

РОЗРОБКА УЗАГАЛЬНЕНОГО МЕТОДУ АНАЛІЗУ ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ ДЛЯ АДЕКВАТНОГО ВИБОРУ СТАТИСТИЧНОГО МЕТОДУ

Статтю присвячено проблемі підвищення якості навчання майбутніх інженерів-педагогів комп'ютерних технологій статистичного опрацювання експериментальних даних. Визначено, що для реалізації вимог державних стандартів вищої освіти методична система навчання дисципліни "Комп'ютерні технології статистичного опрацювання експериментальних даних" має розроблятися на основі способу адекватного вибору статистичного методу аналізу. Теоретично обґрунтовано та розроблено узагальнену ознакову модель та метод аналізу експериментальних даних.

Ключові слова: Інженер-педагог, методична система навчання комп'ютерних технологій статистичного опрацювання експериментальних даних, ознакова модель експериментальних даних.

Форм 7. Рис 7. Літ 10.

Постановка проблеми. Стрімкий розвиток сучасної освіти, економіки, техніки, призводить до необхідності переосмислення підходів до процесу навчання студентів інженерно-педагогічних спеціальностей аналізу і розв'язання задач з різних предметних галузей за допомогою статистичних методів з використанням програмних засобів опрацювання експериментальних даних. При удосконаленні підходів навчання необхідно зробити наголос на процедурному представленні процесу аналізу експериментальних даних, що дозволяє розробляти, оптимізувати і реалізувати складні динамічні послідовності розв'язання задач з психолого-педагогічної, економічної та технічної галузей діяльності інженера-педагога.

Розвиток процедурного підходу спирається на постійно зростаючу необхідність виділення і представлення, позиціонування і взаємодії складових ознак даних умови задачі при їх описі і використанні в типових, програмованих процедурах. Це пов'язано з ускладненням процесу аналізу даних, виконанням вимог формалізації і прозорості процесів, складанням, адаптацією і впровадженням процедурних моделей, які реалізуються на логіко-процедурній основі.

Незважаючи на ускладнення аналізу експериментальних даних в умовах задач з різних предметних галузей, а багато в чому саме завдяки йому, ключовим аспектом моделювання стає необхідність поглиблення системного і комплексного підходів до застосування ознакової моделі аналізу даних. Особливого значення це набуває при практичному опрацюванні експериментальних даних перед застосування відповідних комп'ютерних технологій, зокрема програмних засобів Statistica, SPSS, MS Excel та вибором необхідних і адекватних статистичних методів аналізу, які охоплюють усі сфери діяльності інженерів-педагогів. Отже, побудова процедурної моделі необхідна для забезпечення гнучкості, модульності, можливості змін, досліджень і коригувань процедур аналізу експериментальних даних. У праці Гуржія [2] зазначається, що вибір методів аналізу для опрацювання даних для конкретного об'єкта становить певну складність, викликану необхідністю аналізу та оцінки його ефективності в залежності від класу об'єкту та його статистичних характеристик.

Аналіз останніх досліджень та публікацій. Теоретичне підґрунтя цієї роботи склали науково-практичні дослідження Галазунова А.Т. [1], Донеллі-мл. Р. [3], Дорошенка Ю.О. та Ротаєнка П.А. [4], Жалдака М.І., Кізьмінової Н.М. та Михаліна Г.О. [5], Захожай В.Б. та Чорного А.Ю. [6], Лазарєва М.І. [7], Наследова А.Д. [8], Мармози А.Т. [9], Томашевського О.В. [10].

Постановка завдання. На нашу думку, одним із перспективних шляхів аналізу експериментальних даних в умові задачі є розробка процедурної моделі у вигляді алгоритму. Тому метою нашого дослідження є розробка процедурної моделі аналізу експериментальних даних в умовах задач з різних предметних галузей діяльності майбутнього інженера-педагога, на основі раніше створеної ознакової моделі аналізу експериментальних даних, яка становитиме основу адекватного вибору статистичного методу опрацювання експериментальних даних.

Виклад основного матеріалу. Розглянемо процес аналізу експериментальних даних в умові задачі. На основі раніше розробленої узагальненої ознакової моделі аналізу експериментальних даних в умовах задач (рис. 1), яка має таку структуру: встановлення факторної та результативної змінних, опис даних на основі шкал виміру, визначення закону розподілу, якому підпорядковуються дані.

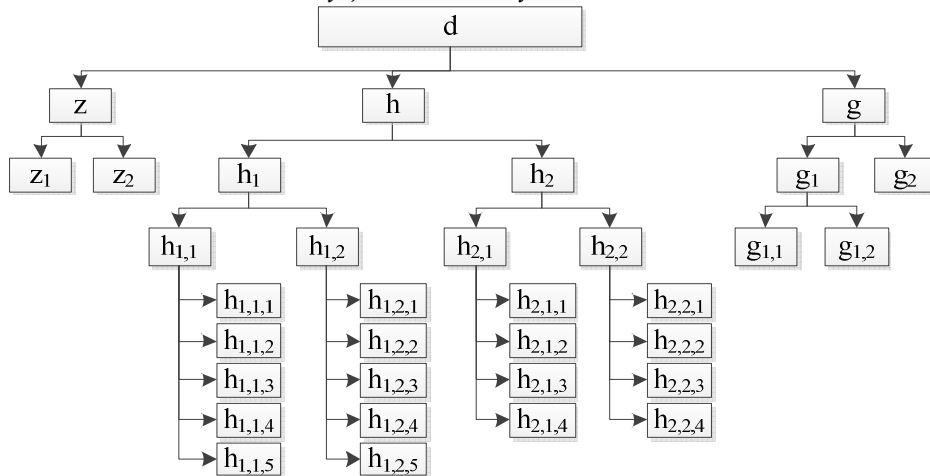


Рис. 1. Узагальнена ознакова модель аналізу експериментальних даних в умовах задач Авторська розробка

В даній моделі:

- d – експериментальні дані в умові задачі;
- $z = \{z_1, z_2\}$ – модель ознак, які представляють факторну і результативну ознаки;
- $h = \{h_1, h_2\}$ – модель ознак, які представляють якісні і кількісні змінні;
- $h_1 = \{h_{1,1}, h_{1,2}\}$ – модель ознак якісних даних;
- $h_2 = \{h_{2,1}, h_{2,2}\}$ – модель ознак кількісних даних;
- $h_{1,1} = \{h_{1,1,1}, h_{1,1,2}, h_{1,1,3}, h_{1,1,4}, h_{1,1,5}\}$ – модель ознак шкали найменувань;
- $h_{1,2} = \{h_{1,2,1}, h_{1,2,2}, h_{1,2,3}, h_{1,2,4}, h_{1,2,5}\}$ – модель ознак порядкової шкали;
- $h_{2,1} = \{h_{2,1,1}, h_{2,1,2}, h_{2,1,3}, h_{2,1,4}\}$ – модель ознак інтервальної шкали;
- $h_{2,2} = \{h_{2,2,1}, h_{2,2,2}, h_{2,2,3}, h_{2,2,4}\}$ – модель ознак шкали відношень;
- $g = \{g_1, g_2\}$ – модель ознак, які представляють закон розподілу даних;
- $g_1 = \{g_{1,1}, g_{1,2}\}$ – модель ознак нормального закону розподілу.

Розв'язання задач з різних галузей діяльності майбутнього інженера-педагога слід починати з визначення зв'язків між змінними, в залежності від постановки умови задачі та особливостей предметних галузей, так як змінні можуть змінювати свій вплив, результативна змінна може стати факторною, а факторна – результативною. При цьому слід бути дуже уважним, неправильний вибір напрямку змінних приведе до безглузвих результатів.

Встановлення зв'язку між змінними, як в умові задачі, так і при виборі статистичного методу аналізу слід починати з визначення факторної змінної – z_1 , якщо сукупність являється однорідною (одна чи декілька ознак, що вивчаються, є загальними для всіх одиниць, і різнорідна сукупність об'єднує явища різного типу), то факторна змінна завжди наявна. Якщо сукупність є різнорідною то факторну змінну визначити не можливо, необхідно провести уточнення умови задачі. У переважній більшості задач з усіх предметних галузей експериментальні дані є однорідними і ми можемо ввести z_1 до ознакової моделі даних задачі.

Аналогічно визначаємо наявність результативної змінної в даних задачі – z_2 . Якщо результативна ознака наявна в експериментальних даних умови задачі, то її необхідно включити до ознакової моделі, що репрезентує зв'язки між даними.

В результаті визначення факторної та результативної ознак даних в умові задачі, отримуємо таку модель:

$$z(X,Y)=\{z_1, z_2\}.$$

Процес визначення факторної і результативної змінних можна представити у вигляді процедурної моделі (алгоритму) з використанням відповідних позначень ознак даних (рис. 2).

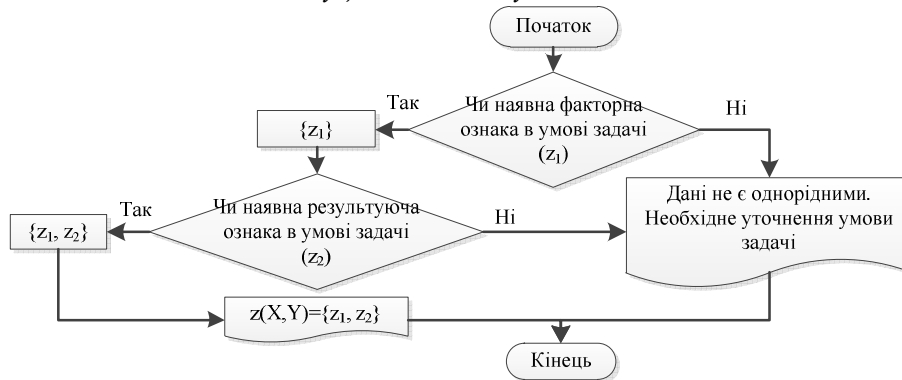


Рис. 2. Процедурна модель визначення факторної і результуючої змінних
Авторська розробка

Отже, розроблена процедурна модель дозволяє визначити наявність змінних в задачі та встановити їх роль та в подальшому співставити дані ознаки умови задачі з відповідними виділеними ознаками статистичних методів.

Проаналізуємо змінну X , яка наявна в умові задачі та визначимо характер типу даних до якого вона належить відповідно до ознакової моделі аналізу даних умови задачі на основі шкал виміру.

Для того, щоб побудувати процедурну модель необхідно детально перевірити кожну ознаку шкали виміру та встановити тип шкали і визначити тип даних (кількісні- h_2 або якісні- h_1 дані).

Розглянемо шкалу найменувань ($h_{1,1}$) та проаналізуємо ознаки, які її характеризують. Дослідимо ознаку $h_{1,1,1}$ (відсутність числових характеристик), якщо дані змінної X відрізняються одне від одного тільки якісною характеристикою (фізика, математика, хімія – предмети, що вивчаються учнями) то вони відповідають даній ознаці. Якщо ця умова виконується то переходимо до розгляду наступної ознаки $h_{1,1,2}$ (неможливо проводити з даними математичні операції крім співставлення (дорівнює «=», не дорівнює « \neq »)), в протилежному випадку необхідно перейти до розгляду ознак порядкової шкали ($h_{1,2}$). Тобто будь-яка не відповідність ознак шкали виміру даних задачі призводить до переходу і аналізу наступної шкали виміру.

Розглянемо випадок, коли ознака $h_{1,1,1}$ шкали найменувань виконується і переходимо до дослідження ознаки $h_{1,1,2}$. Якщо з даними змінної X неможливо виконувати математичні операції крім співставлення (математика \neq фізика, хімія = хімія) то умова даної ознаки виконується. Переходимо до наступної ознаки цієї ж шкали виміру – $h_{1,1,3}$ (в номінальній шкалі досліджуваним об'єктам приписуються певні літерні або кодові (літерно-цифрові) значення), наприклад дані змінної X представлені: математика (літерне) або 1А (літерно-цифрове) або 4-11-23 (кодове).

Проаналізуємо наступну ознаку шкали $h_{1,1} - h_{1,1,4}$ (неможливо впорядкувати об'єкти ознаки та відразу визначити їх кількість). Перевіряємо змінну X на наявність цієї ознаки, якщо в даних не можливо визначити більший або менший елементи то дані відповідають цій ознаці (номер футболіста в команді, шкільні предмети).

Розглянемо ознаку $h_{1,1,5}$ (дані приймають лише два протилежних значення (дані виміряні в дихотомічній шкалі)), дана ознака є підмножиною номінальної шкали. Наприклад: ліворукий - праворукий; стать: чоловіча - жіноча; 0 або 1.

Проаналізувавши всі ознаки номінальної шкали виміру і з'ясувавши, що змінна X задовольняє всім ознакам цієї шкали та належить до якісного типу даних. Можемо записати таку декларативну модель:

$$h_1 = h_{1,1} = \{h_{1,1,1}, h_{1,1,2}, h_{1,1,3}, h_{1,1,4}, h_{1,1,5}\}.$$

Записану декларативну модель можна представити у вигляді процедурної моделі з використанням відповідних позначень ознак даних (рис. 3).

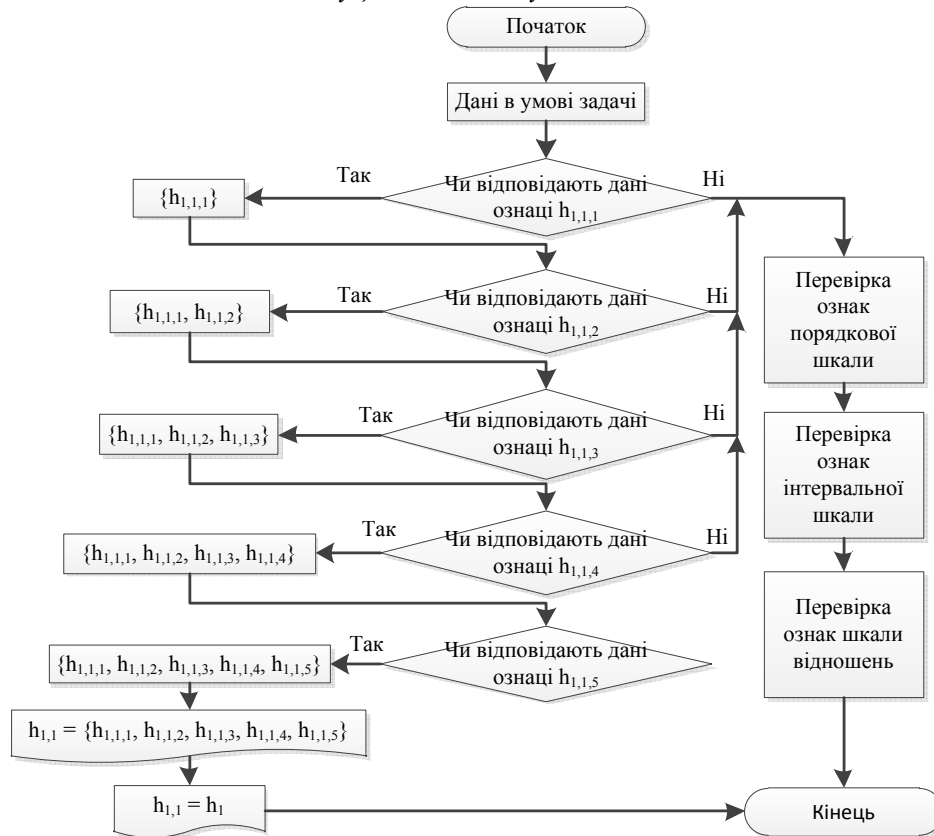


Рис. 3. Процедурна модель аналізу змінної X на основі ознак номінальної шкали
 Авторська розробка

Аналіз змінної X на відповідність ознакам порядкової шкали здійснюється за аналогічним алгоритмом. Декларативна модель для порядкової шкали матиме наступний вигляд:

$$h_1 = h_{1,2} = \{h_{1,2,1}, h_{1,2,2}, h_{1,2,3}, h_{1,2,4}, h_{1,2,5}\}$$

Тобто, перевіряється наявність ознак, які належать до даної моделі:

- $h_{1,2,1}$ – відсутність числових характеристик, які мають одиницю виміру.
- $h_{1,2,2}$ – можливе здійснення порівняння та співставлення об'єктів за величиною ознаки (дорівнює «=», не дорівнює «≠», більше «>», менше «<»).
- $h_{1,2,3}$ – можливе приписування об'єктам певних літерних, кодових або числових значень.
- $h_{1,2,4}$ – дані можливо впорядкувати (за зростанням або спаданням значень даних).
- $h_{1,2,5}$ – дані можливо представити у вигляді рангів (проранжувати).

Записану декларативну модель можна представити у вигляді процедурної моделі з використанням відповідних позначень ознак даних (рис. 4).

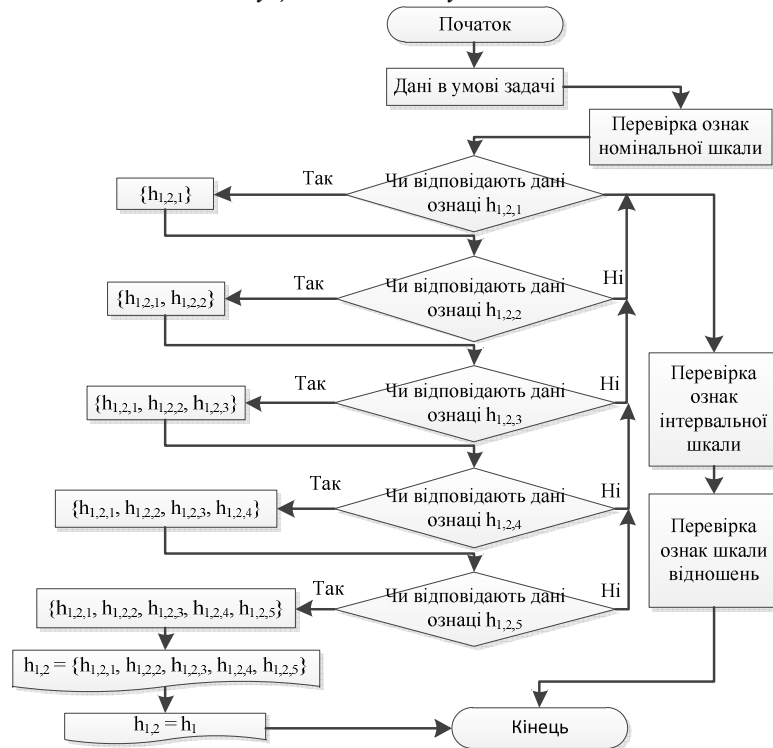


Рис. 4. Процедурна модель аналізу змінної X на основі порядкової шкали
 Авторська розробка

Якщо змінна X не відповідає ознакам порядкової шкали переходимо до аналізу цієї змінної на основі ознак кількісних типів даних, до яких відносяться інтервальна шкала та шкала відношень.

Перші ознаки ($h_{2,1,1}$ та $h_{2,2,1}$), які характеризують ці дві шкали виміру (дані виражені кількісно і мають одиницю виміру) – співпадають, тому їх аналіз здійснюється одночасно. Якщо дані змінної X в умові задачі виражені за допомогою числових значень то можемо ввести $h_{2,1,1}$ та $h_{2,2,1}$ до ознакової моделі даних задачі і перейти до розгляду наступних ознак.

Ознаки $h_{2,1,2}$ та $h_{2,2,2}$ (з динамі можливо проводити порівняння в кількісній мірі (дорівнює «=», не дорівнює « \neq », більше «>», менше «<»)) також співпадають і свідчать про можливість здійснення логічних операцій з даними змінної X. Якщо ці ознаки виконуються то можна перейти до розгляду наступної ознаки.

Розглянемо ознаку $h_{2,1,3}$ (відсутня точка абсолютного відліку) яка виконується тільки для інтервальної шкали виміру ($h_{2,1}$). Якщо в даних змінної X відсутня точка абсолютного відліку то ознаку $h_{2,1,3}$ можемо ввести до ознакової моделі даних задачі. У протилежному випадку переходимо до розгляду шкали відношень ($h_{2,2}$) і вводимо в ознакову модель даних задачі ознаку $h_{2,2,3}$ (наявна точка абсолютного відліку).

Повертаємось до інтервальної шкали виміру даних ($h_{2,1}$) і перевіряємо ознаку $h_{2,1,4}$ (над даними можна виконувати наступні математичні операції: додавання «+», віднімання «-»). Якщо умова ознаки виконується то відносимо змінну X до інтервальної шкали і кількісного типу даних.

Отже, можемо записати таку модель:

$$h_2 = h_{2,1} = \{h_{2,1,1}, h_{2,1,2}, h_{2,1,3}, h_{2,1,4}\}.$$

Розглянемо етап коли виконується ознака $h_{2,2,3}$, в даних наявна точка абсолютного відліку та перейдемо до розгляду наступної ознаки шкали відношень ($h_{2,2}$) – $h_{2,2,4}$ (над даними можна виконувати наступні математичні операції: додавання «+», віднімання «-», множення «*», ділення «/»). Якщо умова ознаки виконується то відносимо змінну X до шкали відношень і кількісного типу даних.

Отже, можемо записати таку модель:

$$h_2 = h_{2,2} = \{h_{2,2,1}, h_{2,2,2}, h_{2,2,3}, h_{2,2,4}\}.$$

Отримані моделі представимо у вигляді процедурної моделі аналізу кількісних даних та вимірюються в інтервальній шкалі або шкалі відношень (рис. 5).

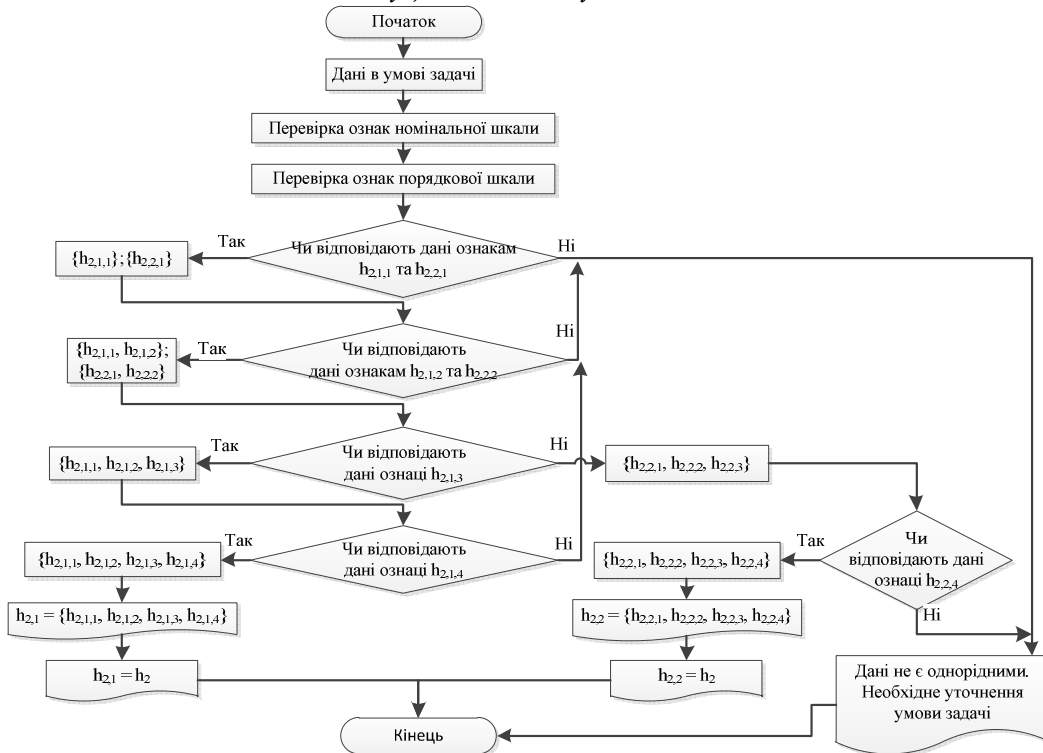


Рис. 5. Процедурна модель аналізу змінної X на основі ознак інтервальної шкали та шкали відношень
 Авторська розробка

На основі вищерозглянутих процедурних моделей побудуємо узагальнену модель аналізу даних змінної X на основі ознак типів шкал виміру (рис. 6).

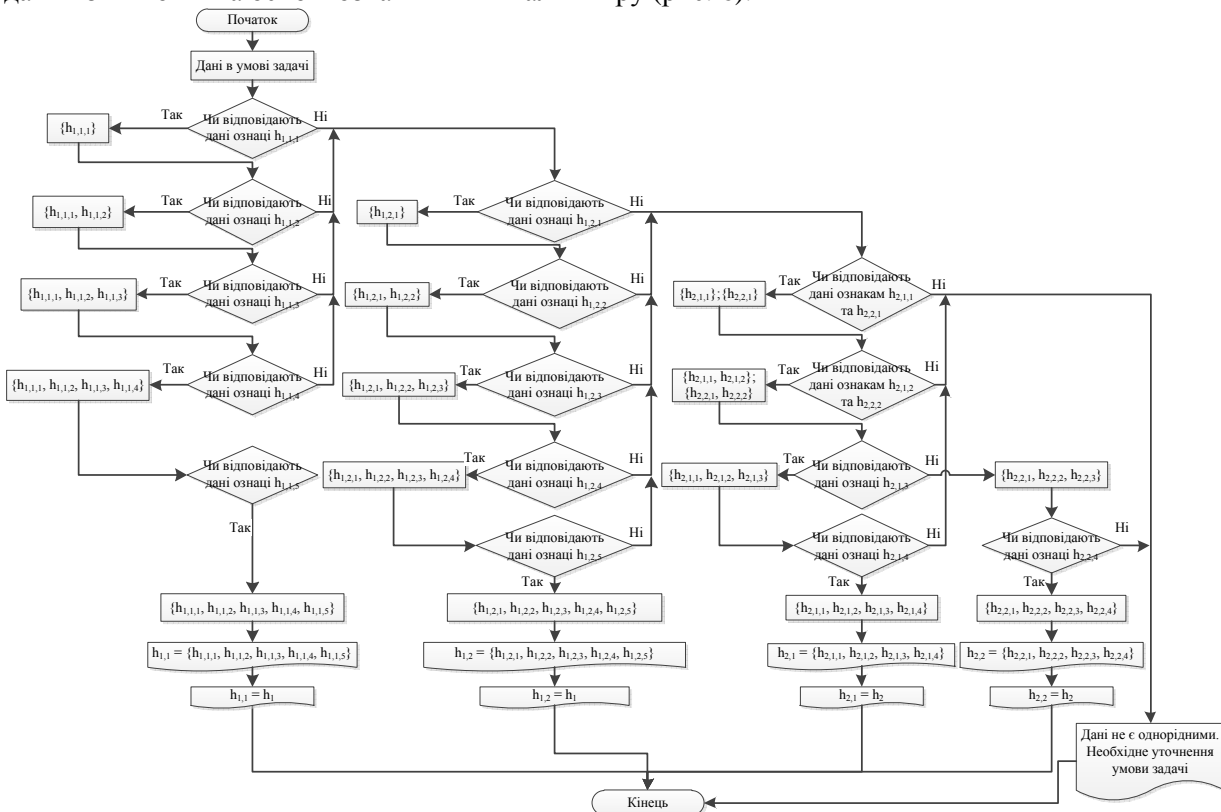


Рис. 6. Узагальнена процедурна модель аналізу змінної X на основі ознак шкал виміру
 Авторська розробка

Аналіз змінної Y в умові задачі виконується за аналогічною процедурною моделлю на основі ознак шкал виміру.

Представимо наступний етап аналізу даних у умові задачі на основі ознак закону розподілу (g).

Аналіз закону розподілу слід починати з визначення кількості даних в умові задачі (ознака $g_{1,1}$). Якщо кількість даних перевищує 25 спостережень то можна вважати, що дані підпорядковуються нормальному закону розподілу то одержуємо декларативну модель:

$$g(X, Y) = g_1 = \{g_{1,1}\}$$

Якщо кількість спостережень є меншою за 25, тобто ознака $g_{1,1}$ не виконується, то необхідно розглянути ознаку $g_{1,2}$ (побудова нормального ймовірнісного графіку). У випадку коли побудований нормальний ймовірнісний графік засвідчує що розподіл даних підпорядковуються нормальному закону то одержуємо наступну декларативну модель:

$$g(X, Y) = g_1 = \{g_{1,2}\}.$$

В протилежному випадку дані не підпорядковуються нормальному закону розподілу і декларативна модель набуває наступного вигляду:

$$g(X, Y) = g_2.$$

Процес визначення закону розподілу даних можна представити у вигляді процедурної моделі з використанням відповідних позначень ознак (рис. 7)

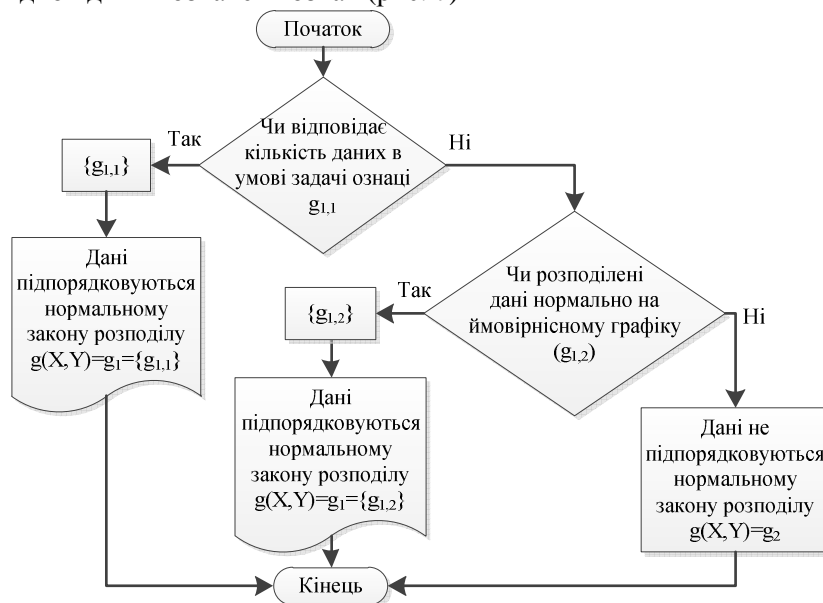


Рис. 7. Процедурна модель визначення закону розподілу, якому підпорядковуються змінні в умові задачі
 Авторська розробка

Висновки. Розглянувши всі етапи аналізу та розробивши окремі процедурні моделі для найбільш поширених варіантів представлення даних в умовах задач нами розроблена узагальнена процедурна модель у вигляді алгоритму аналізу експериментальних даних на основі відповідних ознак.

Після детального, обґрунтованого, коректного аналізу даних умови задачі студентам необхідно перейти до вибору статистичного методу аналізу експериментальних даних. Існує дуже велика кількість статистичних методів аналізу експериментальних даних які будуть необхідними у майбутній діяльності інженерів-педагогів.

На основі виділених ознак аналізу даних умови відповідної задачі та детального проходження кожного етапу алгоритму, студенти зможуть обрати адекватний метод аналізу експериментальних даних, співставивши ознакову модель даних умови задачі та ознакові моделі статистичних методів аналізу які ми плануємо розробити в подальших дослідженнях.

1. Галазунов А.Т. Педагогические исследования: содержание, организация и обработка результатов / А.Т. Галазунов. – Издательский центр АПО: – М., 2003. – 41 с.
2. Гуржій А. М. Електричні і радіотехнічні вимірювання: посіб. для пед. працівників та учнів проф.-техн. навч. закл. / Гуржій А.М., Поворознюк Н.І. – К. : Навчальна книга, 2002. – 267 с. : рис.
3. Донелли-мл. Р. Статистика / Роберт А. Донелли-мл.; пер. с англ. Н.А. Ворониной. – М. : Астрель: АСТ, 2007. – XIV, 367, [3] с. : ил.
4. Дорошенко Ю.О. Достовірність комп'ютерного тестування : Навч.-метод. посіб. / Ю.О. Дорошенко, П.А. Ротаєнко. За ред. Ю.О. Дорошенка. – К. : педагогічна думка, 2007. – 176 с.
5. Жалдак М.І. Теорія ймовірностей і математична статистика: підручник для студентів педагогічних університетів / Жалдак М.І., Кузьміна Н.М., Михалін Г.О. – Полтава-Довкілля: – К., 2009. – 498 с.
6. Захожай В.Б. Статистичне забезпечення управління якістю : навч. посіб. / В.Б. Захожай, А.Ю Чорний. – К. : Центр навчальної літератури, 2005. – 340 с.
7. Лазарєв М.І. Полісистемне моделювання змісту технологій навчання загальноінженерних дисциплін : Монографія / М.І. Лазарєв. – Х. : Вид-во НФаУ, 2003. – 356 с.
8. Наследов А. Д. SPSS: компьютерный анализ данных в психологии и социальных науках / Наследов Андрей Дмитриевич; 2-е изд. – СПб. : Питер, 2007. – 416 с. : ил.
9. Мармоза А.Т. Теорія статистики : Навч. посіб. ВНЗ / А.Т. Мармоза. – К. : Ельга, Ніка-Центр, 2003. – 392 с.
10. Томашевський О. В. Комп'ютерні технології статистичної обробки даних : навч. посібник для студ. вищ. навч. закладів, які навчаються за спец. «Якість, стандартизація та сертифікація» / О. В. Томашевський, В. П. Рисіков. – Запоріжжя : Запорізький національний технічний ун-т, 2006. – 174 с.