

УДК 378.147
Л.В.Павленко
Бердянський державний педагогічний університет

МЕТОДИКА СТВОРЕННЯ ОЗНАКОВИХ МОДЕЛЕЙ АНАЛІЗУ ПСИХОЛОГО-ПЕДАГОГІЧНИХ, ЕКОНОМІЧНИХ ТА ТЕХНІЧНИХ ДАНИХ В УМОВАХ ЗАДАЧ

У статті обґрунтована та розроблена узагальнена ознакова модель аналізу даних умови задачі для опрацювання статистичної інформації майбутніми інженерами-педагогами з різних предметних галузей їх майбутньої професійної діяльності.

Постановка проблеми. Статистичний аналіз масових явищ і процесів є необхідною ланкою в системі управління економікою та державою в цілому.

Передусім за допомогою статистичних даних здійснюється «зворотний зв'язок», тобто потік інформації йде від об'єкта до суб'єкта управління – керівництва підприємств, об'єднань, територіальних, галузевих і центральних органів влади. Без вірогідної, всебічної і своєчасної інформації ефективні управлінські рішення неможливі. В даний час існує ряд спеціальних комп'ютерних програм за допомогою яких можливо провести статистичний аналіз даних, але жодна з програм не може визначити структуру даних, ведення даних в спеціальну комп'ютерну програму здійснюється після попереднього проведення аналізу статистичних даних (визначення структури даних і враховуючи вимоги програми). Підготовчий та заключний етап аналізу даних не здатна виконати жодна комп'ютерна програма - їх дослідник виконує сам [5, с. 14]

Вибір методу аналізу виходячи зі статистичних даних в умові задачі є досить складним етапом дослідження і цьому студентів необхідно навчати, а для цього необхідні спеціальні методичні розробки які на даний час відсутні. Тому в даному дослідженні зупинимось на глибокому аналізі даних в залежності від постановки задачі дослідження.

На сучасному етапі розвитку технологій студентам інженерно-педагогічних спеціальностей доводиться мати справу з даними та інформацією про них яка отримуються з різних джерел їх майбутньої професійної діяльності, насамперед психолого-педагогічних, економічних та технічних галузей [1]. Застосування тих чи інших статистичних методів до аналізу статистичних даних відповідних предметних галузей є глибокою **проблемою**, перш за все тому що існує дуже велике різноманіття статистичних методів і залишається не розробленою методика адекватного вибору статистичного методу відповідно до тих даних з якими студент буде мати справу в своїй професійній діяльності.

Аналіз останніх досліджень та публікацій. Існують ряд методик в яких йде мова про педагогічні дані, або про економічні, або технічні дані. Зокрема проблемою аналізу психолого-педагогічних даних займались Гроленко Г.Л., Захожай В.Б., Єфімова М.Р., Лугініна О.Є., Білоусова С.В., напрямку статистичного аналізу економічних явищ і процесів приділяли увагу такі автори Равікевич Є.І., Білова Н.Ф., Елесеєва І.І., Юзбашенва М.М., Волкова А.Г., Лисенко С.Н., Дмитреєва І.А., Мармоза А.Т. В галузі статистичного аналізу технічних даних працювали такі автори: Моторин Р.М., Чекотовський Е.В., Опря Я.Г., Захожая В.Б., Акімова О.В., Маркевич О.В., Бек В.Л., Гончарук А.Г., в даних джерелах аналіз даних і вибір методу описуються фрагментарно та неповно і враховуються особливості тільки якоїсь однієї предметної галузі, але студентам інженерно-педагогічних спеціальностей в їхній практичній діяльності доведеться мати справу з даними з усіх трьох предметних галузей, тому ми ставимо на **меті** розробити таку методику яка буде допомагати їм легко орієнтуватися в виборі статистичного методу аналізу даних з будь-якої предметної галузі враховуючи специфіку кожного типу даних. Адже на даний час така методика відсутня, раніше цією проблемою ніхто не займався. Отже звернемо нашу увагу на взагалі існуючі підходи до визначення типу даних та спробуємо розробити шляхи представлення навчального матеріалу та методів навчання, які будуть направлені на раціональний вибір представлення навчального матеріалу та методів навчання з виділенням в ньому основної базової та додаткової частини. Які дозволять нам перейти на якісно новий рівень підготовки майбутніх інженерів - педагогів.

Виклад основного матеріалу. Дані являються основою статистики [7, с. 14]. Достовірність будь якого статистичного дослідження самим тісним чином пов'язана з достовірністю даних, причому на самому початку етапу дослідження.

Аналіз даних включає ряд необхідних кроків які студент має здійснити перш чим перейти до розв'язання конкретної задачі з застосуванням певного типу статистичного аналізу. Кожна змінна має значення для дослідження, вона є основою, яка відрізняє об'єкти один від одного. Дані можливо досить просто визначити як значення присвоєне певному спостереженню або виміру. Якщо ми маємо справу з даними про активність студента на занятті то можливо підрахувати скільки раз студент відповів протягом 1 години 20 хвилин, можливо підрахувати тривалість кожної відповіді в секундах, можливо також вимірити змістовність відповіді за допомогою опису «Змістовна відповідь», «не перевершено», «без активний». В кожному випадку ми маємо справу з даними про одне й те ж явище, але в різній формі. В першому випадку вимірюється частота або кількість випадків. В другому випадку тривалість або часовий відрізок. А в третьому випадку описується явище по змістовності з застосуванням слів а не цифр. Кожен з випадків демонструє як можна представити дані. Дані являються основним компонентом усіх статистичних досліджень [7, с. 14] Вибір способу вимірювання даних на початковому етапі дослідження визначає які статистичні методи ми будемо використовувати.

Дані можуть класифікуватися декількома способами А.Т.Мармоза розглядає дані як рад специфічних властивостей, відмінних особливостей. Одна з таких особливостей є наявність варіацій. Варіюючі ознаки поділяються на атрибутивні(якісні) та кількісні [3, с.8].

Класифікувати дані можливо за допомогою різних **способів їх виміру**.

О.Є.Лугінін, С.В. Білоусова розглядають ознаки вимірювання, за допомогою типу шкал. Вони приводять таку класифікацію шкал ознак: номінальна, яка встановлює шкалу найменувань; порядкова, яка встановлює відношення подібності і послідовності; матрична, де за допомогою звичайних чисел вимірюються явища, ресурси, результати господарсько-фінансової діяльності [2, с. 20].

А.Т. Опря вважає, що обробка даних статистичних спостережень залежить від характеру статистичного виміру явищ, які вивчаються, і мети аналізу, яка їм надає тієї чи іншої форми. І виділяє три категорії завдань аналізу які включають певну систему методів: 1. Встановлення закону розподілу. 2.Перевірка статистичних гіпотез, яка є логічним продовженням попередньої. 3.Оцінка невідомих параметрів різних розподілів [6, с.25].

Р.М. Моторін, Е.В. Чекатовський зазначають, що завдання статистичного дослідження полягає у відшуканні узагальнюючих показників і встановленні закономірностей суспільного життя, які виявляються лише в певній множині. Вивчення матеріалу вони розпочинають з одиниць сукупності, кожен елемент сукупності характеризується низкою ознак, значення яких змінюється від одного періоду до іншого. Ознаки поділяються на кількісні та атрибутивні(описові, словесні, щодо яких існують відповідні типи шкал). Кількісна ознака може бути виражена числом. Кількісні ознаки можуть бути дискретними або неперервними. До дискретних належать ознаки, що набувають лише конкретних, ізольованих значень (кількість кімнат в квартирі, кількість членів родини). Неперервні ознаки мають будь-які значення в певних межах варіації (сума податків, урожайність). Шкала це засіб упорядкування та кількісного вираження ознак. Виділяють такі типи шкал: Номінальна, порядкова, метрична. Вибір одиниці вимірювання залежить від природи, матеріального змісту явища, конкретних завдань дослідження та практичної доцільності [4, с.7]

З проаналізованого матеріалу можна зробити висновок, що за допомогою розглянутих способів не можливо провести змістовний аналіз даних. Кожним автором підходи розглядаються частково, фрагментарно, неповно та відсутній необхідний комплексний, системний підхід до аналізу даних. Отже **проблемою** дослідження є теоретичне обґрунтування побудови моделі вибору методів аналізу на основі коректного аналізу досліджуваних даних та **розробка моделей**, які дозволять нам перейти на якісно новий рівень аналізу даних враховуючи специфіку кожної предметної галузі та вибору статистичних метод аналізу

Визначимо шляхи аналізу даних які, на нашу думку, будуть коректними, обґрунтованими, та зрозумілими для студентів інженерно-педагогічних спеціальностей і дозволять їм адекватно здійснювати аналіз даних та на його основі обрати необхідний статистичний метод для аналізу.

На першому етапі визначаємо факторну й результативну ознаки, які позначимо через – z. Ці ознаки підрозділяється на такі:

z_1 – незалежні дані (X – факторна ознака) – обирається дослідником;

z_2 – залежні дані (Y- результативна ознака) – вимірюється або реєструється дослідником.

Отже, можемо сформулювати модель причинно-наслідкового зв'язку між змінними (рис. 1):

$$z = \{z_1, z_2\}.$$

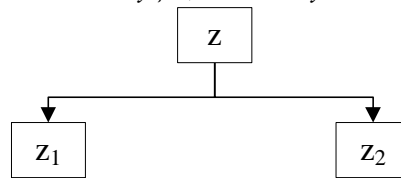


Рис. 1. Модель ознак причинно-наслідкових зв'язків в умові задачі

На другому етапі слід виділити два **типи даних**: кількісні та якісні. Кількісні дані використовують числові значення для опису досліджуваного об'єкту. До їх допустимих значень відносяться: цифри перед якими може стояти знак плюс або мінус і десятковий роздільник (у вигляді крапки або коми). Числа можуть містити максимальну кількість знаків (значення які містять більше десяти розрядів, округляються). Всі допустимі числові значення, включаючи експоненційне представлення, про яке свідчить наявність в числі букви Е або Д, а також знак плюс або мінус. Наприклад вік студентів досліджуваної групи 18,19,20.

Якісні дані використовують описовий вираз для виміру або класифікації об'єкта який нас цікавить. До допустимих значень відносяться: букви, слова, спеціальні символи. Наприклад високий, красивий, дешевий.

Шкала найменувань і порядкова шкала – шкали якісних ознак. Шкали кількісних ознак – це шкали інтервалів та відношень.

Залежність ознак виміру, тобто всі варіанти даних які можуть бути в умові задачі і підлягають аналізу позначимо через h . Дані які мають якісний характер позначимо через h_1 , а кількісні дані – h_2 .

Отже, можемо сформувати перший рівень моделі ознак даних (рис. 2):

$$h = \{h_1, h_2\}.$$

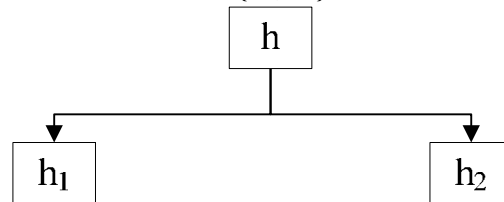


Рис. 2. Перший рівень моделі ознак даних

На третьому етапі пропонуємо виділити аналіз даних в залежності від ознак виміру (типу шкали).

Розглянемо дані які вимірюються в шкалі **найменувань**. Так вимірюється, наприклад, наявність інтересу до певного виду діяльності: 1-математика, 2-фізика, 3-література, 4-біологія, 5-техніка, 6-спорт.... Віднесені до шкали найменувань об'єкти можна розміщати в будь-якій послідовності в залежності від мети дослідження. До цієї шкали відносяться матеріали у яких досліджувані об'єкти відрізняються один від одного за їх якістю. У номінальній шкалі різним об'єктам приписуються певні числові чи літературні значення в залежності від назв об'єктів (тобто числа використовуються лише як мітки, наприклад, номери телефонів). З даними вимірними у цій шкалі проводити математичні операції не коректно. Простий випадок номінальної шкали є дихотомічна шкала, дані вимірні в цій шкалі можуть приймати лише два значення (стать чоловіча, жіноча; праворукий або лівша).

На основі характеристики шкали найменувань сформуємо елемент моделі ознак виміру в умові задачі. З метою побудови такої моделі за допомогою якої, на основі ознак даних, студенти зможуть змістовно, логічно, проаналізувати дані і як наслідок коректно і адекватно вибирати статистичний метод.

Дані які вимірюються в шкалі найменувань позначимо через $h_{1,1}$. А ознаки які їх описують відповідно $h_{1,1,1}, h_{1,1,2}, h_{1,1,3}, \dots, h_{1,1,n}$.

Визначимо ознаки, які характеризують дану шкалу.

$h_{1,1,1}$ – відсутність числових характеристик.

$h_{1,1,2}$ – неможливо проводити з даними математичні операції крім співставлення (дорівнює «=»), не дорівнює « \neq »).

$h_{1,1,3}$ – в номінальній шкалі досліджуваним об'єктам приписуються певні літерні або кодіві (літерно-цифрові) значення.

$h_{1,1,4}$ – неможливо впорядкувати об'єкти ознаки та відразу визначити їх кількість.

$h_{1,1,5}$ – дані приймають лише два протилежних значення (дані виміряні в дихотомічній шкалі).

Отже, вигляд моделі для шкали найменувань набуває наступного виду (рис. 3):

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

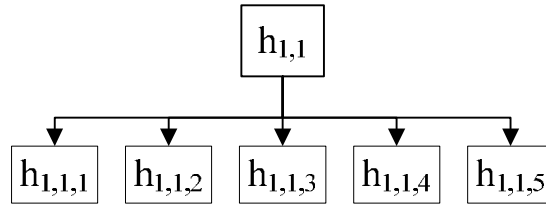


Рис. 3. Модель ознак, що репрезентують шкалу найменувань

Розглянемо дані які вимірюються в **порядковій** шкалі. Якщо в шкалі найменувань розміщення деяких об'єктів практично не грає ніякої ролі, то в шкалі порядку це видно з її назви – саме на цю послідовність переключается вся увага. Типовим прикладом порядкової шкали є шкільна оцінка: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 в балах. Різниця між «12» і «11» не така сама, як між «3» і «4», крім того, у різних учителів підходи до оцінки дещо різні. Можна сказати, що у цьому випадку одиниця вимірювання не постійна. **Порядкова шкала** використовується у тому випадку, коли певна ознака може проявлятися в більшій або меншій мірі, але важко сказати на скільки. За цією шкалою можна здійснювати упорядкування об'єктів за величиною певної ознаки (так вимірюються: знання учнів, кваліфікація вчителів, менше, ніжче...). Окремим випадком порядкової шкали є **рангова шкала**, яка передбачає приписування об'єктам певних числових значень (рангів) на основі їх упорядкування за ступенем прояву певної ознаки. Причому необхідно дотримуватись таких вимог: кількість рангів повинна дорівнювати кількості об'єктів; однаковим об'єктам (тим, які неможливо диференціювати) приписуються однакові ранги.

Рангову шкалу віднесемо до порядкової шкали і розглядатимемо її, як ознаку порядкового виміру.

Дані які відносяться до порядкової шкали позначимо через $h_{1,2}$, ознаки які характеризують даний вид виміру позначимо через $h_{1,2,1}, h_{1,2,2}, \dots, h_{1,2,n}$.

Визначимо ознаки, які характеризують дану шкалу.

$h_{1,2,1}$ – відсутність числових характеристик, які мають одиницю виміру.

$h_{1,2,2}$ – можливе здійснення порівняння та співставлення об'єктів за величиною ознаки (дорівнює «=», не дорівнює « \neq », 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).

Отже, вигляд моделі для порядкової шкали набуває наступного виду (рис. 4):

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

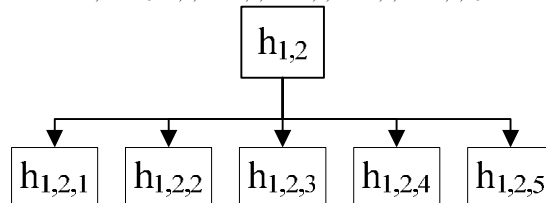


Рис. 4. Модель ознак, що репрезентують порядкову шкалу

Перейдемо до розгляду **інтервальної шкали** в якій одиниця вимірювання стала. В даній шкалі передбачається зіставлення величини досліджуваної ознаки з певним стандартним інтервалом, який приймають за одиницю міри (так вимірюють, наприклад, швидкість читання учнем тексту, за шкалою інтервалів вимірюють величину потенційної енергії або координату точки на прямій, на якій не відмічені ні початок, ні одиниця виміру: вага (1 кг), довжина (1м), кількість помилок, кількість виконаних завдань і т.д.). З наведених ознак легко можна зрозуміти що даний вид шкали застосовується в тих випадках де наявна одиниця виміру.

Дані які відносяться до інтервальної шкали позначимо через $h_{2,1}$, ознаки які характеризують даний вид виміру позначимо через $h_{2,1,1}, h_{2,1,2}, h_{2,1,3}, \dots, h_{2,1,n}$.

Визначимо ознаки, які характеризують дану шкалу.

$h_{2,1,1}$ – дані виражені кількісно і мають одиницю виміру.

$h_{2,1,2}$ – з динамі можна проводити порівняння в кількісній мірі (дорівнює «=», не дорівнює « \neq », більше «>», менше «<»).

$h_{2,1,3}$ – відсутня точка абсолютного відліку.

$h_{2,1,4}$ – над даними можна виконувати наступні математичні операції: додавання «+», віднімання «-».

Отже, вигляд моделі для інтервальної шкали набуває наступного виду (рис. 5):

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

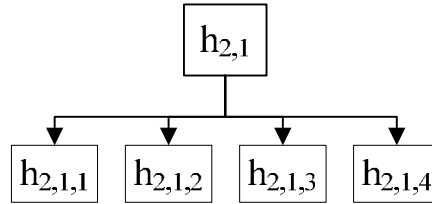


Рис. 5. Модель ознак, що репрезентують інтервальну шкалу

Розглянемо шкалу **відношень**, за допомогою якої передбачається співставлення величини ознаки з інтервалом можливих її значень. Така шкала має певне нульове значення. За шкалою відношень вимірюється більшість фізичних одиниць: масу тіла, довжину, заряд, а також ціни в економіці. Час вимірюється за шкалою різниць, якщо рік приймаємо природною одиницею виміру, і за шкалою інтервалів в загальному випадку. Дані які відносяться до шкали відношень позначимо через $h_{2,2}$, ознаки які характеризують даний вид виміру позначимо через $h_{2,2,1}, h_{2,2,2}, h_{2,2,3}, \dots, h_{2,2,n}$.

Визначимо ознаки, які характеризують дану шкалу.

$h_{2,2,1}$ – дані виражені кількісно і мають одиницю виміру.

$h_{2,2,2}$ – з динамі можна проводити порівняння в кількісній мірі (дорівнює «=», не дорівнює « \neq », більше «>», менше «<»).

$h_{2,2,3}$ – відсутня точка абсолютного відліку (абсолютного нуля «0»).

$h_{2,2,4}$ – над даними можна виконувати наступні математичні операції: додавання «+», віднімання «-», множення «*», ділення «/».

Отже, вигляд моделі для шкали відношень набуває наступного виду (рис. 6):

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

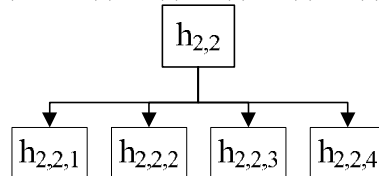


Рис. 6. Модель ознак, що репрезентують шкалу відношень

На основі виділених ознак типів шкал виміру даних побудуємо узагальнену модель аналізу (рис. 7).

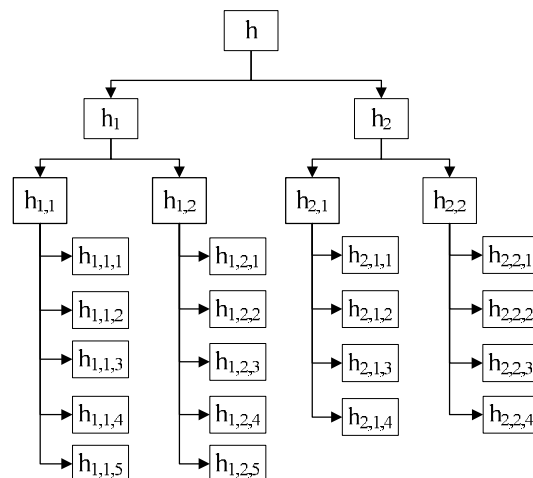


Рис. 7. Узагальнена ознакова модель опису даних в задачах на основі шкал виміру

Отже згідно з розглянутими типами шкал та їх ознаками студент повинен коректно визначити до якого типу шкали відносяться ті дані з якими вони матимуть справу у своїй практичній діяльності. До такого висновку майбутні інженери-педагоги можуть прийти за допомогою тих ознак які їм надаються в описі шкал виміру, а не на основі інтуїтивного, хаотичного аналізу.

Наступним, четвертим етапом, виділимо визначення закону розподілу, якому підпорядковуються дані.

Для того щоб проаналізувати значення змінних необхідно встановити – яка ймовірність того що змінна прийме дане значення з даного інтервалу. Іншими словами, ми цікавимося тим, як розподілені значення змінної. Описові статистики дають загальну інформацію про розподіл змінної. Повну відповідь дає функція розподілу. Таким чином, позначимо ознаку «закон розподілу», як g .

Нормальний закон розподілу (закон розподілу Гауса). Цей закон є одним із *найпоширеніших законів розподілу похибок*, що пояснюється центральною граничною теоремою теорії ймовірностей, котра твердить, що розподіл випадкових похибок буде близьким до нормального, якщо результати спостереження формуються під впливом великої кількості незалежних факторів впливу, кожний із котрих створює лише незначну дію в порівнянні із сумарною дією всієї решти.

Зазвичай у статистиці вимагають щоб досліджувані змінні самі були нормально розподіленими в сукупності, тобто задовольняли б «припущенню». При збільшенні обсягу вибірки форма розподілу статистики критерію наближається до нормального, навіть якщо розподілу досліджуваних змінних не являється нормальним, цей принцип називається центральною граничною теоремою.

Таким чином, позначимо ознаку «нормальний закон розподілу», як g_1 .

Виділимо ознаки нормального закону розподілу:

$g_{1,1}$ – кількість спостережень (>25);

$g_{1,2}$ – побудова нормального ймовірнісного графіку (дані розподілені нормально).

Дані які не підпорядковуються нормальному закону розподілу позначимо через ознаку –

g_2 .

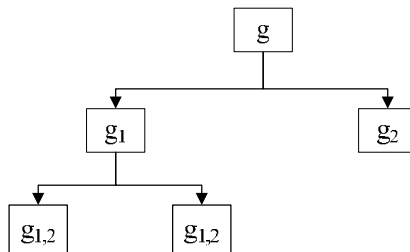


Рис. 8. Модель ознак закону розподілу даних в умові задачі

На основі виділених ознак побудуємо узагальнену модель аналізу умови задач для опрацювання психолого-педагогічних, технічних та економічних даних майбутніми інженерами-педагогами (рис. 9).

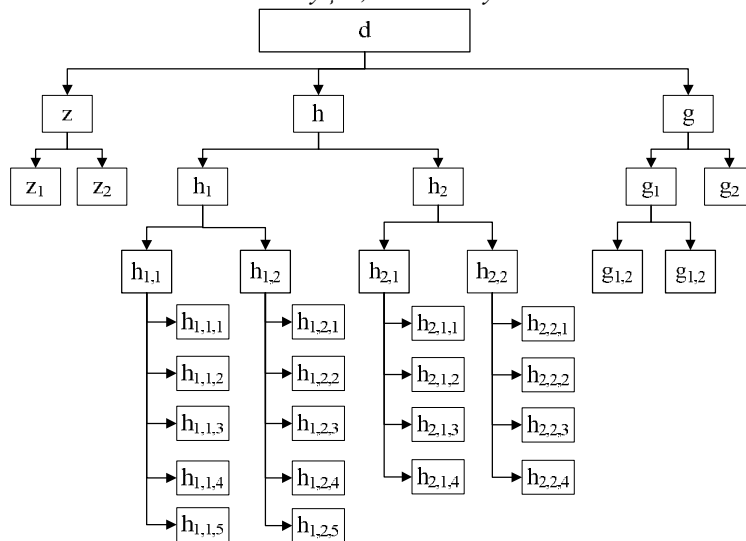


Рис. 9. Узагальнена ознакова модель аналізу даних умови задачі: причинно-наслідкових зв'язків між даними, опису даних на основі шкал виміру, закону розподілу даних

Висновки. Аналіз психолого-педагогічної та методичної літератури дозволив виявити та обґрунтувати необхідність і доцільність розробки методики навчання аналізу даних на основі виділення ознак які характеризують залежність і незалежність змінних, тип даних, шкали виміру та закон розподілу даних. Розроблені ознакові моделі аналізу даних умови задачі: причинно-наслідкових зв'язків між даними, опису даних на основі шкал виміру, закону розподілу даних, дозволять студентам інженерно-педагогічних спеціальностей перейти на якісно новий рівень аналізу даних враховуючи специфіку кожної предметної галузі та на їх основі коректно обирати статистичні методи аналізу.

У подальших дослідженнях планується виділення ознак які характеризують статистичні методи аналізу з метою співставлення їх з ознаками даних в умові задачі і для адекватного вибору необхідного статистичного методу аналізу даних задач з різних предметних галузей.

1. Коваленко Е.Э. Методика профессионального обучения : учеб. для инженеров-педагогов, преподавателей спецдисциплин системы профессионально-технического и высшего образования / Е.Э. Коваленко. – Х. : ЧП «Штрих», 2003. – 480 с.
2. Лігінін О.Є. Статистика : підрч. / О.Є. Лігінін, С.В. Білоусова. – К. : Центр навчальної літератури, 2005. – 580 с.
3. Мармоза А.Т. Теорія статистики : Навч. посіб. ВНЗ / А.Т. Мармоза. – К. : Ельга, Ніка-Центр, 2003. – 392 с.
4. Моторин Р.М. Статистика. Збірник індивідуальних завдань з використанням Excel : Навч.-метод. посіб. для самост. вивч. дисц. у вузах / Р.М. Моторин, Е.В. Чекатовський. – К. : КНЕУ, 2005 – 268 с.
5. Наследов А. Д. SPSS: компьютерный анализ данных в психологии и социальных науках / Наследов Андрей Дмитриевич; 2-е изд. – СПб. : Питер, 2007. – 416 с. : ил.
6. Опря А.Т. Статистика (з програмованою формою контролю знань). Математична статистика. Теорія статистики : Навч. посіб. / А.Т. Опря. – К. : Центр навчальної літератури, 2005 – 472 с.
7. Донелли-мл. Р. Статистика / Роберт А. Донелли-мл.; пер. с англ. Н.А. Ворониной. – М. : Астрель: АСТ, 2007. – XIV, 367, [3] с. : ил.