

УДК 004.03

DOI <https://doi.org/10.32782/IT/2023-3-8>

Владислав СІДАНЧЕНКО

аспірант кафедри безпеки інформації та телекомунікацій, Національний технічний університет «Дніпровська політехніка», просп. Дмитра Яворницького, 19, м. Дніпро, Україна, 49005, sidanchenko.vl.v@nmu.one

ORCID: 0000-0001-5581-9177

Бібліографічний опис статті: Сіданченко, В. (2023). Перевірка характеру розподілу даних про хімічний склад чавуну на випуску. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 3, 65–69, doi: <https://doi.org/10.32782/IT/2023-3-8>

ПЕРЕВІРКА ХАРАКТЕРУ РОЗПОДІЛУ ДАНИХ ПРО ХІМІЧНИЙ СКЛАД ЧАВУНУ НА ВИПУСКУ

Вступ. Забезпечення заданого хімічного складу чавуну на виробництві грає вирішальну роль у задачі визначення якості металургійної продукції і важливих економічних показників підприємства. Для досягнення цієї мети виникає необхідність розробки та впровадження ефективних методів прогнозування, які відіграють ключову роль оптимального управління процесом виплавки чавуну заданої якості.

Складні технологічні процеси, такі як доменна плавка, піддаються впливу багатьох факторів різної природи. Ці фактори впливають як на загальний перебіг та розвиток процесу, так і на його окремі властивості та кількісні характеристики.

Методологія. У процесі дослідження найбільша увага приділялася ключовим аспектам первинної обробки даних про хімічний склад чавуну та методам описової статистики. Головною метою описової статистики є надання інформації про досліджувані дані в компактній та зрозумілій формі. Однак, перш ніж перейти до опису доступних даних, необхідно проаналізувати їх тип і характер розподілу, оскільки різні типи даних вимагають різних методів опису та обробки. Це також буде корисним при виборі відповідного статистичного методу для перевірки гіпотез.

Наукова новизна. На даний момент дослідники не дійшли єдиної думки про те, яким законом розподілу описуються дані про хімічний склад чавуну на випуску, крім цього автору невідомі наукові публікації, які докладно розкривають цю тему. У зв'язку з чим виникає необхідність у проведенні дослідження стохастичних властивостей реальних даних про хімічний склад чавуну на випуску доменної печі.

Ключові слова: типи даних, закон розподілу, описова статистика, нестационарні процеси.

Vladyslav SIDANCHENKO

Postgraduate Student of the Department of Information Security and Telecommunications, Dnipro University of Technology, 19 Dmytra Yavornytskoho ave., Dnipro, Ukraine, 49005, sidanchenko.vl.v@nmu.one

ORCID: 0000-0001-5581-9177

To cite this article: Sidanchenko, V. (2023). Perevirka kharakteru rozpodilu danykh pro khimichniy sklad chavunu na vypusku [Examination of the data distribution nature on the chemical composition of cast iron at the output]. *Information Technology: Computer Science, Software Engineering, and Cyber Security*, 3, 65–69, doi: <https://doi.org/10.32782/IT/2023-3-8>

EXAMINATION OF THE DATA DISTRIBUTION NATURE ON THE CHEMICAL COMPOSITION OF CAST IRON AT THE OUTPUT

Introduction. Ensuring the desired chemical composition of cast iron in manufacturing plays a crucial role in determining the quality of metallurgical products and important economic indicators of an enterprise. To achieve this goal, there is a need for the development and implementation of effective forecasting methods that play a key role in the optimal management of the cast iron smelting process to a specified quality.

Complex technological processes, such as blast furnace smelting, are influenced by various factors of different natures. These factors affect both the general course and development of the process, as well as its individual properties and quantitative characteristics.

Methodology. In the course of the research, the primary focus was on the key aspects of data preprocessing and descriptive statistics methods. The main purpose of descriptive statistics is to provide information about the investigated data in a concise and understandable form. However, before proceeding to describe the available data, it is necessary to analyze their type and distribution, as different types of data require different methods

of description and processing. This will also be useful when choosing an appropriate statistical method for hypothesis testing.

Scientific Novelty. At present, researchers have not reached a consensus on the distribution law that describes the data on the chemical composition of cast iron at the output. In addition, the author is unaware of scientific publications that thoroughly cover this topic. Hence, there is a need for conducting a study of the stochastic properties of real data on the chemical composition of cast iron in the output of a blast furnace.

Key words: data types, distribution law, descriptive statistics, non-stationary processes.

Аналіз попередніх досліджень. До характерних особливостей доменного виробництва належать:

- випадковий характер змін у часі фізичних та хімічних властивостей шихтових матеріалів;
- велика кількість чинників (зокрема неконтрольованих), які впливають на кінцевий результат доменної плавки.

Зазначені особливості зумовлюють необхідність проведення досліджень властивостей часових рядів, якими представлені результати хімічного аналізу чавуну на випуску. Такі дослідження необхідні для розробки рекомендацій зі створення методик прогнозування хімічного складу чавуну за умов діючого виробництва, адекватних характеру прогнозованого процесу (Гусев, Сіданченко, 2022, с. 24-31).

Як правило, часовими рядами є випадкові зміни величин, що дозволяють послідовно уявити еволюцію складних систем на основі отриманих даних (Boffetta, Cencini, Falconi, Vulpiani, 2002, с. 367-374). Найчастіше статистичний аналіз ґрунтується на припущенні, що досліджувана система є випадковою, тобто причинний процес, що створив часовий ряд, має багато складових частин або ступенів свободи. Взаємодія цих компонентів настільки комплексна, що детерміноване пояснення неможливе. При цьому об'єктом дослідження є клас моделей, які, як правило, відповідають класу випадкового гаусівського процесу. Однак, багато реальних часових рядів характеризуються інваріантністю щодо масштабних перетворень (властивість самоподібності), у зв'язку з чим стандартна гаусова статистика виявляється неспроможною і проблема дослідження часових рядів зводиться до аналізу стохастичних самоподібних процесів, які можуть бути описані фрактальними множинами (Mandelbrot, 2002; Feder, 1988).

Основне дослідження. Перед тим як розпочинати роботу з даними, завжди слід проводити перевірку розподілу. Під видом розподілу розуміють функцію, яка зв'язує значення змінної випадкової величини з ймовірністю їх появи в сукупності.

У статистичних дослідженнях найчастіше проводиться перевірка розподілу на нормаль-

ність. Під нормальним розподілом розуміють симетричний розподіл дзвоноподібної форми, при якому близько 68% даних відрізняється від середнього арифметичного не більше ніж на одне, а приблизно 95% – не більше ніж на два стандартні відхилення в кожну сторону. Незважаючи на те, що нормальний (гаусовий) розподіл зустрічається дуже часто і відіграє важливу роль у статистиці, існують і інші розподіли даних (біноміальний, Пуассона, Максвелла, Шарльє, та ін). Для перевірки існують графічні методи та статистичні критерії.

Дане дослідження проводилося у програмному середовищі Matlab, на основі реальних даних про відсотковий вміст кремнію в чавуні, які були отримані в різні моменти часу, на доменній печі № 3 (ДП-3) Маріупольського металургійного комбінату ім. Ілліча (ММК) (Сіданченко, Нікольська, 2023, с. 77).

На першому етапі дослідження було проведено тест Колмогорова – Смирнова для досліджуваних часових рядів, з використанням функції "ksstat", ця функція використовується для обчислення значення статистики Колмогорова – Смирнова (KS-статистики) при порівнянні двох вибірок або однієї вибірки з теоретичним розподілом. Статистика KS демонструє, наскільки добре емпірична функція розподілу (ЕФР) вибірки відповідає теоретичній функції розподілу (ТФР). Отримані рівні значущості склали $p = 0,00012$ і $p = 0,00004$ для часових рядів 1 і 2, дозволяють відкинути гіпотезу про відповідність даних, закону нормального розподілу, оскільки критичне значення рівня значущості становить 0,05.

До аналогічного висновку можна дійти на підставі результату візуального аналізу гістограм (рис. 1), де помітні відхилення від нормального розподілу, що демонструється у таких аспектах як:

1. Асиметрія: У нормальному розподілі дані симетричні щодо середнього значення. Відхилення вліво або вправо від симетрії свідчить про асиметрію даних.

2. Викиди: Відхилення від нормального розподілу проявляється у вигляді викидів – значень, які сильно відрізняються від інших у наборі даних.

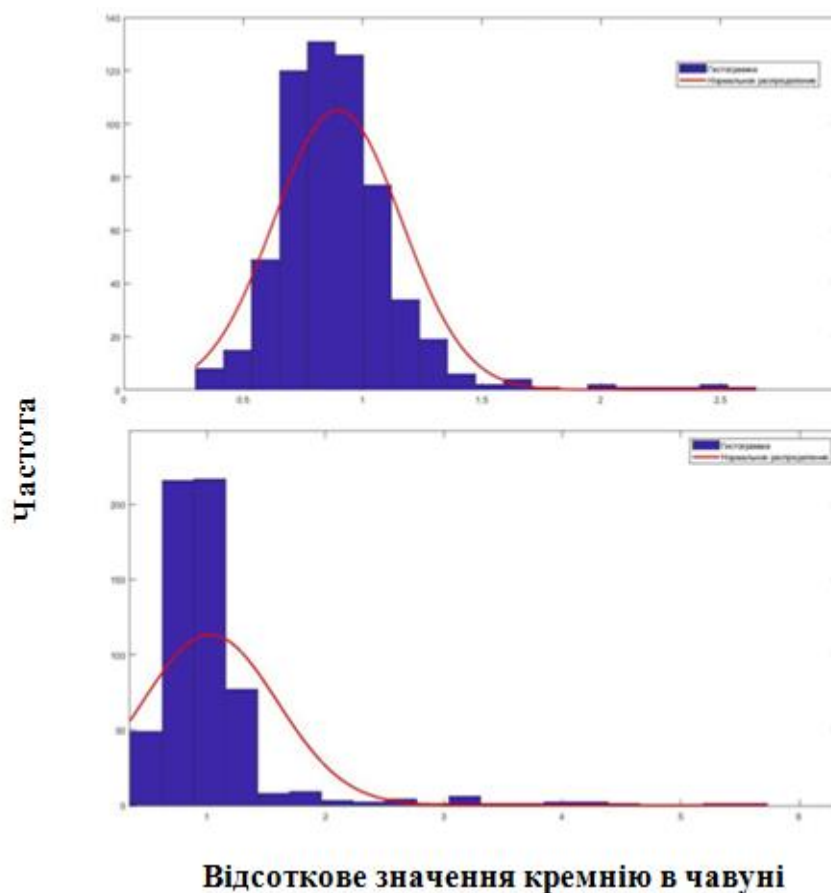


Рис. 1. Гістограми розподілу даних

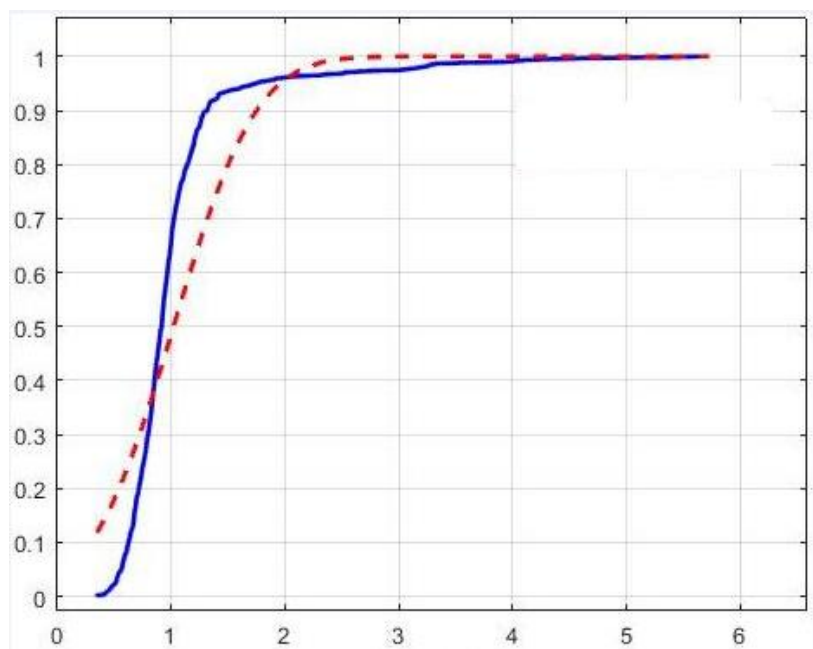


Рис. 2. Порівняння емпіричної функції розподілу та теоретичної функції розподілу

Крім того, з візуального аналізу графіків емпіричної функції розподілу (ЕФР – синя лінія) та теоретичної функції розподілу (ТФР – червона

пунктирна лінія) (рис. 2), помітна їх суттєва відмінність, що так само може вказувати на відхилення даних від нормального розподілу.

Незважаючи на те, що гістограма є добрим способом перевірки нормальності розподілу, більш чітку картину дають квантильні діаграми (рис. 3). У випадку нормального розподілу даних квантильна діаграма має вигляд прямої лінії. Будь-яке відхилення від прямої лінії свідчить про відхилення даних від нормальності.

З отриманих квантильних діаграм чітко видно істотне відхилення від лінії нормального розподілу (червона пунктирна лінія).

Висновки. Узагальнюючи отримані нетривіальні результати дослідження стохастичних

властивостей часових рядів, якими представлені результати хімічного аналізу чавуну на випуску можна зробити наступні висновки:

1. Експериментально підтверджено, що класична гіпотеза про нормальний (гаусівський) розподіл даних хімічного аналізу чавуну на випуску може вважатися необґрунтованою.

2. Аналізуючи отримані результати дослідження, можна дійти висновку про необґрунтованість класичних методів оцінки і прогнозу хімічного складу чавуну, які не можуть дати необхідного за своєю точністю результату,

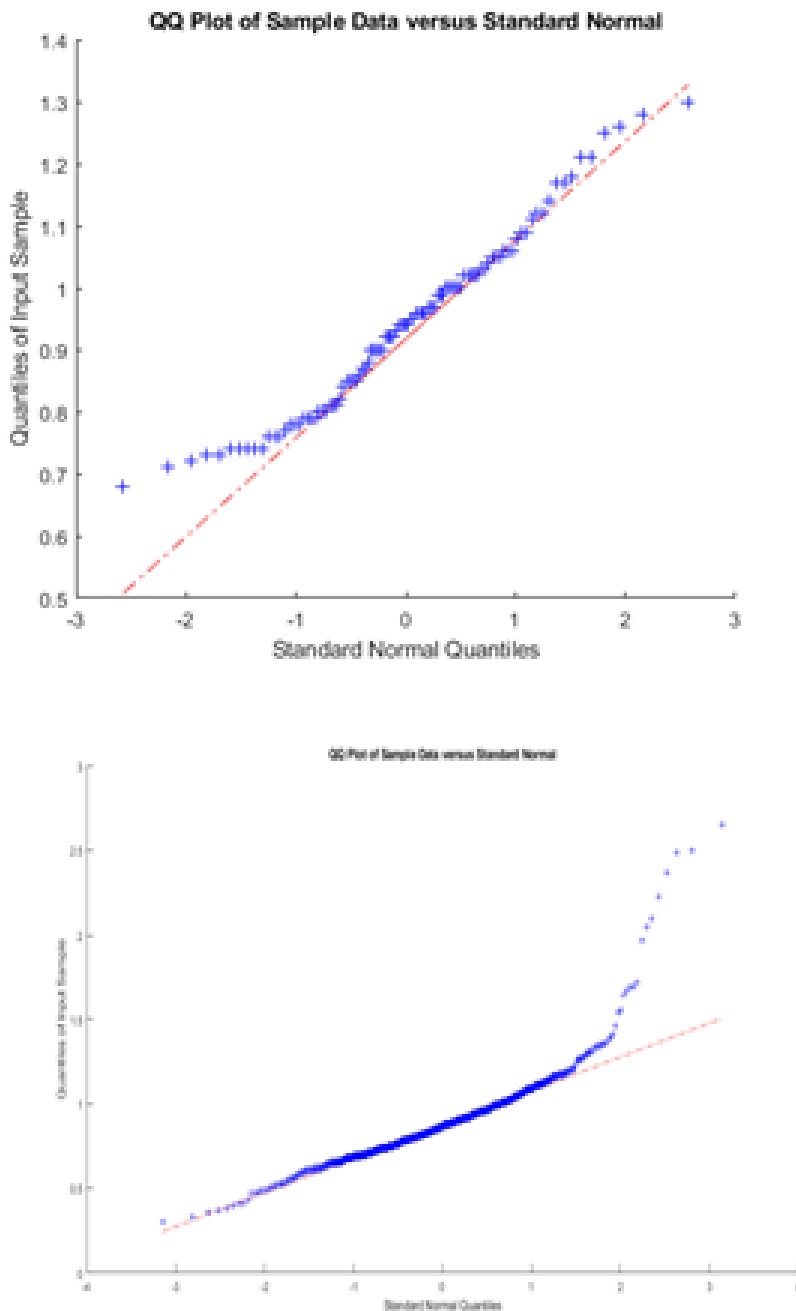


Рис. 3. Квантильні діаграми досліджуваних часових рядів

оскільки більшість із них, як згадувалося раніше, ґрунтуються на припущенні, що об'єктом дослідження є клас моделей, які відповідають класу випадкового (гаусівського) процесу.

3. Точні відомості про характер розподілу даних є важливим для вибору відповідних ста-

тистичних методів та моделей для подальшого аналізу. Оскільки дані не відповідають нормальному характеру розподілу, виникає необхідність альтернативного підходу до аналізу даних, з цілю побудови прогнозних моделей адекватних характеру досліджуваного процесу.

ЛІТЕРАТУРА:

1. Гусев О.Ю., Сіданченко В.В. Фрактальний аналіз реальних даних про хімічний склад чавуну на випуску доменної печі. *Information Technology: Computer Science, Software Engineering and Cyber Security*. 2022. №. 2. С. 24-31.

2. Boffetta G., Cencini M., Falconi M., Vulpiani A. Predictability: a way to characterize complexity // *Phys. Rep.* 2002. V. 356. P. 367-374.

3. Mandelbrot B. (2002). Fractal geometry of nature. *The institute of Computer Research*.

4. J. Feder. (1988). *Fractals*. Plenum Press, New York.

5. Сіданченко В.В., Нікольська О.І. Методи нелінійної динаміки в задачі прогнозування хімічного складу чавуну на випуску. *Information Technology: Computer Science, Software Engineering and Cyber Security*. 2023. №. 2. С. 76-83.

REFERENCES:

1. Gusev O.Yu., Sidanchenko V.V. (2022). Fraktalniy analiz realnykh danykh pro khimichniy sklad chavunu na vypusku domennoi pechi. [Fractal analysis of real data on the chemical composition of cast iron at the output of a blast furnace]. *Information Technology: Computer Science, Software Engineering and Cyber Security*. №. 2. P. 24-31 [in Ukrainian].

2. Boffetta G., Cencini M., Falconi M., Vulpiani A. (2002). Predictability: a way to characterize complexity // *Phys.Rep.* V. 356. P. 367-374.

3. Mandelbrot B. (2002). Fractal geometry of nature. *The institute of Computer Research*.

4. J. Feder. (1988). *Fractals*. Plenum Press, New York.

5. Sidanchenko V.V., Nikolska O.I. (2023). Metody neliniinoi dynamiky v zadachi prohozuvannia khimichnoho skladu chavunu na vypusku. [Methods of non-linear dynamics in the problem of forecasting the chemical composition of cast iron at the output]. *Information Technology: Computer Science, Software Engineering and Cyber Security*. №. 2. P. 76-83 [in Ukrainian].