МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ЛЬВІВСЬКА ПОЛІТЕХНІКА»

Кафедра «Інформаційні системи та мережі»



АНАЛІЗ СТАТИСТИЧНИХ ДАНИХ

Розрахункова робота

Виконала:

Студентка групи КН-28

Гаврилів Р. М.

Прийняла:

к.т.н., доцент

Висоцька В.А.

Львів 2016

**Мета роботи:** ознайомлення з основними методами візуалізації, графічного відображення та первинної статистичної обробки числових даних, представлених вибірковою сукупністю або часовим рядом. Метою даної роботи є ознайомлення з основними методами висвітлення тенденції поведінки досліджуваного показника, яка представлена характером його тренду, з допомогою методів згладжування часових рядів та подання отриманих результатів засобами табличного процесора MS Excel. Метою роботи також є ознайомлення з методами кореляційного аналізу експериментальних даних, поданих часовими послідовностями.

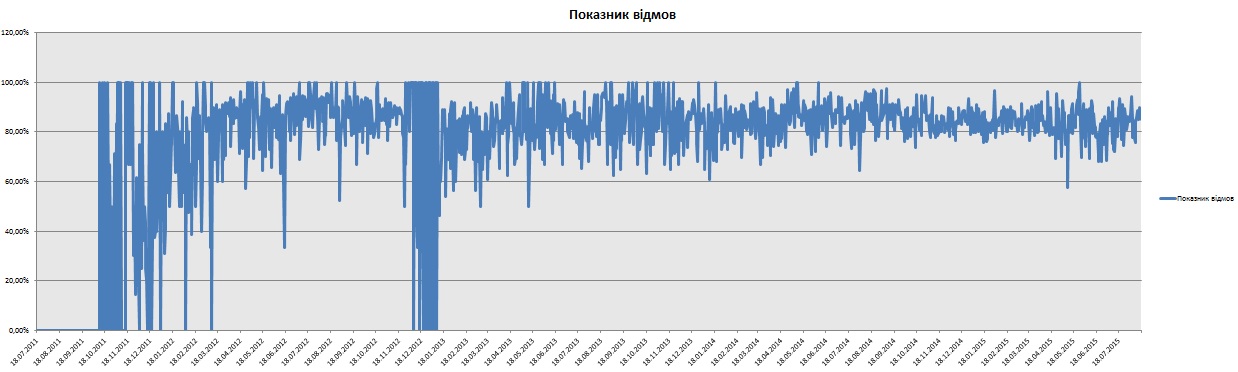
**Теоретичнівідомості**

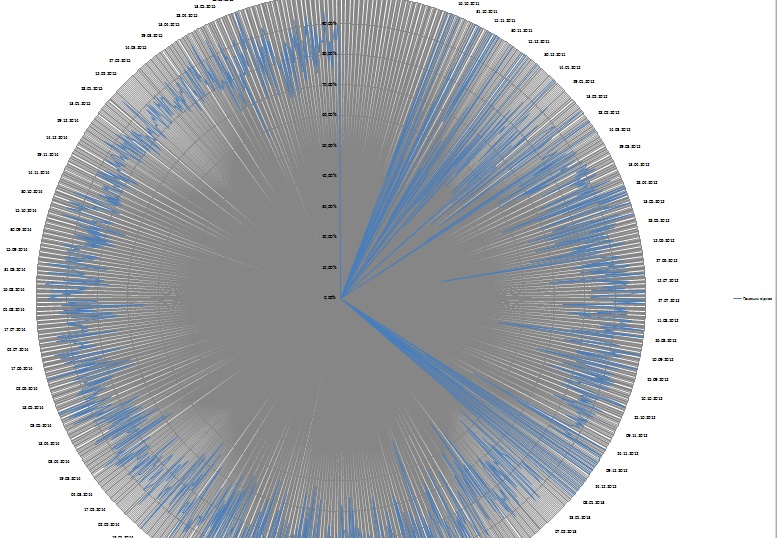
Статистичне зведення – наукове опрацювання матеріалів статистичного спостереження для одержання узагальнюючих показників (характеристик). Визначною рисою ознак окремих одиниць спостереження є їх мінливість (варіація). Якби ознаки окремих одиниць спостереження мали однакове значення, тобто була б відсутня їх мінливість, то відпала б сама необхідність збору цих ознак (характеристик) для кожної одиниці спостереження.

Аналізування варіації характеристик (показників) одиниць спостереження має вагоме позначення. Головним постає завдання - підсумувати, узагальнити таким чином результати, щоб стало можливим виявити характерні істотні властивості тих чи інших типів явищ, викрити необхідності досліджуваних процесів. Статистичне зведення включає не лише обрахунки матеріалу, але й певне групування даних та представлення їх у вигляді статистичних рядів, таблиць і графіків.

**Хід виконання розрахункової роботи:**

Опираючись на задану умову та вхідні дані згідно варіанту, будуємо відповідні графіки у декартовій та полярній системах координат (рис.1, 2).

*Рис. 1. Графік, побудований у декартовій системі*

  
*Рис. 2. Графік у полярній системі (пелюсткова діаграма)*

У рядку Меню, вкладці Дані, використовуємо надбудову Пакет Аналізу та обираємо команду Описова статистика. Нам буде подана Описова статистика заданої вибірки (див. Таблиця 1).

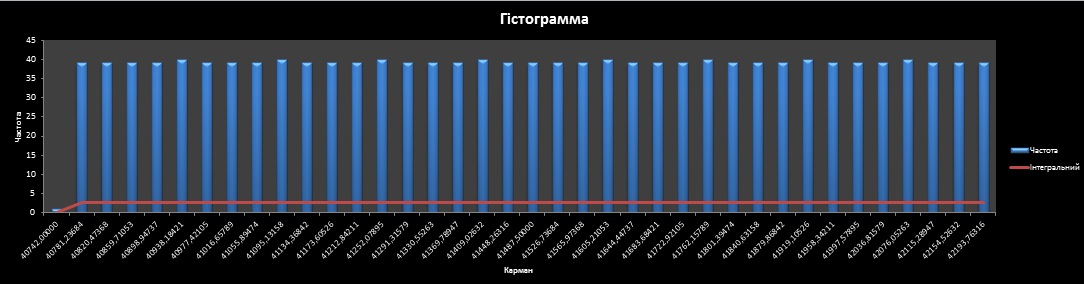
*Таблиця 1.Описова статистика*

|  |  |
| --- | --- |
| ***Показник відмов*** | |
|
| Среднее | 41487,50000 |
| Стандартная ошибка | 11,15422192 |
| Медиана | 41487,50000 |
| Мода | 0 |
| Стандартное отклонение | 430,8476142 |
| Дисперсия выборки | 185629,6667 |
| Эксцесс | -1,2 |
| Асимметричность | 5,84563E-17 |
| Интервал | 1491 |
| Минимум | 40742 |
| Максимум | 42233 |
| Сумма | 61899350 |
| Счет | 1492 |

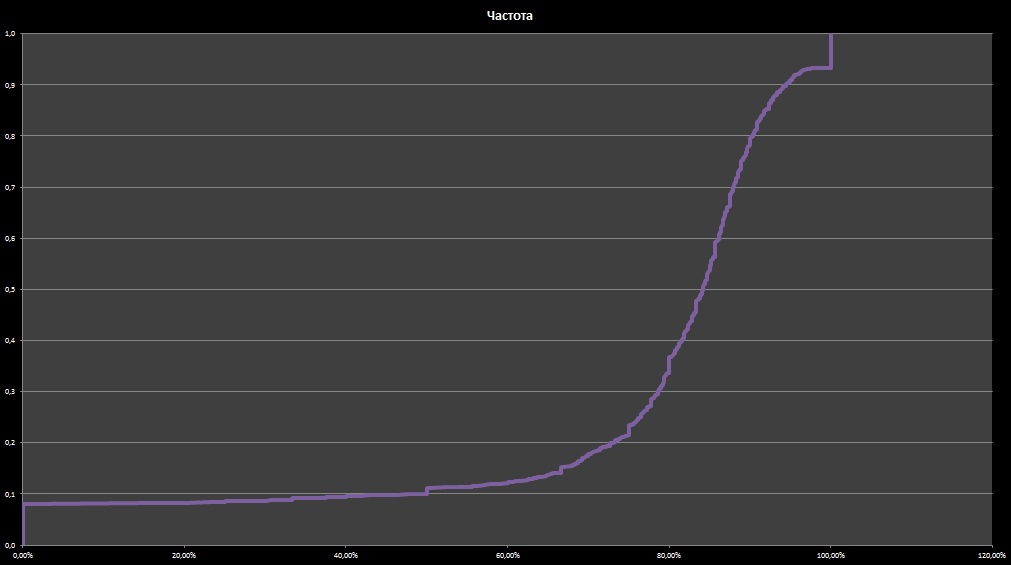
Наступним кроком необхідно побудувати Гістограму (рис. 3) та Кумуляту. Попередньо формуємо таблицю із полями: *Частота* та обрахованими *Інтегральними відсотками* кожного з інтервалів (табл.2).

*Таблиця 2. Гістограма (Пакет Аналізу)*

|  |  |  |
| --- | --- | --- |
| *Карман* | *Частота* | *Интегральный %* |
| 0 | 107 | 3,59% |
| 1111,394737 | 1385 | 50,00% |
| 2222,789474 | 0 | 50,00% |
| 3334,184211 | 0 | 50,00% |
| 4445,578947 | 0 | 50,00% |
| 5556,973684 | 0 | 50,00% |
| 6668,368421 | 0 | 50,00% |
| 7779,763158 | 0 | 50,00% |
| 8891,157895 | 0 | 50,00% |
| 10002,55263 | 0 | 50,00% |
| 11113,94737 | 0 | 50,00% |
| 12225,34211 | 0 | 50,00% |
| 13336,73684 | 0 | 50,00% |
| 14448,13158 | 0 | 50,00% |
| 15559,52632 | 0 | 50,00% |
| 16670,92105 | 0 | 50,00% |
| 17782,31579 | 0 | 50,00% |
| 18893,71053 | 0 | 50,00% |
| 20005,10526 | 0 | 50,00% |
| 21116,5 | 0 | 50,00% |
| 22227,89474 | 0 | 50,00% |
| 23339,28947 | 0 | 50,00% |
| 24450,68421 | 0 | 50,00% |
| 25562,07895 | 0 | 50,00% |
| 26673,47368 | 0 | 50,00% |
| 27784,86842 | 0 | 50,00% |
| 28896,26316 | 0 | 50,00% |
| 30007,65789 | 0 | 50,00% |
| 31119,05263 | 0 | 50,00% |
| 32230,44737 | 0 | 50,00% |
| 33341,84211 | 0 | 50,00% |
| 34453,23684 | 0 | 50,00% |
| 35564,63158 | 0 | 50,00% |
| 36676,02632 | 0 | 50,00% |
| 37787,42105 | 0 | 50,00% |
| 38898,81579 | 0 | 50,00% |
| 40010,21053 | 0 | 50,00% |
| 41121,60526 | 380 | 62,73% |
| Еще | 1112 | 100,00% |

*Рис. 3. Гістограма*

На рис. 4 зображена Кумулята, побудована за відсортованими даними вхідної таблиці.

*  
Рис. 4. Кумулята*

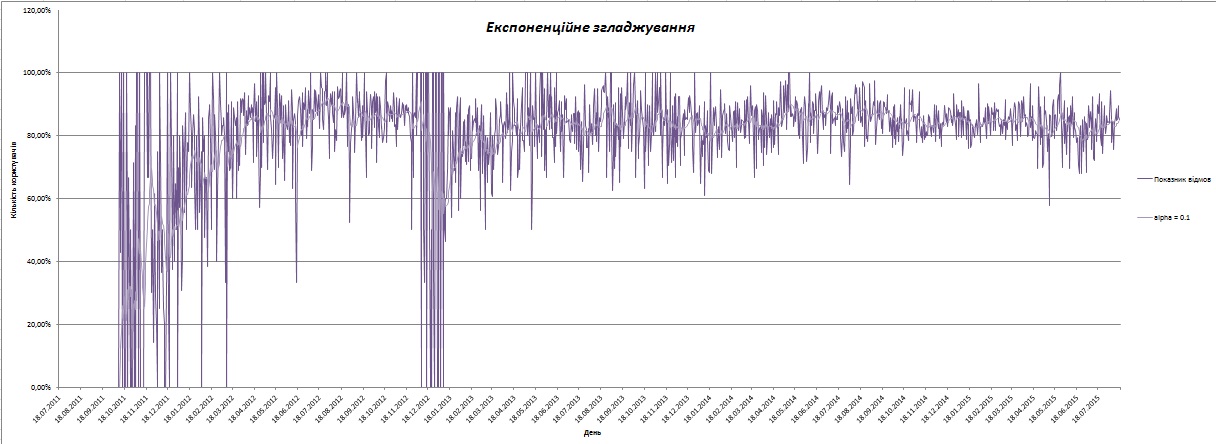
Заданий динамічний ряд, що дуже коливається, тому використовуємо згладжування - це так зване вирівнювання цих рядів. Запроваджуємо такі інтервали для згладжування за Кенделом та при медіанному згладжуванні.

Удруге здійснюємо згладжування, уже попередньо згладженої функції для інтервалів, розміром *5, 7, 9, 11, 13, 15.*

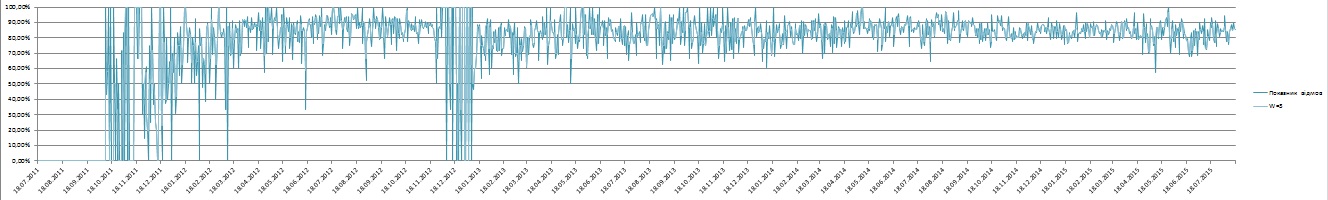
Стосовно значення параметра згладжування, то в ході експоненціального згладжування приймемо такі : .

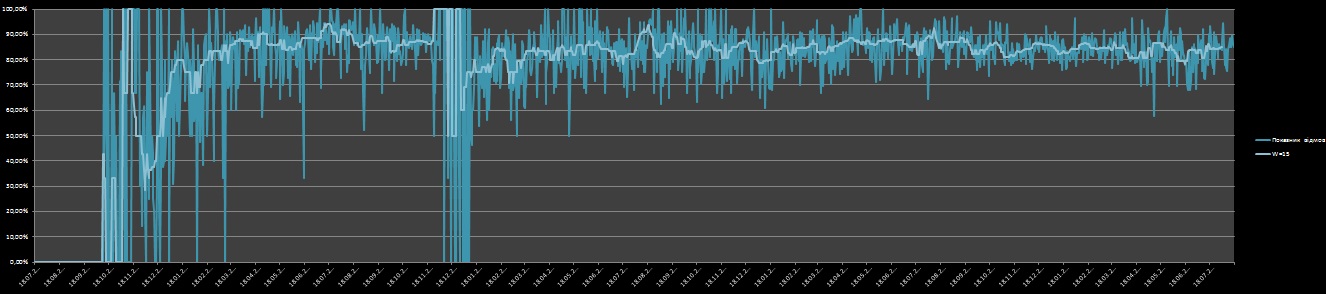
Поворотні точки — це такі точки в часі, коли проходять суттєві зміни в рядах даних. Включаючи і звичайні, й повторні згладжування, були сформовані таблиці поворотних точок.

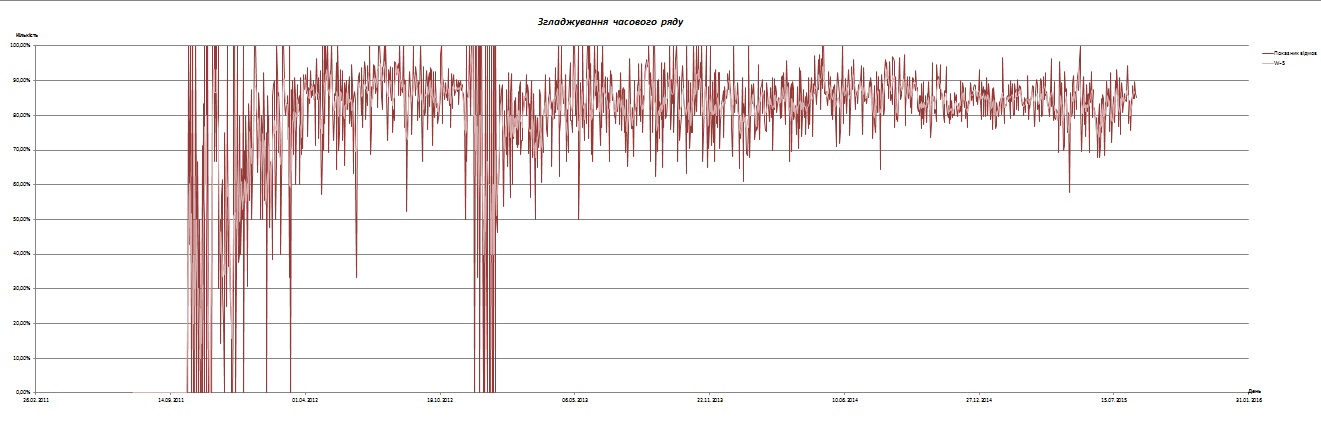
Надалі у вигляді зображень показуємо діаграми по кожному з методів. Оскільки ми знай шли коефіцієнти кореляції для відповідних методів згладжування, то на їх основі збудуємо кореляційні матриці. В результаті виводимо табличку «об’єкт-властивість», у якій сортуємо дані по роках (2011-2015). Щоб пересвідчитись, див. *Таблиця 6-7 (Об’єкт-властивість).*

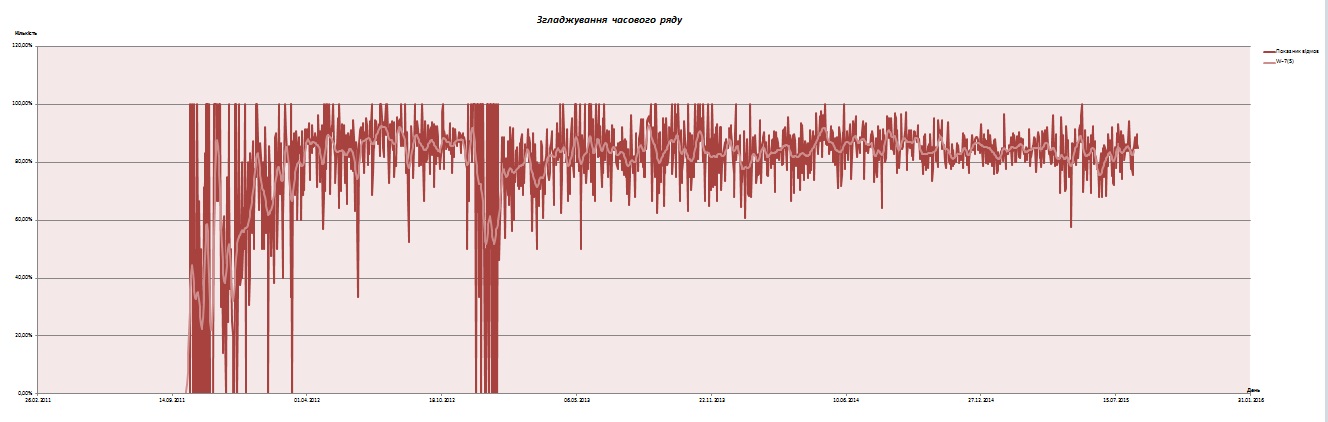


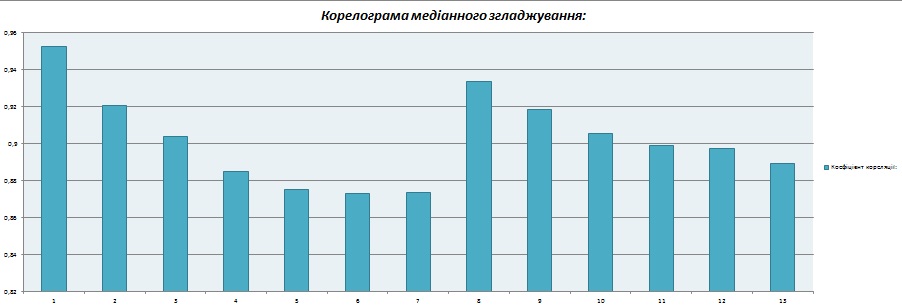
*Рис. 5. Експоненціальне згладжування з параметром α = 0.1*

 *Рис. 6. Медіанне згладжування, при w=5*

*  
Рис. 7. Медіанне згладжування, при w=15*

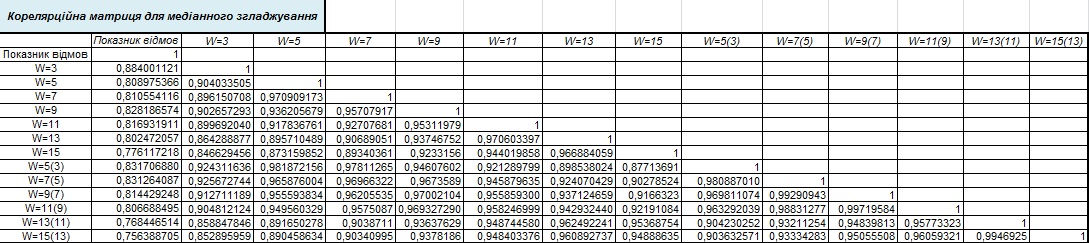
*  
Рис. 8. Згладжування за Кенделом, при w=5*

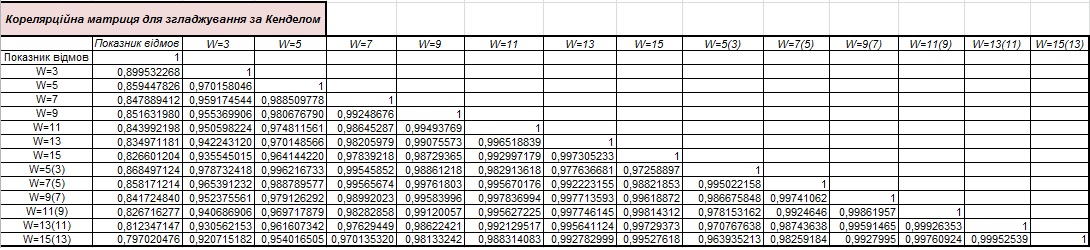
  
*Рис. 9. Повторне згладжування*

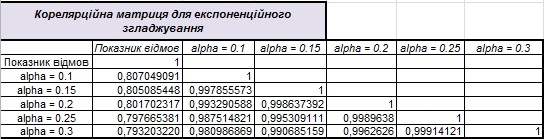
  
*Рис. 10 Корелограма для медіанного згладжування*

  
*Рис. 11. Корелограма для згладжування Кендела*

  
*Рис. 12. Корелограма для експоненційного згладжування*

*Таблиця 3. Кореляційна матриця для медіанного згладжування*

*Таблиця 4. Кореляційна матриця для згладжування за Кенделом*

*Таблиця 5. Кореляційна матриця для експоненційного згладжування*

*Таблиця 6. Об’єкт-властивість*



*Таблиця 7. Об’єкт-властивість(нормована)*



*Таблиця 8. Матриця близкостей*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Матриця близкостей*** | | | | | |
|  | **2011** | **2012** | **2013** | **2014** | **2015** |
| **2011** | 0,00 | 3,21 | 3,20 | 45,07 | 45,07 |
| **2012** | 3,21 | 0,00 | ***1,65*** | 46,24 | 46,24 |
| **2013** | 3,20 | ***1,65*** | 0,00 | 45,39 | 45,39 |
| **2014** | 45,07 | 46,24 | 45,39 | 0,00 | 0,00 |
| **2015** | 45,07 | 46,24 | 45,39 | 0,00 | 0,00 |

*Таблиця 9. Перший етап кластерного аналізу*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Кластерний аналіз (1)*** | | | | |
|
|  | *2011* | *2012, 2013* | *2014* | *2015* |
| *2011* | 0 | ***3,2*** | 45,07 | 45,07 |
| *2012, 2013* | ***3,2*** | 0 | 45,39 | 45,39 |
| *2014* | 45,07 | 45,39 | 0 | 0 |
| *2015* | 45,07 | 45,39 | 0 | 0 |

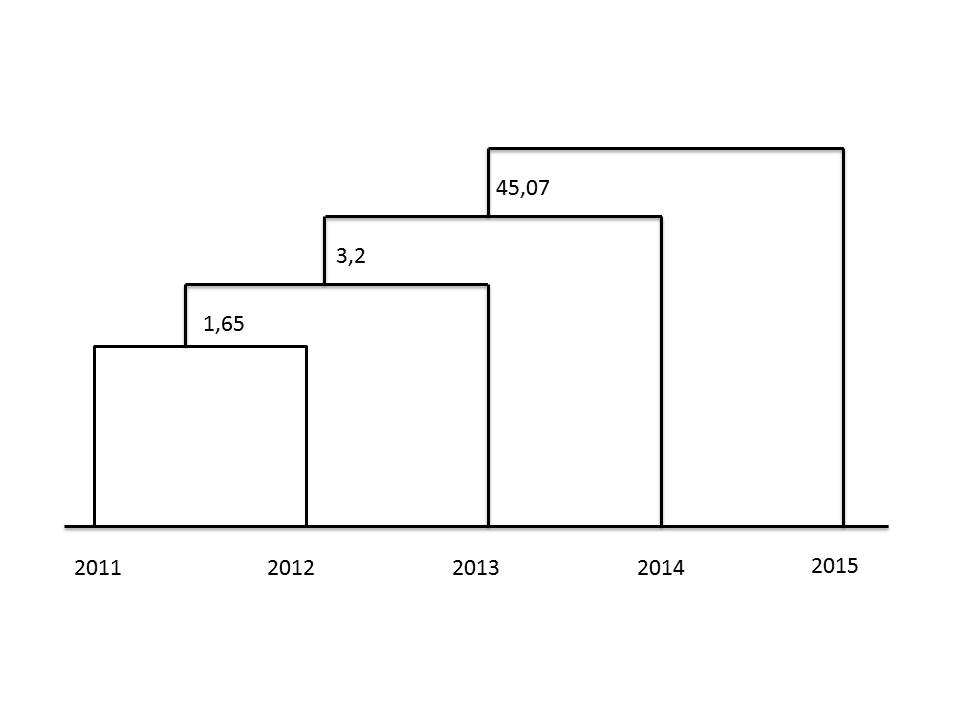
*Таблиця 10. Другий етап кластерного аналізу*

|  |  |  |  |
| --- | --- | --- | --- |
| ***Кластерний аналіз (2)*** | | | |
|
|  | *2011,2012, 2013* | *2014* | *2015* |
| *2011, 2012, 2013* | 0 | ***45,07*** | 45,07 |
| *2014* | ***45,07*** | 0 | 0 |
| *2015* | 45,07 | 0 | 0 |

*Таблиця 11. Третій етап кластерного аналізу*

|  |  |  |
| --- | --- | --- |
| ***Кластерний аналіз (3)*** | | |
|
|  | *2011, 2012,  2013, 2014* | *2015* |
| *2011, 2012, 2013, 2014* | 0 | ***45,07*** |
| *2015* | ***45,07*** | 0 |

За обрахованими кластерами, можемо сформувати та побудувати дендрограму *(див. Рис. 13).*

 *Рис. 13. Дендрограма*

**Висновок**

Виконавши дану розрахункову роботу, я ознайомлення з основними методами візуалізації, графічного відображення та первинної статистичної обробки числових даних. Залежно від поставленого завдання, я встановила певні ознаки, кількість груп, макети розроблюваних таблиць. Вміщена статистична сукупність розділяється за окремими ознаками, показниками. Такі характеристики можуть визначатись у формі підсумкових показників, величин.

Отож, у своїй розрахункові роботі я ознайомилась із основними можливості табличного процесора Excel. Застосування цього табличного процесора дозволяє легко оптимізувати і прискорити роботу з обробки інформації, угрупування та компактного подання статистичних даних та вибірки великого об’єму. Тому доведено, що подання інформації в табличному вигляді, а також у вигляді графіків, діаграм дозволяє більш наочно і краще візуально представити інформацію.