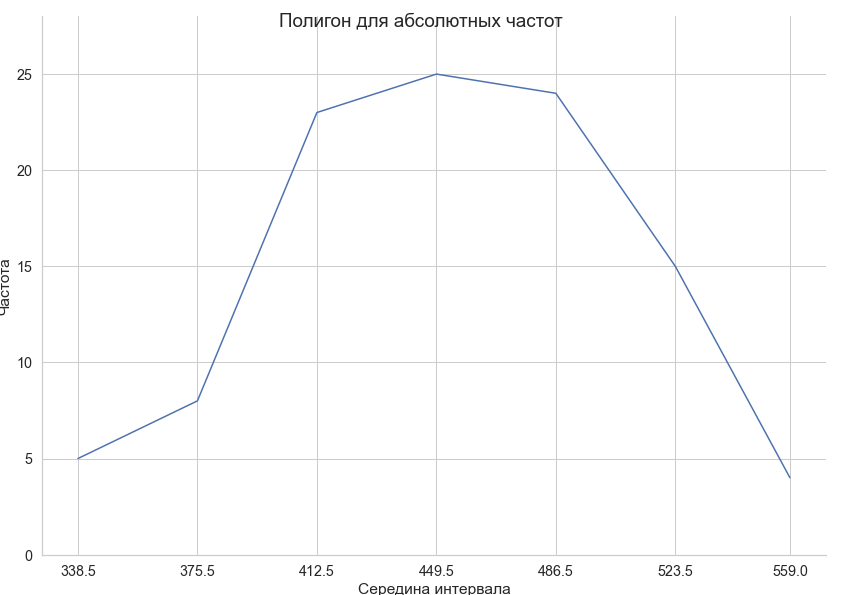


Рисунок 1 – Полигон для абсолютных частот

Гистограмма, представлена на рис. 2.

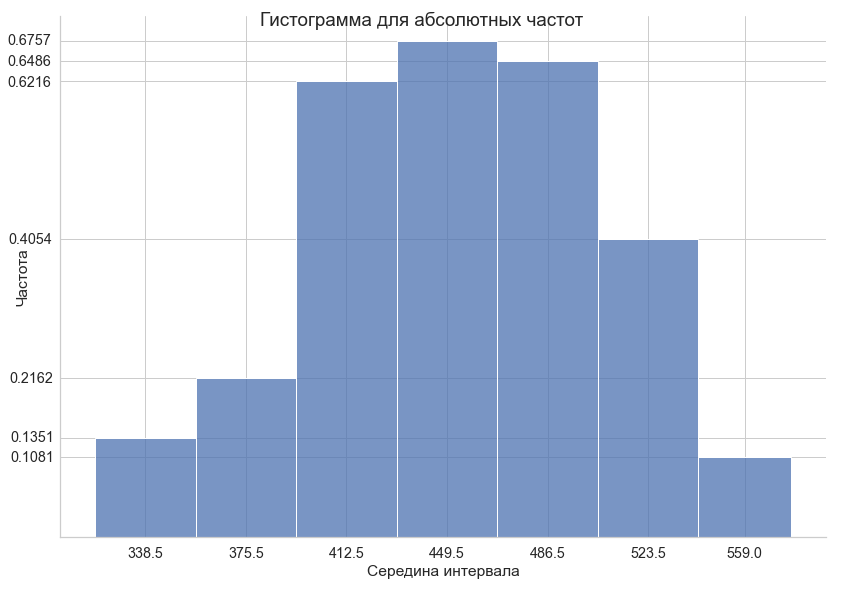


Рисунок 2 – Гистограмма для абсолютных частот

* Графики для интервального ряда относительных частот

Полигон представлен на рис. 3.

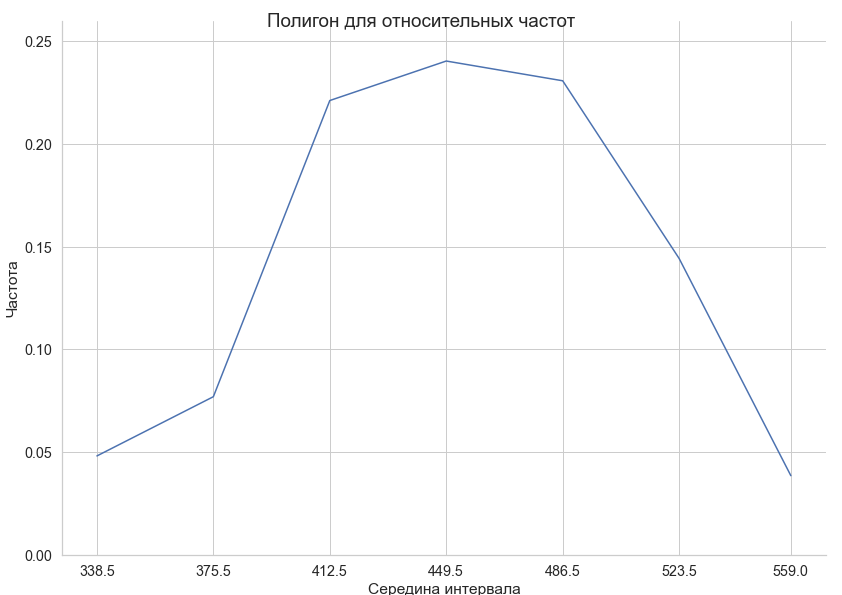


Рисунок 3 – Полигон для относительных частот

Гистограмма, представлена на рис. 4.



Рисунок 4 – Гистограмма для относительных частот

Эмпирическая функция распределения, построенная применительно к интервальному ряду для относительных частот представлен на рис. 5.

Функция распределения:

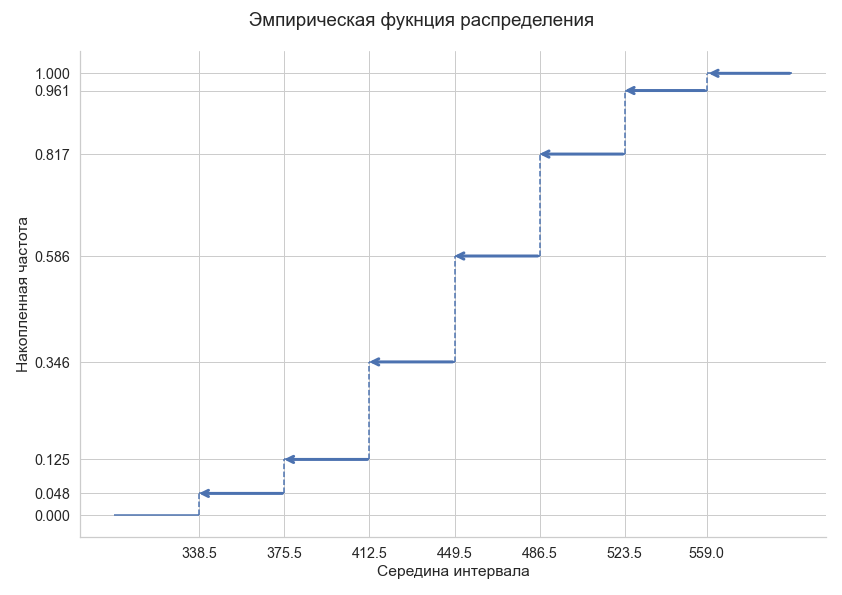


Рисунок 5 – График эмпирической функции распределения

**Выводы**

В ходе данной лабораторной работы была выбрана двумерная генеральная совокупность, согласованная с преподавателем. Из генеральной совокупности была сформирована репрезентативная выборка.

Выборка была преобразована в ранжированный, вариационный и интервальный ряды. Используя полученный интервальный ряд был построен полигон и гистограмма для абсолютных и относительных частот. Для интервального ряда относительных частот был построен график эмпирической функции распределения.

Элементы ранжированного ряда расположены в порядке возрастания их значений, поэтому можно определить минимальный и максимальный элемент выборки. Для данной выборки , .

Вариационный ряд получается в результате объединения одинаковых элементов, поэтому можно определить варианту с наибольшей частотой повторения в выборке. Для данной выборки это с абсолютной частотой и относительной частотой .

Интервальный ряд был построен с помощью деления вариационного ряда на интервалы. По формуле Стерджесса было получено нечетное количество интервалов . По сформированному интервальному ряду можно увидеть, что наибольшая частота попадания значений вариант в интервале .

Такой же результат можно увидеть на построенных полигоне и гистограмме. Форма графиков не меняется для абсолютных и относительных частот, меняется ось ординат, которая для полигонов обозначает частоты (абсолютные или относительные), а для гистограмм уже площадь прямоугольника обозначает частоты, что можно проверить путем умножения высоты столбца на ширину. Сумма площадей прямоугольников гистограммы для абсолютных частот равна объему выборки, а для относительных частот равна . На графике эмпирической функции распределения можно увидеть отношение накопленных частот до середины интервалов к объему выборки.

По виду полигона и гистограммы можно сделать предположение о том, что анализируемая переменная имеет примерно нормальное распределение.

**ПРИЛОЖЕНИЕ А**

**ИСХОДНЫЙ КОД**

# %%

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from IPython.core.interactiveshell import InteractiveShell

InteractiveShell.ast\_node\_interactivity = "all"

# %% [markdown]

# ## Выборка

# %%

raw = pd.read\_csv('c:/Users/gandh/dev/unv/smoed/data/sample.csv')

df = pd.read\_csv('c:/Users/gandh/dev/unv/smoed/data/main\_data.csv')

df.to\_csv('data/data1.csv', index=False)

n = len(df)

n

# %% [markdown]

# ## Распределение

# %%

sns.set\_theme(style="whitegrid", palette='deep', context='notebook', font\_scale=1.3)

ax = sns.catplot(data=raw, kind='box', height=8.27, aspect=11.7/8.27)

plt.savefig('pics/1.png')

# %%

sns.set\_theme(style="whitegrid", palette='deep', context='notebook', font\_scale=1.3)

ax = sns.catplot(data=df, kind='box', height=8.27, aspect=11.7/8.27)

plt.savefig('pics/2.png')

# %% [markdown]

# ## Одна переменная

# %%

df2 = df.drop('E', axis=1)

df2.to\_csv('data/data2.csv', index=False)

df2.head()

# %% [markdown]

# ## Ранжированный ряд

# %%

df2 = df2.sort\_values(by=['nu'], ignore\_index = True)

df2.to\_csv('data/data3.csv', index=False)

df2.head()

# %%

df2.min()

df2.max()

# %%

X = df2['nu']

# %% [markdown]

# ## Вариационный ряд

# %%

table\_af = X.value\_counts().sort\_index()

table\_rf = X.value\_counts(normalize=True).sort\_index()

table\_af = pd.DataFrame({'nu': table\_af.index, 'af': table\_af.values})

table\_rf = pd.DataFrame({'nu': table\_rf.index, 'rf': table\_rf.values})

table\_rf2 = table\_rf.copy()

table\_rf2['rf'] = np.round(table\_rf2['rf'], 4)

table\_af.to\_csv('data/data4.csv', index=False)

table\_rf2.to\_csv('data/data5.csv', index=False)

# %% [markdown]

# ## Интервальный ряд

# %%

k = 1+3.31\*np.log10(n)

k = int(np.floor(k))

k

# %%

min(X)

max(X)

# %%

h = (max(X)-min(X))/k

h = int(np.ceil(h))

h

# %%

data\_interval = pd.concat([table\_af, table\_rf], ignore\_index=True, axis=1).drop(2, axis=1)

data\_interval.columns = ['nu', 'af', 'rf']

data\_interval.to\_csv('data/data6.csv', index=False)

# %%

ivs = np.hstack((np.arange(min(X), max(X), h), np.array(max(X))))

# %%

data\_interval['inter'] = pd.cut(data\_interval['nu'], bins=ivs,

right=False)

data\_interval['inter'].value\_counts().sort\_index()

data\_interval.iloc[83, 3] = data\_interval.iloc[82, 3]

# %%

f\_inter = data\_interval.groupby(['inter'])[['af', 'rf']].apply(sum).reset\_index()

f\_inter['avg\_inter'] = np.array([np.mean([ivs[i], ivs[i+1]], axis=0) for i in range(k)])

f\_inter = f\_inter[['inter', 'avg\_inter', 'af', 'rf']]

f\_inter.to\_csv('data/data7.csv', index=False)

# %% [markdown]

# ## Графики абсолют

# %%

ax = sns.relplot(data=f\_inter, x='avg\_inter', y='af', kind='line', height=8.27, aspect=11.7/8.27)

ax.set\_axis\_labels('Середина интервала', 'Частота')

ax.set(ylim=[0,28], xticks=f\_inter['avg\_inter'])

ax.fig.suptitle('Полигон для абсолютных частот')

plt.savefig('pics/3.png')

# %%

ax = sns.displot(data=df, x='nu', bins=ivs, kind='hist', height=8.27, aspect=11.7/8.27, stat='probability')

ax.set\_axis\_labels('Середина интервала', 'Частота')

ax.set(ylim=[0,.26],xticks=f\_inter['avg\_inter'])

ax.fig.suptitle('Гистограмма для абсолютных частот')

plt.savefig('pics/4.png')

# %%

f\_inter['sum\_rf'] = f\_inter['rf'].cumsum()

f\_inter

# %%

ax = sns.relplot(data=f\_inter, x='avg\_inter', y='sum\_rf', s=80,

kind='scatter', height=8.27, aspect=11.7/8.27, color='b')

for i in range(6):

plt.hlines(f\_inter['sum\_rf'][i], f\_inter['avg\_inter'][i], f\_inter['avg\_inter'][i+1], color='b')

plt.hlines(1, 559, 589, color='b')

for i in range(6):

plt.vlines(f\_inter['avg\_inter'][i+1], f\_inter['sum\_rf'][i], f\_inter['sum\_rf'][i+1], color='b', linestyle='--')

plt.vlines(338.5, 0, 0.048, color='b', linestyle='--')

ax.set\_axis\_labels('Середина интервала', '')

ax.set(xticks=f\_inter['avg\_inter'])

ax.fig.suptitle('Эмпирическая фукнция распределения')

plt.savefig('pics/5.png')

# %% [markdown]

# ## Графики относительно

# %%

ax = sns.relplot(data=f\_inter, x='avg\_inter', y='rf', kind='line', height=8.27, aspect=11.7/8.27)

ax.set\_axis\_labels('Середина интервала', 'Частота')

ax.set(ylim=[0,0.26], xticks=f\_inter['avg\_inter'])

ax.fig.suptitle('Полигон для относительных частот')

plt.savefig('pics/6.png')

# %%

ax = sns.displot(data=df, x='nu', bins=ivs, kind='hist', height=8.27, aspect=11.7/8.27, stat='density')

ax.set\_axis\_labels('Середина интервала', 'Частота')

ax.set(xticks=f\_inter['avg\_inter'])

ax.fig.suptitle('Гистограмма для относительных частот')

plt.savefig('pics/7.png')

# %%

f\_inter['af']/h

f\_inter['rf']/h