

Розв'язування задач з математичної статистики з використанням програми Gran1

Горошко Юрій Васильович,
Чернігівський педуніверситет імені Т.Г.Шевченка

Наведені в цій статті задачі відповідають методиці вивчення елементів статистики, запропонованій в [1, 2]. Задачі поділено на групи відповідно до тем, що вивчаються, причому основну увагу приділено задачам з практичним змістом. Наведені в статті команди програми Gran1 стосуються її Windows-версії.

Задачі на створення і аналіз частотної таблиці

Наведені тут задачі досить прості. Оскільки частотна таблиця для статистичної вибірки будується автоматично після виклику команди “Операції/Статистика/Частотна таблиця”, основна мета таких задач – сформувати у учнів навички введення статистичних вибірок, заданих у різний спосіб, правильно визначивши тип даних та модель даних (дискретну чи неперервну) для них.

У задачах 1-5 статистичну вибірку пропонується сформувати самостійно, провівши необхідні вимірювання чи експерименти.

Задача 1. З колоди карт навмання вибирають карту і записують її масть. Необхідно скласти частотну таблицю випадання мастей карт на основі 30 карт, що вибрано.

Зауваження. У цій задачі модель даних дискретна (4 масті). Оскільки програма Gran1 працює тільки з числовими вибірками, необхідно пронумерувати кожен масть, наприклад 1 - пікові, 2 - хрестові, 3 – червові, 4 – бубнові і працювати далі уже з числовою вибіркою. Тип даних в цій та інших експериментальних задачах – варіанти, оскільки інші типи можна отримати тільки після початкового опрацювання експериментальних даних шляхом групування варіант.

Задача 2. Гральний кубик кидають 50 разів. Скласти частотну таблицю випадання граней кубика.

Зауваження. Як і у попередній задачі модель даних – дискретна, тип даних – варіанти.

Задача 3. Виміряти довжину долоні 30 учнів і скласти частотну таблицю довжин.

Задача 4. Виміряти зріст 30 учнів і скласти частотну таблицю зростів.

Задача 5. Виміряти масу 40 учнів і скласти частотну таблицю, визначивши довжину інтервалу самостійно.

Зауваження. У задачах 3-5 досліджувана величина може набувати довільних значень з певного діапазону, тому модель даних – неперервна; тип даних – варіанти. У такому випадку програма при створенні вибірки запитає у додатковому вікні кількість підвідрізків, на які буде поділено відрізок задання вибірки. Тут доцільно буде скористатися формулою Стерджеса, натиснувши відповідну кнопку у додатковому вікні введення статистичної вибірки.

Задачі 6-9 більш складні, оскільки в них учні повинні робити висновки про генеральну сукупність на основі запропонованої вибірки шляхом виконання певних обчислень.

Задача 6. Коректор вибрав 40 сторінок з 400-сторінкового рукопису книги і перевірів ці сторінки на помилки. Кількість помилок на кожній сторінці така:

0	1	1	3	0	0	1	1
3	2	0	2	1	0	0	2
0	0	0	2	1	0	2	0
0	1	0	1	0	2	3	0
1	0	0	2	1	0	1	0

1) Потрібно побудувати частотну таблицю для кількості помилок на сторінці.

2) Для досліджуваної книги розрахувати наближену кількість:

- a) сторінок без помилок;
- b) сторінок з 3 і більше помилками;

- c) сторінок з 2 і менше помилками;
- d) загальну кількість помилок.

Зауваження. Модель даних – дискретна, тип даних - варіанти. Для розрахунків можна скористатися послугою “Операції/Калькулятор”.

Відповіді:

x	n	Накопич. n	P_n^*	Накопич. P_n^*
0	19	19	0.475	0.475
1	11	30	0.275	0.75
2	7	37	0.175	0.925
3	3	40	0.075	1

- a) Для визначення кількості сторінок без помилок необхідно перемножити відносну частоту сторінок без помилок ($P_n^*(0)$) на об'єм генеральної сукупності (n), тобто $0.475 \cdot 400 = 190$
- b) $P_n^*(3) \cdot n$, тобто $0.075 \cdot 400 = 30$
- c) $\text{Накопич. } P_n^*(2) \cdot n$, тобто $0.925 \cdot 400 = 370$
- d) $(0.275 + 0.175 \cdot 2 + 0.075 \cdot 3) \cdot 400 = 340$

Задача 7. Контейнер містить 5000 піддонів з яйцями (у кожному піддоні 20 яєць). На перевірку взято 100 піддонів і підраховано кількість розбитих яєць. Отримана вибірка:

0	0	0	1	0	20	0	0	0	20
0	1	2	1	0	0	3	20	1	0
1	0	1	1	0	0	1	0	0	0
0	0	20	0	2	3	1	2	0	0
2	0	0	3	0	0	2	0	3	0
1	1	0	0	0	3	3	2	0	0
0	0	0	1	0	0	2	2	0	20
0	0	1	2	2	20	20	1	0	0
0	0	0	1	1	0	0	20	0	1
1	0	0	0	20	1	2	0	0	0

Потрібно:

- 1) побудувати частотну таблицю;

2) розрахувати:

- кількість піддонів без розбитих яєць;
- кількість піддонів з всіма розбитими яйцями;
- кількість піддонів з більш ніж одним розбитим яйцем;
- прогнозовану загальну кількість цілих яєць у контейнері.

Зауваження. Модель даних – дискретна, тип даних – варіанти.

Відповіді:

x	n	Накопич. n	Pn*	Накопич. Pn*
0	55	55	0.55	0.55
1	19	74	0.19	0.74
2	11	85	0.11	0.85
3	6	91	0.06	0.91
20	9	100	0.09	1

- 55;
- 9;
- $11+6+19=36$;
- $(0.55*20+0.19*19+0.11*18+0.06*17)*5000=88050$

Задача 8. Комп'ютерна програма згенерувала 100 випадкових чисел у проміжку від 0 до 1 (з трьома знаками після коми):

0.184	0.398	0.941	0.953	0.357	0.690	0.035	0.116	0.226	0.168
0.421	0.687	0.540	0.415	0.966	0.503	0.809	0.346	0.794	0.691
0.930	0.512	0.745	0.633	0.454	0.631	0.970	0.621	0.299	0.332
0.581	0.808	0.947	0.344	0.537	0.322	0.435	0.614	0.181	0.636
0.669	0.098	0.105	0.577	0.880	0.186	0.257	0.559	0.244	0.031
0.474	0.919	0.985	0.227	0.847	0.451	0.495	0.602	0.606	0.505
0.273	0.094	0.438	0.264	0.722	0.946	0.410	0.046	0.118	0.712
0.918	0.751	0.278	0.796	0.345	0.810	0.736	0.524	0.328	0.243
0.472	0.965	0.831	0.684	0.486	0.005	0.427	0.490	0.774	0.389
0.472	0.290	0.487	0.736	0.081	0.569	0.339	0.738	0.096	0.177

Потрібно:

- побудувати частотну таблицю з відрізком поділу 0,1;

2) прокоментувати отриману частотну таблицю (правильно чи не правильно працює програма генерації випадкових чисел).

Зауваження. В задачі 8 наведено готову статистичну вибірку, але можна запропонувати учням написати програму на одній з мов програмування для отримання текстового файлу з значеннями випадкових чисел. В такому випадку, по-перше, учні навчаться вводити вибірку завантаженням її з файлу, що значно економить час в порівнянні з введенням вручну, а по-друге, вони зможуть розглянути і вибірки більшого об'єму, наприклад 1000, для отримання більш обгрунтованого висновку про роботу генератора випадкових чисел. При розв'язуванні необхідно встановити неперервну модель даних і тип даних – варіанти. У додатковому вікні треба встановити відрізок задання $[0; 1]$, а кількість відрізків поділу – 10.

Задача 9. Певну компанію цікавить час життя деяких компонентів приладів. 30 компонентів тестували до їх зламу і записували (у годинах) час роботи цих компонент. Отримали таблицю:

1,2	21,0	34,7	13,2	3,6	14,7
31,0	17,1	22,1	16,4	21,2	15,2
11,3	6,8	2,7	31,2	9,0	6,8
23,7	16,3	30,0	28,6	19,0	29,0
4,4	44,3	18,5	5,3	5,5	10,0

Потрібно:

- 1) створити частотну таблицю з довжиною інтервалу 10 год.;
- 2) компанія стверджує, що не менше 90% компонентів працює не менше 10 год. Чи можна довіряти цьому твердженню?
- 3) фірма замовила 150 компонентів; необхідно розрахувати кількість компонентів, які пропрацюють більше 20 годин.

Зауваження. Модель даних – неперервна, тип даних – варіанти. У додатковому вікні створення вибірки потрібно вказати відрізок задання $[0; 50]$ і ширину інтервалу 10.

Відповіді.

Відрізок	n	Накопич. n	Pn*	Накопич. Pn*
0 - 10	9	9	0.3	0.3
10 - 20	10	19	0.3333	0.6333
20 - 30	6	25	0.2	0.8333
30 - 40	4	29	0.1333	0.9667
40 - 50	1	30	0.03333	1

2) Оскільки 30% деталей з вибірки працюють не більше 10 год., твердження компанії можна вважати неправдивим.

3) $(1-0.6333)*150=55.01$, тобто можна сподіватися, що в партії з 150 деталей більше за 20 год. будуть працювати 55 деталей.

Побудова графіків

Для ознайомлення з різними видами статистичних графіків можна запропонувати учням побудувати такі графіки для вибірок, розглянутих раніше (тому важливо записувати на диск всі вибірки, які створюють учні).

Задача 10. Для порівняння двох добрив при вирощуванні соняхів вибрали двісті ділянок однієї площі і на перших ста з них застосували перший вид добрив, а на інших ста ділянках – другий вид добрив. Обчислили кількість соняшникової олії, отриманої з кожної ділянки і отримали наступну таблицю:

Мітка класу	600	800	1000	1200	1400	1600	1800
1-ше добриво	3	9	19	39	21	9	0
2-ге добриво	3	2	11	32	30	20	2

Необхідно:

- 1) За допомогою гістограм, побудованих в одній системі координат, відповісти, яке з добрив дозволяє отримати більшу кількість олії.
- 2) За допомогою функцій розподілу, побудованих в одній системі координат, порівняти для кожного виду добрив процент ділянок, де кількість олії більша за 1200 л..

Зауваження. Модель даних – неперервна, тип даних – частоти. Вводити потрібно мітки класу (тобто середини інтервалів, на які поділено відрізок задання вибірки) та відповідні їм частоти. Вид графіка для кожної вибірки

встановлюють при її створенні або зміні. Функція розподілу в програмі називається “Функція $F^*(x)$ ”.

Відповіді

- 1) З побудованих в одній системі координат гістограм видно, що гістограма для другого виду добрив має стовпчики, нижчі для малих врожаїв і вищі для великих (у порівнянні з іншою гістограмою), тому можна робити висновок про більшу ефективність саме другого добрива.
- 2) Для аргументу 1200 функції розподілу для першого і другого добрив мають значення 0.51 та 0.32 відповідно (ці значення можна отримати за допомогою курсору). Це означає, що 51% ділянок для першого виду добрив і 32% ділянок для другого дають олії менше за 1200 л., тому процент ділянок з більшою за 1200 л. врожайністю буде відповідно 49% та 68%.

Числові характеристики вибірок

Оскільки числові характеристики вибірки програма Gran1 визначає автоматично і записує їх у вікні властивостей вибірки, немає сенсу розв’язувати велику кількість задач для формування навичок знаходження цих чисел. Краще розглянути більш цікаві задачі, в яких результати обчислень за програмою числових значень вибірки є джерелом для подальших роздумів учнів щодо розв’язування задачі.

Задача 11. В таблиці наведено дані про зарплати співробітників певної фірми.

Зарплата (грн.)	230	380	560	2000	3000
Кількість співробітників	20	7	4	2	1

Необхідно визначити, яка “в середньому” є зарплата співробітників.

Зауваження. В задачі потрібно визначити міру центральної тенденції для наведеної вибірки. Для дискретної моделі даних програма Gran1 визначає дві таких характеристики: середнє арифметичне (в програмі це число називається математичним сподіванням) та моду - 485.29 та 230 відповідно.

Якщо побудувати для вибірки полігон, можна помітити, що він є дуже асиметричним, тому в цьому випадку досліджувану величину краще характеризує мода (тобто число 230). Такий же висновок можна зробити і з наступних міркувань: оскільки декілька співробітників фірми мають дуже високу зарплату у порівнянні з іншими, значення середнього арифметичного значно відрізняється в сторону збільшення від зарплати основної маси співробітників і не може бути використане в якості відповіді.

Задача 12. У таблиці, наведеній нижче, дано округлений до одного см розмір ноги учнів трьох класів.

см	15	16	17	18	19	20	21	22	23	24	25	26	27	28
n	1	3	5	6	5	4	6	8	7	6	5	3	2	1

Необхідно визначити моду.

Зауваження. При формальному підході до розв'язування задачі відповіддю буде число 22, запропоноване програмою. Але тут потрібно задуматись, наскільки така характеристика має реальне значення. Насправді, відомо, що розмір ноги хлопців і дівчат значно відрізняється. Якщо за допомогою послуги “Графік/Побудувати” побудувати полігон частот для цієї вибірки, ми побачимо два локальних максимуми; тому такі вибірки називають бімодальними. За допомогою полігону можна визначити і значення цих двох мод; це будуть числа 18 та 22, вони і будуть відповіддю задачі. На основі вищезгаданого також можна зробити висновок про те, що такі вибірки, якщо це можливо, краще розділяти на дві і кожен з них розглядати окремо.

Критерій Пірсона

Задача 13. Для задачі 8 (про генератор випадкових чисел) перевірити гіпотезу про рівномірний розподіл з рівнем значущості $\alpha=0,01$.

Зауваження. Якщо потрібно дослідити рівномірність неперервного розподілу ймовірностей випадкової величини, гістограму потрібно порівнювати з функцією $y=1/(b-a)$, де a, b – кінці відрізка, на якому задано

вибірку. В нашому випадку це буде функція $y=1$. Для розв'язування задачі учень повинен:

- завантажити вибірку, записану раніше на диск;
- створити функцію $y=1$, причому відрізок задання функції повинен співпадати з відрізком задання вибірки, тобто з відрізком $[0; 1]$;
- Скористатися послугою “Операції/Статистика/Критерій Пірсона”, встановивши рівень значущості $\alpha=0.01$.

Після виконання обчислень програма повідомить про те, що гіпотеза не підтвердилася.

Задача 14. Для перевірки справжності рукописів листів К.К.Снодграсса, які приписують М.Твену, Ч.С.Грінегар підрахував кількість слів довжиною в $k=1, 2, \dots, 12$ літер в творах М.Твена і в листах К.К.Снодграсса. З'ясувалося, що слова з 2, 3, 4 і більше літер зустрічаються в творах М.Твена з відносними частотами

2	3	4	Решта
0,177	0,232	0,191	0,400

В 13175 словах рукописів К.К.Снодграсса слова тієї ж довжини зустрічаються з наступними частотами:

2	3	4	Решта
2685	2752	2302	5436

За допомогою критерія Пірсона перевірити гіпотезу про те, що листи К.К.Снодграсса написані М.Твеном.

Зауваження. Для обох вибірок модель даних – дискретна, тип даних для першої вибірки – відносні частоти. Оскільки в задачі не сказано, скільки слів з творів М.Твена було проаналізовано, будемо вважати, що їх було приблизно стільки ж, скільки досліджено слів з рукописів К.К.Снодграсса, наприклад 13000. Це число потрібно вписати у полі “Об’єм вибірки”. Тип даних другої вибірки – частоти. При введенні вибірок в програму можна вважати всі слова з кількістю літер від 5 до 12 як слова з кількістю літер 5. В задачі не вказано рівень значущості, з яким потрібно провести дослідження,

тому його потрібно вибрати самостійно, при цьому слід пам'ятати, що значення α , близькі до нуля, вибирають, щоб не допустити прийняття хибної гіпотези, але в цьому випадку великий ризик відхилити правильну гіпотезу. Значення α , близькі до 1, вибирають, щоб не відхилити правильну гіпотезу, але так значно збільшується ризик прийняття гіпотези неправильної. Тобто при мінімальному значенні α отримуємо найбільш строгий аналіз. Наперед скажемо, що в даній задачі гіпотеза не підтверджується при всіх значеннях α . Послідовність дій для розв'язування може бути такою:

- ввести першу вибірку;
- ввести другу вибірку;
- відмітити вибірки і звернутися до команди
“Операції/Статистика/Критерій Пірсона”, теоретичною вибіркою
будемо вважати вибірку з даними про твори М.Твена.

Відповідь

Гіпотеза не підтвердилася, тобто можна вважати, що М.Твен не писав листів К.К.Снодграсса.

Задача 15. Результати підрахунку частот появи цифр 0, 1, ..., 9 в 10002 знаках десяткового запису числа π -3 наведено в таблиці:

0	1	2	3	4	5	6	7	8	9
968	1026	1021	974	1014	1046	1021	970	948	1014

За допомогою критерія Пірсона перевірити гіпотезу про рівноможливість появи в даному записі кожної з цифр 0, 1, ..., 9.

Зауваження. В цій задачі потрібно порівнювати дану вибірку з вибіркою рівномірного розподілу всіх цифр, яку і будемо вважати теоретичною:

0	1	2	3	4	5	6	7	8	9
1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

Модель даних для обох вибірок дискретна, тип даних – варіанти.

Відповідь

При рівнях значущості, близьких до 0, гіпотезу потрібно відкинути, при рівнях значущості, близьких до 1, немає причин відкидати гіпотезу.

Задача 16. Для аналізу впливу вітаміну С на профілактику простудних захворювань 200 чоловік випадковим чином було поділено на 2 рівних групи. Одній групі дали вітамін С, іншій – схожу пігулку, але без вітаміну, причому всім пацієнтам було сказано, що їм дали вітамін С. Результати обстеження наведено у наступній таблиці:

	Менше простудних захворювань	Більше простудних захворювань	Без змін
Контрольна група	39	21	40
Група, що приймала вітамін	51	20	29

За допомогою критерія Персона для рівня значущості $\alpha=0,05$ перевірити гіпотезу про незалежність простудних захворювань від прийому вітаміну С.

Зауваження. Модель даних - дискретна, тип даних – частоти. Щоб можна було скористатися програмою Gran1, надамо значенню “Менше простудних захворювань” номер 1, значенню “Більше простудних захворювань” номер 2, значенню “без змін” номер 3. Вибірку для контрольної групи будемо вважати теоретичною.

Відповідь

Гіпотеза не підтвердилася, тобто можна вважати, що кількість простудних захворювань залежить від прийому вітаміну С.

Задача 17. За даними з задачі 5 за допомогою критерію Пірсона для рівня значущості 0.01 перевірити гіпотезу про нормальний розподіл зросту учнів.

Зауваження. При розв’язуванні задачі необхідно завантажити файл з вибіркою (звернувшись до послуги програми “Файл/Відкрити”), відмітити вибірку і звернутися до послуги “Операції/Статистика/Функція щільності норм. розподілу за вибіркою”. Ця послуга записує до списку об’єктів

функцію щільності вибірки, вважаючи цю вибірку нормально розподіленою. Далі необхідно скористатися послугою "Операції/Статистика/Критерій Пірсона". Відповідь до задачі залежить від експериментальних даних, визначених учнями, тому ми її не наводимо.

Література

1. Жалдак М.І., Горошко Ю.В. Комп'ютер і елементи стохастики у шкільному курсі математики / Комп'ютер у школі та сім'ї, №3, 1998 р. С.16-20.
2. Жалдак М.І., Горошко Ю.В. Комп'ютер і елементи стохастики у шкільному курсі математики / Комп'ютер у школі та сім'ї, №4, 1998 р. С.22-27.
3. Чиби́сов Д.М., Пагурова В.И. Задачи по математической статистике.- Изд-во Моск. Ун-та, 1990.- 171 с.
4. Alfred P.Hanwell, Marshall P.Bye, Thomas J.Grifiths. Holt mathetatics 4.- Holt, Rinechart and Winston of Canada, Limited, 1980.- 470 с.
5. Kennth D.Fryer, Ronald G.Dunkley, H.A.Elliott, Norman J.Hill R.Jock MacKay. Holt mathetatics 5.- Holt, Rinechart and Winston of Canada, Limited, 1980.- 373 с.
6. Dino Dottori, George Knill, John Seumour. Applied mathematics for today: senior si metric edition.- McGraw –Hill Ryerson Limited, 1977.- 486 с.