

UDC 004(92)+631:3.06(075.8)

Yu. O. Sotnykov, Cand. Sci. (Econ.), Assistant Professor

D. V. Gavva, Cand. Sci. (Agric.)

K. B. Novosad, Cand. Sci. (Agric.), Assistant Professor

S. V. Resnik, Postgraduate Student

*Kharkiv National Agrarian University named after V. V. Dokuchaev,
Kharkiv, e-mail: yesaul.halt@knau.kharkov.ua, pochvoved@ukr.net*

VISUALIZATION MULTIDIMENSIONAL DATA IN THE EXPERIMENT OF AGRONOMICAL SOIL SCIENCE

The possibilities of using visualization and analysis of multidimensional data obtained as a result of the agronomic experiment are considered. Effective methods for rapid and efficient data analysis with the help of visualization methods are demonstrated, in particular, using the "Chernov's face" method. A step-by-step algorithm is considered for constructing a visual image of the elements of the soil formation system.

The main task of visualization is its practical purpose, namely, facilitating the perception of data for their further analysis, for example, obtaining patterns of influence of parameters of the system under study. In this sense, visualization is essentially a support for decision making. Therefore, there is an inextricable connection between the application and the method of visualization. For visualization, both primary data and pre-processed data can be used.

The main idea of such a method is based on the ability of a person to "automatically" record complex relationships between many variables if they appear in a sequence of elements. Intuitive pattern recognition of a person from childhood on a subconscious level. Sometimes the understanding that some elements "something" are similar to each other comes earlier than an analyst can explain what variables determine this similarity, that is, the analysis of information using such a reflection method is based on the ability of a person to intuitively find similarities and differences in the features of object – face.

The algorithm for constructing visual representations is based on the methods of cognitive analysis and the construction of a cognitive image, the main purpose of which is the allocation of new knowledge from already existing visual images.

It is proved that the application of the "Chernov's face" method in the process of analyzing the data of the agronomic experiment is effective at all stages of scientific research.

Keywords: *visualization of multidimensional data, diagram, pictogram, "Chernov's faces", Statgraphics, cognitive graphics, soil.*

УДК 004(92)+631:3.06(075.8)

Ю. А. Сотников, канд. экон. наук, доцент

Д. В. Гавва, канд. с.-х. наук

К. Б. Новосад, канд. с.-г. наук, доцент

С. В. Резник, аспирант

*Харьковский национальный аграрный университет им. В. В. Докучаева
г. Харьков e-mail: yesaul.halt@knau.kharkov.ua, pochvoved@ukr.net*

ВИЗУАЛИЗАЦИЯ МНОГОМЕРНЫХ ДАННЫХ В ЭКСПЕРИМЕНТЕ АГРОНОМИЧЕСКОГО ПОЧВОВЕДЕНИЯ

Рассмотрены возможности использования визуализации и анализа многомерных данных, полученных в результате проведения агрономического эксперимента. Продемонстрировано эффективные способы быстрого и эффективного анализа данных с помощью методов визуализации, в частности – с помощью метода «лица Чернова». Пошагово рассмотрен алгоритм построения визуального образа элементов системы почвообразования. Доказано, что применение метода «лица Чернова» в процессе анализа данных агрономического эксперимента является эффективным на всех этапах научного исследования.

Ключевые слова: визуализация многомерных данных, диаграмма, пиктограмма, «лица Чернова», Statgraphics, когнитивная графика, почва.

УДК 004(92)+631:3.06(075.8)

Ю. О. Сотников, канд. экон. наук, доцент

Д. В. Гавва, канд. с.-г. наук

К. Б. Новосад, канд. с.-г. наук, доцент

С. В. Резник, аспирант

*Харківський національний аграрний університет ім. В. В. Докучаєва
м. Харків e-mail: yesaul.halt@knau.kharkov.ua, pochvoved@ukr.net*

ВИЗУАЛІЗАЦІЯ БАГАТОМІРНИХ ДАНИХ В ЕКСПЕРИМЕНТІ АГРОНОМІЧНОГО ҐРУНТОЗНАВСТВА

Розглянуто можливості використання візуалізації та аналізу багатомірних даних, отриманих у результаті проведення агрономічного експерименту. Продемонстровано ефективні способи швидкого та ефективного аналізу даних за допомогою методів візуалізації, зокрема –

за допомогою методу «обличчя Чернова». Покроково розглянуто алгоритм побудови візуального образу елементів системи ґрунтоутворення, що аналізується. Доведено, що застосування методу «обличчя Чернова» в процесі аналізу даних агрономічного експерименту є ефективним на всіх етапах наукового дослідження

Ключові слова: візуалізація багатомірних даних, діаграма, піктограма, обличчя Чернова, Statgraphics, когнітивна графіка, ґрунт.

Вступ. Після первинної систематизації даних наукового експерименту (Сотников Ю. О., 2017) дуже важливо зробити перший крок у правильному напрямку досягнення поставленої мети. І тут можуть стати в нагоді інтуїтивно зрозумілі прийоми та методи когнітивної графіки (*Cognitive graphics*), що дають можливість, по-перше, наочно побачити існування потенційної відповіді на поставлені запитання. По-друге, отримати підказку для вирішення проблеми (Du Toit С.Н.С., 1986; Spencer Neil Н., 2014). На сьогоднішній день існує досить багато програмних продуктів, у яких вдало реалізовані методи та прийоми когнітивної графіки, серед найбільш розповсюджених – Statsoft Statistica, SPSS, JMP SAS, Origin, Statgraphics та ін. Використовуючи реманент когнітивної графіки, що реалізовані в цих програмних продуктах, дослідник, маючи мінімальні знання в галузі математики та програмування, може вирішити одне з найважчих завдань дослідження – ідентифікація та класифікація об'єктів.

Одним з найкращих методів реалізації такого підходу і є візуалізація даних, що дозволяє охопити створений масив даних «одним поглядом, але залишаючи можливість за необхідності сконцентруватися на окремих його елементах» (Yau Nathan, 2011). Одним з найбільш розповсюджених інструментів комп'ютерної реалізації цієї властивості є так званий метод «обличчя Чернова» (*Chernoff faces*) – уявлення багатомірних об'єктів у вигляді піктографіків. «Цей метод не є особливо точним і непідготовленого користувача він може ошелешити. Однак в деяких ситуаціях «обличчя Чернова» можуть бути дуже корисними, та й людям, які багато працюють з даними, застосування цього методу часто приносить масу задоволення, так що овоїти його безперечно варто» (Yau Nathan, 2011).

Інформація з навколишнього середовища, яка сприймається людиною, викликає в нього певні емоції. У людини виявлено шість таких універсальних станів емоцій: смуток, гнів, радість, страх, огида і здивування. Кожна емоція відображається на обличчі, тому виразу обличчя є надійним індикатором емоційного стану людини. Брови, очі і рот є головними елементами на обличчі, за допомогою яких виражаються й орієнтуються емоційні стани людини. Інформація отримана на основі таких емоції послуговує підвалиною для прийняття рішення. Однак, крім об'єктивної інформації людина має потребу і в суб'єктивній інформації, тобто емоцій. Вона може виходити від людей, з якими людина контактує, або синтезованих емоцій, які виробляються індикаторами емоцій у технічній системі у вигляді так званих «обличчя Чернова» (Bernhard

Flury, 1981).

Першим аспектом завдання візуалізації є її практичне призначення, а саме: полегшення сприйняття даних для їх подальшого аналізу, наприклад, отримання закономірностей впливу параметрів досліджуваної системи. У цьому сенсі візуалізація по суті є підтримкою прийняття рішень. Тому є нерозривний зв'язок між прикладною задачею і способом візуалізації. Для візуалізації можуть використовуватися і первинні дані, і дані, що пройшли попередню обробку. Це другий аспект візуалізації. Його слід відрізнити від третього аспекту – розробки власне програмних алгоритмів і кодів для реалізації візуалізації на засобах обчислювальної техніки (Романова І., 2016).

Головна ідея такого методу заснована на здатності людини «автоматично» фіксувати складні зв'язки між багатьма змінними, якщо вони проявляються в послідовності елементів. Інтуїтивне розпізнавання образів властиве людині з дитинства на підсвідомому рівні. Для кожного спостереження викреслюється «обличчя», де відносні значення обраних змінних представлені як форми і розміри окремих рис: довжина носа, кут між бровами, ширина обличчя та ін. (Davenport Thomas H., 2013).

Іноді розуміння, що деякі елементи «чимось схожі» один на одного, приходить раніше, ніж аналітик може пояснити, які саме змінні обумовлюють цю подібність, тобто аналіз інформації за допомогою такого способу відображення заснований на здатності людини інтуїтивно знаходити подібності та відмінності в рисах об'єкта – обличчя (рис. 1).

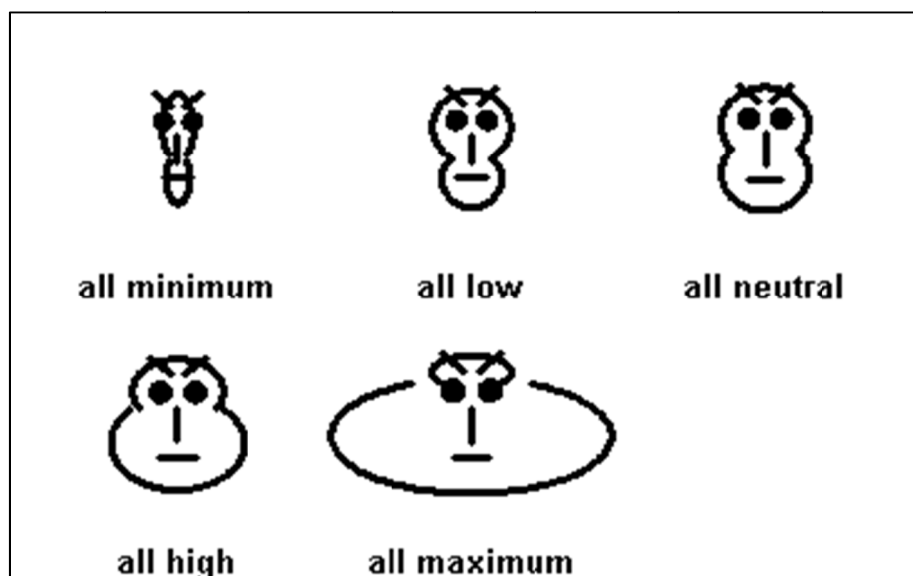


Рис. 1. Ключі рис «облич Чернова»

Алгоритм побудови візуальних представлень заснований на методиках когнітивного аналізу та побудові когнітивного образу, основна мета яких – виділення нового знання із вже існуючих візуальних образів.

Методика когнітивного аналізу складних ситуацій містить такі етапи:

1) формулювання завдання та цілі дослідження;

- 2) вивчення виробничого процесу із позицій поставленої мети;
- 3) збір, систематизація та аналіз існуючої статистичної та якісної (повної, чіткої) інформації проблеми;
- 4) виділення основних ознак, взаємозв'язків процесу дослідження та визначення дій основних об'єктів, законів розвитку досліджуваної ситуації (це уможливить виділити об'єктивні залежності й тенденції в процесах);
- 5) визначення основних вимог, цілей і обмежень ситуації;
- 6) виділення основних суб'єктів, їх інтересів – це дасть змогу окреслити можливі зміни в розвитку ситуації;
- 7) визначення шляхів, механізмів дії, реалізації виробничих інтересів основних суб'єктів, що уможливить у подальшому визначити стратегії поведінки.

Метод «обличчя Чернова» застосовують, як правило, у двох випадках:

- коли потрібно виявити характерні залежності або групи спостережень (при цьому цей метод використовують для класифікації спостережуваних даних аналогічно кластерному аналізу);
- коли необхідно дослідити складні взаємозв'язки між кількома змінними (Маковецька С., 2015).

Метою досліджень є візуалізація та опис даних експерименту з агрономічного ґрунтознавства за допомогою методу «обличчя Чернова».

Об'єктами дослідження було обрано масив даних, що характеризують властивості ґрунтів південно-східного Лісостепу України та напрям їх еволюційних змін залежно від використання.

Методика досліджень полягала в аналізі ефективності інструментів візуалізації експериментальних даних на підготовчому етапі дослідження.

Результати та обговорення. Розглянемо застосування методу «обличчя Чернова» для аналізу та ідентифікації зразків ґрунту чорнозему типового південно-східного Лісостепу України. Попередньо було сформовано масив даних, що характеризують 36 зразків ґрунту за 15 хімічними, фізичними, фізико-хімічними показниками, а також ознаками, що характеризують мікробіологічну активність чорноземних ґрунтів.

Далі, за допомогою реманенту Chernoff faces (Plot – Multivariate Visualization) створюємо відповідні піктограми облич кожного зразка ґрунту (рис. 2). Які висновки можна зробити на основі цих піктограм «на перший погляд»?

1. Існує істотна різниця в «емоціональності облич» між деякими групами зразків. Існує можливість продовження дослідження, шляхом виділення визначення та виділення цих груп.

2. Зразки *значно* відрізняються між собою обрисами та розмірами облич. Це повинно наштовхнути на думку, що на різницю між зразками ґрунту істотно впливають не всі параметри, що досліджуються, *а лише ключові* з них.

Таким чином, існує можливість та необхідність стиснення масиву даних

та за рахунок цього отримати більш стислі, аргументовані, «рельєфні висновки».



Рис. 1. «Обличья Чернова» – візуалізація досліджуваних варіантів ґрунтів залежно від їх властивостей

Шляхом наступного проведення *факторного аналізу* визначмо чинники, що визначають різницю в зразках ґрунту (методика проведення факторного аналізу не входить до мети цього дослідження – наведемо лише висновки).

За результатами дослідження ознаки були стиснені в чотири групи чинників:

1) чинник азотно-низькотрофної акумуляції (мікроскопічні гриби, актиноміцети, оліготрофи та олігонітрофіли);

2) чинник кислотності ґрунту (рНводний, рНсольовий, Ca^{2+} , Mg^{2+} , гідролітична кислотність, мікроскопічні гриби);

3) чинник органогенної акумуляції (вміст гумусу, Сфк, Na^+ , K^+ , мікроорганізми, що асимілюють органічні та мінеральні форми азоту);

4) чинник азотно-багатотрофної акумуляції (вміст гумусу, Сгк, Сфк, Ca^{2+} , мікроскопічні гриби, гетеротрофи – мікроорганізми, що асимілюють органічні та мінеральні форми азоту).

Отже, отримано *чотири* групи чинників, що, скоріш за все і характеризують (визначають) таксономічні (класифікаційні) групи ґрунтів чорнозему типового залежно від їх використання.

Наступним кроком є визначення ключових параметрів, що характеризують цю різноманітність. За результатами проведення дискримінантного аналізу отримуємо, що на диференціацію чорноземів типових найбільше впливають лише 7 з 15 показників, що досліджуються, (відповідно за зменшенням величини R^2):

1) груповий склад гумусу (Сгк, Сфк);

2) кислотність ґрунту (рНводний), обмінні катіони (Ca^{2+} , Mg^{2+} , Na^{2+} , K^{+}).

Інші досліджувані показники суттєво не впливають на диференціацію подібного типу ґрунтів. У результаті виконаної дискримінантної диференціації достовірно класифіковано 94,87 %

Таким чином виділено лише сім *ключових ознак* (Сгк, Сфк, Ca^{2+} , Mg^{2+} , Na^{2+} , K^{+}), які також визначають таксономічні (класифікаційні) групи ґрунтів.

Отримані попередні результати про наявність взаємозв'язків кількісних показників складу ґрунту і варіантів його використання є підставою для поглибленого пошуку закономірностей. З метою підтвердження висунутої нами гіпотези та виділення результуючих ознак в однорідні групи за впливом на зміни показників чорноземів типових було використано метод ієрархічного кластерного аналізу. Кластеризацію (виділення однорідних груп) проводили за методом Варда, використовуючи метрику Сіті-блок для оцінки міжкластерних відстаней. Кількість ключових ознак – сім (за результатами дискримінантного аналізу). Кількість кластерів – 4 (за результатами факторного аналізу) (табл. 1).

1. Центроїди груп, отриманих за результатами кластерного аналізу

Кластер	Сгк	Сфк	рНводний	Ca^{2+}	Mg^{2+}	Na^{+}	K^{+}
1	1,20	0,72	6,81	34,49	5,72	0,24	1,21
2	0,87	0,43	7,39	33,53	4,88	0,25	0,73
3	0,83	0,33	6,59	32,91	6,15	0,37	0,53
4	0,88	0,38	7,07	38,01	4,39	0,35	0,42

Тепер знову настав час повернутися до використання процедури «обличчя Чернова» за показниками табл. 1 (рис. 3).

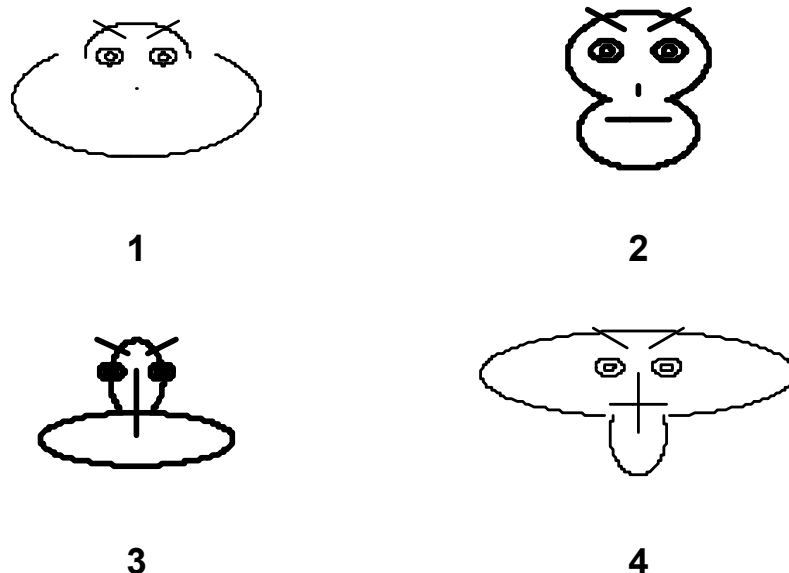


Рис. 3. Результати класифікації ґрунтів за допомогою «обличчя Чернова»

Як бачимо, рис. 3 істотно відрізняється від рис. 1, який став більш компактним та зрозумілим за рахунок проведення операцій стиснення даних

(проведених факторного, кластерного та дискримінантного аналізів). Це дозволило більш точно визначити групи ґрунтів за сьома ключовими ознаками:

$S_{\text{гк}}$ – радіус до кута обличчя,

$S_{\text{фк}}$ – кут нахилу горизонтальний,

$r_{\text{Нводний}}$ – вертикальний розмір обличчя,

Ca^{2+} – ексцентриситет верхньої частини обличчя,

Mg^{2+} – ексцентриситет нижньої частини обличчя,

Na^+ – довжина носа,

K^+ – вертикальне положення рота

та чотири типи ґрунтів, що відповідає результатам факторного аналізу.

Отримані зображення (рис. 3) у подальшому доцільно використовувати як еталон, для швидкої ідентифікації окремих зразків ґрунту відповідно до отриманої класифікації.

Висновки. Продемонстровано можливість візуальної інтерпретації статистичних даних агрономічного експерименту «Обличчя Чернова».

Розглянуто алгоритм побудови візуального образу елементів системи ґрунтоутворення, що аналізується. Доведено, що застосування методу «Обличчя Чернова» в процесі аналізу даних агрономічного експерименту є ефективним на всіх етапах наукового дослідження.

СПИСОК ЛІТЕРАТУРИ / REFERENCES

Сотников Ю. О. Систематизація даних агрономічного експерименту за допомогою електронних таблиць / Ю. О. Сотников, Д. В. Гавва // Вісник ХНАУ. Сер. «Ґрунтознавство, агрохімія, землеробство, лісове господарство, екологія ґрунтів». – Харків: ХНАУ, 2017. – № 1. – С. 96-103.

Sotnykov Yu. O., Gavva D. V., 2017, "Systematization of agronomic experiment data using spreadsheets», Visnyk KHNAU ser. "Soil science, agrochemistry, agriculture, forestry, soil ecology", Kharkiv, KHNAU, Vol 1, pp. 96-103.

Романова И. К. Современные методы визуализации многомерных данных: анализ, классификация, реализация, приложения в технических системах / И. К. Романова // Наука и Образование. – Электрон. журн. МГТУ им. Н. Э. Баумана. – 2016. – № 3. – С. 133–167.

Romanova I. K., 2016, "Modern Methods of Multidimensional Data Visualization: Analysis, Classification, Implementation and Applications in Technical Systems, Science and Education of the Bauman MSTU, Vol. 3, pp. 133–167.

Маковецька С. Візуальна інтерпретація технологічних процесів цукрового виробництва з використанням методу «обличчя Чернова» / С. Маковецька, О. М'якшило // Товари і ринки. – 2015. – № 2(20). – С. 156–167.

Maikovetska S., Myakshylo O., 2015, "Visual interpretation of technological processes of sugar production using Chernoff face method", Commodities and Markets, Vol. 2, pp. 156–167.

Spencer Neil H., 2014, "Essentials of Multivariate Data Analysis", CRC Press, NY, 180 p.

Du Toit C.H.C., Steyn A.G.V., Stumph R.H., 1986, "Graphical Exploratory Data Analysis", Springer-Verlag, 322 p.

Yau Nathan, 2011, "Visualize This: The FlowingData Guide to Design, Visualization, and Statistics", Wiley Publishing Inc., 384 p.

Davenport Thomas H., Kim Jinho, 2013, "Keeping Up with the Quants: Your Guide to Understanding and Using Analytics", Harvard Business Press, 240 p.

Flury B., Riedwyl H., 1981, "Graphical Representation of Multivariate Data by Means of Asymmetrical Faces", Journal of the American Statistical Association, Vol. 76, No. 376, pp. 757-765.