

УДК 004.9

*Антонюк А. О., здобувачка вищої освіти,
Луценко А. В., д-р філософії з математики,
старший викладач, в. о. завідувача кафедри
прикладної математики та кібербезпеки*

ЗАСТОСУВАННЯ МАТЕМАТИЧНОЇ СТАТИСТИКИ В АНАЛІЗІ ДАНИХ

Донецький національний університет імені Василя Стуса, м. Вінниця

Анотація. У роботі йдеться про застосування математичної статистики в аналізі даних. Звернено увагу на її значущість у сучасному світі. Описано значення описової статистики, регресійного та кластерного аналізу, перевірки гіпотез і аналізу числових рядів. Та на основі цього підкреслено важливість математичної статистики для аналізу даних у сучасному світі.

Ключові слова: математична статистика, аналіз даних, регресійний аналіз, перевірка гіпотез, кластерний аналіз, аналіз числових рядів.

Вступ. Через стрімкий розвиток інформаційних технологій ми отримуємо доступ до великої кількості даних у різних сферах. Проте звичайне накопичення інформації не забезпечує її ефективність без застосування належних аналітичних методів. Дослідники використовують математичну статистику для отримання потрібної інформації з даних, прогнозування подій та оптимізації процесів.

Метою цієї роботи є вивчення методів математичної статистики, які використовуються для аналізу даних, і показ їх практичного значення в різних сферах діяльності.

Основний текст. Описова статистика. Основною метою описової статистики є опис набору даних. Вона використовує числові та графічні методи для пошуку закономірностей у наборі даних, узагальненні інформації і представленні її в зручній формі. Описовий статистичний аналіз використовує спеціальні інструменти для опису даних, до них належать частота, відсотки та показники центральної тенденції. Описова статистика дає змогу виявити загальні тенденції, перевірити гіпотези про розподіл даних та порівняти різні групи даних [1].

Регресійний аналіз. Регресійний аналіз – це статистичний метод, який використовується для моделювання зв'язку між залежною змінною та однією чи кількома незалежними змінними. Він широко застосовується в аналізі даних для визначення закономірностей, прогнозування та розуміння зв'язків між змінними. В аналізі даних регресійний аналіз використовується для визначення факторів, які впливають на зміни відповіді, і прогнозування майбутніх результатів. У фінансовому аналізі регресія використовується для оцінки впливу ринкових факторів на дохід компанії, а в медицині – для визначення впливу різних факторів на стан здоров'я людей [2].

Перевірка гіпотез. Перевірка гіпотез – це фундаментальна концепція математичної статистики, яка передбачає перевірку гіпотези про сукупність на основі вибірки даних. Вона широко використовується в аналізі даних, щоб

приймати обґрунтовані рішення та робити висновки щодо населення. Під час аналізу даних перевірка гіпотез використовується, щоб визначити, чи є результати випадковими, чи вони відображають реальний ефект [3].

Кластерний аналіз. Кластерний аналіз – це статистичний метод, який використовується для групування схожих об’єктів або випадків у кластери на основі їх характеристик. Він широко використовується в аналізі даних для виявлення шаблонів і структур у даних. Під час аналізу даних кластерний аналіз використовується для визначення базових сегментів у даних і прогнозування поведінки нових випадків. До того ж кластерний аналіз використовується для сегментації клієнтів у маркетингу чи виявлення шахрайства у фінансовій сфері [4].

Аналіз числових рядів. Аналіз числових рядів – це статистичний метод, який використовується для аналізу та прогнозування даних, зібраних за певний час. Математична статистика використовується для визначення закономірностей, тенденцій і аномалій у даних числових рядів, що дає змогу аналітикам робити прогнози та приймати обґрунтовані рішення. Аналіз числових рядів є важливим в аналізі даних, зокрема у фінансах та економіці допомагає досліджувати динаміку цін та акцій і прогнозувати економічні показники [5].

Висновки. Математична статистика відіграє ключову роль в аналізі даних і широко застосовується в різних галузях. Завдяки її методам, як-от описова статистика, регресійний та кластерний аналіз, перевірка гіпотез, існує можливість ефективно працювати з великими обсягами даних, виявляти закономірності та прогнозувати події.

Сучасні технології підвищили важливість статистичних методів, зробивши їх доступними для використання у різних сферах. Використання математичної статистики допомагає приймати обґрунтовані рішення, оптимізувати процеси та знаходити нові методи для вирішення складних задач.

Отже, застосування математичної статистики в аналізі даних є невід’ємною частиною сучасного інформаційного суспільства, що сприяє розвитку багатьох сфер людської діяльності.

Список використаних джерел

1. Applied A. Z. Statistics: Basic Principles and Application. *International Journal of Innovation and Economic Development*. 2021. № 7(3). P. 27–33.
2. Shahab Araghinejad Regression-Based Models Dordrecht: Springer, 2014. P. 49–83.
3. Sheldon M. Ross Chapter 8 – Hypothesis Testing, 2014. P. 297–356.
4. Voges K. E. Cluster Analysis Using Rough Clustering and k-Means Clustering. IGI Global, 2005. P. 435–438.
5. Velicer W. F., Fava J. L. Time Series Analysis. John Wiley & Sons, 2003. P. 581–606.