

МІЖРЕГІОНАЛЬНА
АКАДЕМІЯ УПРАВЛІННЯ ПЕРСОНАЛОМ



МАУП

**МЕТОДИЧНІ РЕКОМЕНДАЦІЇ
ЩОДО ВИКОНАННЯ КОНТРОЛЬНОЇ РОБОТИ
з дисципліни
“МАТЕМАТИЧНІ МЕТОДИ В СОЦІОЛОГІЇ”
(для бакалаврів)**

МАУП

Київ 2008

Підготовлено професором кафедри математики *Р. К. Чорнеєм*, доцентом кафедри інформатики та інформаційних технологій *А. Б. Телейком*

Затверджено на засіданні кафедри математики (протокол № 5 від 31.01.07)

Схвалено Вченою радою Міжрегіональної Академії управління персоналом

Телейко А. Б., Чорней Р. К. Методичні рекомендації щодо виконання контрольної роботи з дисципліни “Математичні методи в соціології” (для бакалаврів). — К.: МАУП, 2008. — 42 с.

Методичні рекомендації містять пояснювальну записку, теоретичні питання, задачі для контрольної роботи, приклади розв’язання задач, а також список літератури.

© Міжрегіональна Академія
управління персоналом (МАУП),
2008

ПОЯСНЮВАЛЬНА ЗАПИСКА

Контрольна робота складається з двох теоретичних питань і двох задач. Номери теоретичних питань студенти вибирають за останньою цифрою номера своєї залікової книжки (див. таблицю).

| Остання цифра номера залікової книжки студента | Номери теоретичних питань |
|--|---------------------------|
| 1 | 1, 11 |
| 2 | 2, 12 |
| 3 | 3, 13 |
| 4 | 4, 14 |
| 5 | 5, 15 |
| 6 | 6, 16 |
| 7 | 7, 17 |
| 8 | 8, 18 |
| 9 | 9, 19 |
| 0 | 10, 20 |

Для кожної практичної задачі подано 10 варіантів даних. Номер варіанта задачі студенти вибирають за останньою цифрою номера своєї залікової книжки (цифри "0" відповідає варіант 10).

Контрольну роботу студенти виконують у зошиті або на аркушах паперу формату А4 з полями для зауважень викладача. При розв'язанні кожної задачі необхідно вказати її номер і переписати умову. Розв'язання задачі обов'язково потрібно пояснювати. У розрахунках слід використовувати правила наближених обчислень.

ТЕОРЕТИЧНІ ПИТАННЯ ДЛЯ КОНТРОЛЬНОЇ РОБОТИ

1. Шкали вимірювання.
2. Усереднені показники та міри розсіювання, їх порівняльний аналіз.
3. Міри розсіювання та їх порівняльний аналіз.
4. Форми емпіричних розподілів.
5. Методи формування випадкової вибірки.
6. Вибіркові розподіли та їх властивості. Розподіл вибіркового середнього.
7. Спеціальні розподіли: t -розподіли Стьюдента, χ^2 -розподіли Колмогорова, F -розподіли Фішера.

8. Аналіз генеральної сукупності: точкові та інтервальні оцінки, перевірка гіпотез.
9. Процедура вибірки: мета вибіркового дослідження, види вибірок, випадкові вибірки, методи формування випадкових вибірок.
10. Вибіркові розподіли: розподіл вибіркового середнього; спеціальні вибіркові розподіли.
11. Точкові оцінки та їх властивості: лінійність, незміщеність, ефективність, BLUE-оцінки, змістовність. Найуживаніші точкові оцінки.
12. Інтервальні оцінки генерального середнього при відомому та невідомому стандартних відхиленнях, інтервальні оцінки різниці генеральних середніх. Інтервальні оцінки генерального стандартного відхилення.
13. Інтервальні оцінки генеральної пропорції, інтервальні оцінки різниці генеральних пропорцій.
14. Точність інтервального оцінювання. Мінімальний розмір вибірки, необхідний для забезпечення заданої точності.
15. Статистичні критерії порівняння ознак.
16. ϕ^* -критерій (кутове перетворення Фішера), його застосування та обмеження.
17. Статистичні критерії розпізнавання зсувів.
18. χ^2 -критерій Пірсона, його застосування та обмеження.
19. λ -критерій Колмогорова — Смирнова, m -біноміальний критерій, їх застосування та обмеження.
20. Числові характеристики кореляції.

ЗАДАЧІ ДЛЯ КОНТРОЛЬНОЇ РОБОТИ

Задача 1

У деякій країні є чотири політичні партії: білі, червоні, зелені та чорні. Незалежна соціологічна служба провела опитування щодо політичних симпатій виборців країни. Країна складається з двох великих регіонів: Заходу і Сходу. Оскільки кожний з цих регіонів має власні історичні, релігійні та культурні традиції, соціологічна служба вирішила поділити дослідження на дві частини: опитування на Заході та Сході. Обсяг опитаної випадкової вибірки в західному регіоні становив 2500 респондентів, а у східному — 2300. Кожний респондент вказував партію, прихильником якої був у 2000 р. та 2004 р. Дані дослідження наведені в таблиці.

Варіант 1

| Політична партія | Кількість прихильників за роками та регіонами | | | |
|------------------|---|----------------|-----------------|----------------|
| | 2000 р. | | 2006 р. | |
| | Західний регіон | Східний регіон | Західний регіон | Східний регіон |
| Білі | 577 | 581 | 613 | 601 |
| Червоні | 651 | 536 | 610 | 527 |
| Зелені | 600 | 552 | 561 | 488 |
| Чорні | 672 | 631 | 716 | 684 |

Варіант 2

| Політична партія | Кількість прихильників за роками та регіонами | | | |
|------------------|---|----------------|-----------------|----------------|
| | 2000 р. | | 2006 р. | |
| | Західний регіон | Східний регіон | Західний регіон | Східний регіон |
| Білі | 579 | 577 | 633 | 593 |
| Червоні | 659 | 524 | 593 | 539 |
| Зелені | 596 | 553 | 553 | 489 |
| Чорні | 666 | 646 | 721 | 679 |

Варіант 3

| Політична партія | Кількість прихильників за роками та регіонами | | | |
|------------------|---|----------------|-----------------|----------------|
| | 2000 р. | | 2006 р. | |
| | Західний регіон | Східний регіон | Західний регіон | Східний регіон |
| Білі | 583 | 565 | 621 | 605 |
| Червоні | 640 | 533 | 606 | 524 |
| Зелені | 599 | 550 | 541 | 490 |
| Чорні | 678 | 652 | 732 | 681 |

Варіант 4

| Політична партія | Кількість прихильників за роками та регіонами | | | |
|------------------|---|----------------|-----------------|----------------|
| | 2000 р. | | 2006 р. | |
| | Західний регіон | Східний регіон | Західний регіон | Східний регіон |
| Білі | 582 | 566 | 625 | 593 |
| Червоні | 639 | 530 | 609 | 538 |
| Зелені | 602 | 549 | 542 | 492 |
| Чорні | 677 | 655 | 724 | 677 |

Варіант 5

| Політична партія | Кількість прихильників за роками та регіонами | | | |
|------------------|---|----------------|-----------------|----------------|
| | 2000 р. | | 2006 р. | |
| | Західний регіон | Східний регіон | Західний регіон | Східний регіон |
| Білі | 564 | 580 | 626 | 603 |
| Червоні | 657 | 532 | 608 | 539 |
| Зелені | 610 | 553 | 542 | 475 |
| Чорні | 669 | 635 | 724 | 683 |

Варіант 6

| Політична партія | Кількість прихильників за роками та регіонами | | | |
|------------------|---|----------------|-----------------|----------------|
| | 2000 р. | | 2006 р. | |
| | Західний регіон | Східний регіон | Західний регіон | Східний регіон |
| Білі | 585 | 574 | 623 | 599 |
| Червоні | 649 | 538 | 608 | 518 |
| Зелені | 597 | 555 | 545 | 486 |
| Чорні | 669 | 633 | 724 | 697 |

Варіант 7

| Політична партія | Кількість прихильників за роками та регіонами | | | |
|------------------|---|----------------|-----------------|----------------|
| | 2000 р. | | 2006 р. | |
| | Західний регіон | Східний регіон | Західний регіон | Східний регіон |
| Білі | 567 | 580 | 617 | 591 |
| Червоні | 660 | 523 | 594 | 534 |
| Зелені | 592 | 543 | 544 | 484 |
| Чорні | 681 | 654 | 745 | 691 |

Варіант 8

| Політична партія | Кількість прихильників за роками та регіонами | | | |
|------------------|---|----------------|-----------------|----------------|
| | 2000 р. | | 2006 р. | |
| | Західний регіон | Східний регіон | Західний регіон | Східний регіон |
| Білі | 586 | 585 | 634 | 593 |
| Червоні | 658 | 519 | 605 | 536 |
| Зелені | 594 | 561 | 553 | 472 |
| Чорні | 662 | 635 | 708 | 699 |

Варіант 9

| Політична партія | Кількість прихильників за роками та регіонами | | | |
|------------------|---|----------------|-----------------|----------------|
| | 2000 р. | | 2006 р. | |
| | Західний регіон | Східний регіон | Західний регіон | Східний регіон |
| Білі | 575 | 586 | 614 | 600 |
| Червоні | 645 | 529 | 598 | 528 |
| Зелені | 605 | 562 | 545 | 473 |
| Чорні | 675 | 623 | 743 | 699 |

Варіант 10

| Політична партія | Кількість прихильників за роками та регіонами | | | |
|------------------|---|----------------|-----------------|----------------|
| | 2000 р. | | 2006 р. | |
| | Західний регіон | Східний регіон | Західний регіон | Східний регіон |
| Білі | 583 | 570 | 616 | 597 |
| Червоні | 661 | 526 | 607 | 522 |
| Зелені | 607 | 562 | 544 | 486 |
| Чорні | 649 | 642 | 733 | 695 |

1. Враховуючи, що співвідношення між кількостями виборців на Заході та Сході становить від 46 % до 54 %, для кожної партії знайти точкову оцінку генеральної пропорції її прихильників:

- у 2000 р. та 2006 р. окремо в кожному регіоні;
- у 2000 р. та 2006 р. по країні разом.

2. Для кожної партії визначити 95 %-й довірчий інтервал для генеральної пропорції її прихильників:

- у 2000 р. та 2006 р. окремо в кожному регіоні;
- у 2000 р. та 2006 р. по країні разом;

3. Для кожної партії визначити 95 %-й довірчий інтервал для різниці генеральних пропорцій її прихильників на Заході та Сході:

- у 2000 р.;
- у 2006 р.

4. За допомогою χ^2 -критерію Пірсона перевірити на рівні значущості 5 %, чи узгоджений з рівномірним розподілом розподіл кількості прихильників між політичними партіями:

- на Заході у 2006 р.;
- на Сході у 2006 р.;
- разом по країні у 2006 р.

Література [5, розд. 2, 5, 6; 8, розд. 12, 13]

Задача 2

Незалежна соціологічна служба провела опитування 400 випадкових респондентів працездатного віку в центральному регіоні країни про їх місячні прибутки X у 2000 та 2006 р. (з точністю до 10 ум. од.). Дані опитування наведені в таблиці.

Варіант 1

| Місячні прибутки X , ум. од. | Кількість опитаних | | Місячні прибутки X , ум. од. | Кількість опитаних | |
|--------------------------------|--------------------|---------|--------------------------------|--------------------|---------|
| | 2000 р. | 2006 р. | | 2000 р. | 2006 р. |
| 10 | 4 | 0 | 110 | 14 | 21 |
| 20 | 5 | 4 | 120 | 23 | 18 |
| 30 | 8 | 8 | 130 | 14 | 12 |
| 40 | 19 | 17 | 140 | 10 | 25 |
| 50 | 25 | 23 | 150 | 18 | 13 |
| 60 | 56 | 39 | 160 | 10 | 7 |
| 70 | 52 | 40 | 170 | 10 | 16 |
| 80 | 51 | 63 | 180 | 11 | 6 |
| 90 | 25 | 46 | 190 | 8 | 9 |
| 100 | 33 | 25 | 200 | 4 | 8 |

Варіант 2

| Місячні прибутки X , ум. од. | Кількість опитаних | | Місячні прибутки X , ум. од. | Кількість опитаних | |
|--------------------------------|--------------------|---------|--------------------------------|--------------------|---------|
| | 2000 р. | 2006 р. | | 2000 р. | 2006 р. |
| 10 | 6 | 0 | 110 | 14 | 17 |
| 20 | 4 | 4 | 120 | 9 | 19 |
| 30 | 10 | 2 | 130 | 21 | 13 |
| 40 | 22 | 16 | 140 | 10 | 14 |
| 50 | 41 | 35 | 150 | 12 | 10 |
| 60 | 63 | 56 | 160 | 3 | 9 |
| 70 | 51 | 40 | 170 | 4 | 12 |
| 80 | 51 | 57 | 180 | 15 | 8 |
| 90 | 32 | 42 | 190 | 6 | 12 |
| 100 | 21 | 26 | 200 | 5 | 8 |

Варіант 3

| Місячні прибутки X, ум. од. | Кількість опитаних | | Місячні прибутки X, ум. од. | Кількість опитаних | |
|-----------------------------------|--------------------|---------|-----------------------------------|--------------------|---------|
| | 2000 р. | 2006 р. | | 2000 р. | 2006 р. |
| 10 | 4 | 0 | 110 | 18 | 25 |
| 20 | 5 | 1 | 120 | 11 | 16 |
| 30 | 11 | 4 | 130 | 19 | 16 |
| 40 | 19 | 8 | 140 | 9 | 9 |
| 50 | 18 | 22 | 150 | 8 | 11 |
| 60 | 72 | 50 | 160 | 7 | 13 |
| 70 | 57 | 56 | 170 | 12 | 21 |
| 80 | 53 | 57 | 180 | 9 | 13 |
| 90 | 39 | 44 | 190 | 5 | 3 |
| 100 | 22 | 26 | 200 | 2 | 5 |

Варіант 4

| Місячні прибутки X, ум. од. | Кількість опитаних | | Місячні прибутки X, ум. од. | Кількість опитаних | |
|-----------------------------------|--------------------|---------|-----------------------------------|--------------------|---------|
| | 2000 р. | 2006 р. | | 2000 р. | 2006 р. |
| 10 | 5 | 0 | 110 | 15 | 13 |
| 20 | 7 | 4 | 120 | 12 | 24 |
| 30 | 12 | 6 | 130 | 15 | 14 |
| 40 | 19 | 19 | 140 | 10 | 20 |
| 50 | 35 | 17 | 150 | 13 | 12 |
| 60 | 57 | 39 | 160 | 7 | 25 |
| 70 | 55 | 67 | 170 | 8 | 14 |
| 80 | 45 | 46 | 180 | 9 | 9 |
| 90 | 44 | 36 | 190 | 8 | 7 |
| 100 | 19 | 15 | 200 | 5 | 13 |

Варіант 5

| Місячні прибутки X, ум. од. | Кількість опитаних | | Місячні прибутки X, ум. од. | Кількість опитаних | |
|-----------------------------------|--------------------|---------|-----------------------------------|--------------------|---------|
| | 2000 р. | 2006 р. | | 2000 р. | 2006 р. |
| 1 | 2 | 3 | 4 | 5 | 6 |
| 10 | 3 | 0 | 110 | 14 | 19 |
| 20 | 5 | 3 | 120 | 16 | 20 |
| 30 | 7 | 2 | 130 | 12 | 15 |
| 40 | 19 | 16 | 140 | 11 | 12 |

| | | | | | |
|-----|----|----|-----|----|----|
| 1 | 2 | 3 | 4 | 5 | 6 |
| 50 | 31 | 27 | 150 | 13 | 16 |
| 60 | 54 | 45 | 160 | 11 | 8 |
| 70 | 47 | 56 | 170 | 7 | 14 |
| 80 | 60 | 50 | 180 | 8 | 6 |
| 90 | 38 | 50 | 190 | 4 | 13 |
| 100 | 33 | 18 | 200 | 7 | 10 |

Варіант 6

| Місячні прибутки X, ум. од. | Кількість опитаних | | Місячні прибутки X, ум. од. | Кількість опитаних | |
|-----------------------------------|--------------------|---------|-----------------------------------|--------------------|---------|
| | 2000 р. | 2006 р. | | 2000 р. | 2006 р. |
| 10 | 4 | 0 | 110 | 14 | 24 |
| 20 | 4 | 8 | 120 | 19 | 18 |
| 30 | 8 | 2 | 130 | 23 | 15 |
| 40 | 20 | 14 | 140 | 14 | 19 |
| 50 | 25 | 19 | 150 | 9 | 10 |
| 60 | 50 | 34 | 160 | 4 | 11 |
| 70 | 64 | 64 | 170 | 10 | 16 |
| 80 | 45 | 53 | 180 | 9 | 8 |
| 90 | 41 | 44 | 190 | 10 | 5 |
| 100 | 24 | 27 | 200 | 3 | 9 |

Варіант 7

| Місячні прибутки X, ум. од. | Кількість опитаних | | Місячні прибутки X, ум. од. | Кількість опитаних | |
|-----------------------------------|--------------------|---------|-----------------------------------|--------------------|---------|
| | 2000 р. | 2006 р. | | 2000 р. | 2006 р. |
| 10 | 5 | 0 | 110 | 15 | 24 |
| 20 | 5 | 7 | 120 | 13 | 16 |
| 30 | 4 | 7 | 130 | 24 | 14 |
| 40 | 24 | 15 | 140 | 12 | 21 |
| 50 | 37 | 23 | 150 | 6 | 10 |
| 60 | 61 | 37 | 160 | 7 | 14 |
| 70 | 66 | 55 | 170 | 3 | 10 |
| 80 | 42 | 65 | 180 | 5 | 9 |
| 90 | 46 | 35 | 190 | 5 | 5 |
| 100 | 19 | 23 | 200 | 1 | 10 |

Варіант 8

| Місячні прибутки X, ум. од. | Кількість опитаних | | Місячні прибутки X, ум. од. | Кількість опитаних | |
|-----------------------------------|--------------------|---------|-----------------------------------|--------------------|---------|
| | 2000 р. | 2006 р. | | 2000 р. | 2006 р. |
| 10 | 3 | 0 | 110 | 22 | 24 |
| 20 | 6 | 3 | 120 | 12 | 9 |
| 30 | 7 | 3 | 130 | 23 | 19 |
| 40 | 21 | 17 | 140 | 8 | 21 |
| 50 | 29 | 28 | 150 | 11 | 17 |
| 60 | 54 | 57 | 160 | 4 | 11 |
| 70 | 65 | 46 | 170 | 14 | 11 |
| 80 | 49 | 61 | 180 | 5 | 7 |
| 90 | 27 | 39 | 190 | 7 | 6 |
| 100 | 29 | 16 | 200 | 4 | 5 |

Варіант 9

| Місячні прибутки X, ум. од. | Кількість опитаних | | Місячні прибутки X, ум. од. | Кількість опитаних | |
|-----------------------------------|--------------------|---------|-----------------------------------|--------------------|---------|
| | 2000 р. | 2006 р. | | 2000 р. | 2006 р. |
| 10 | 5 | 0 | 110 | 22 | 23 |
| 20 | 3 | 3 | 120 | 9 | 11 |
| 30 | 7 | 4 | 130 | 15 | 17 |
| 40 | 27 | 12 | 140 | 21 | 15 |
| 50 | 24 | 21 | 150 | 10 | 21 |
| 60 | 58 | 39 | 160 | 7 | 10 |
| 70 | 66 | 57 | 170 | 5 | 13 |
| 80 | 43 | 57 | 180 | 8 | 11 |
| 90 | 44 | 43 | 190 | 11 | 12 |
| 100 | 13 | 26 | 200 | 2 | 5 |

Варіант 10

| Місячні прибутки X , ум. од. | Кількість опитаних | | Місячні прибутки X , ум. од. | Кількість опитаних | |
|--------------------------------------|--------------------|---------|--------------------------------------|--------------------|---------|
| | 2000 р. | 2006 р. | | 2000 р. | 2006 р. |
| 10 | 1 | 0 | 110 | 15 | 26 |
| 20 | 4 | 5 | 120 | 16 | 20 |
| 30 | 9 | 4 | 130 | 16 | 13 |
| 40 | 16 | 15 | 140 | 14 | 14 |
| 50 | 32 | 26 | 150 | 12 | 18 |
| 60 | 50 | 43 | 160 | 14 | 13 |
| 70 | 77 | 50 | 170 | 8 | 10 |
| 80 | 39 | 57 | 180 | 4 | 8 |
| 90 | 33 | 44 | 190 | 10 | 8 |
| 100 | 23 | 20 | 200 | 7 | 6 |

1. Здійснити статистичний аналіз отриманих даних окремо для 2000 р. та 2006 р. Для цього:

- знайти вибіркові середнє, моду та медіану місячних прибутків;
- знайти вибіркові дисперсію, стандартне відхилення, середнє абсолютне відхилення, виправлене стандартне відхилення, коефіцієнт варіації та розмах варіації місячних прибутків;
- побудувати полігон, гістограму та кумуляту частот вибіркового розподілу місячних прибутків.

2. Окремо для 2000 р. та 2006 р. за допомогою χ^2 -критерію Пірсона з рівнем значущості 5 % перевірити, чи узгоджений вибірковий розподіл місячних прибутків з пуассоновим розподілом. У разі узгодженості знайти точкову оцінку параметра λ -розподілу Пуассона.

3. Окремо для 2000 р. та 2006 р. знайти 95 %-й довірчий інтервал для середнього значення місячних прибутків.

4. Окремо для 2000 р. та 2006 р. за допомогою λ -критерію Колмогорова — Смірнова на рівні значущості 5 % перевірити, чи узгоджений вибірковий розподіл місячних прибутків з нормальним розподілом. У разі узгодженості знайти:

- точкові оцінки параметрів (μ , σ) нормального розподілу;
- 95 %-й довірчий інтервал для стандартного відхилення місячних прибутків.

5. За допомогою кутового перетворення Фішера в поєднанні з розрахунком λ -критерію Колмогорова — Смірнова на рівні значущості 5 % перевірити наявність позитивного зсуву в місячних прибутках.

Література [5, розд. 6, 7; 8, розд. 15]

ПРИКЛАДИ РОЗВ'ЯЗАННЯ ЗАДАЧ

Приклад 1

Література [5, розд. 2, 5, 6; 8, розд. 12, 13]

У деякій країні є чотири політичні партії: білі, червоні, зелені та чорні. Незалежна соціологічна служба провела опитування щодо політичних симпатій виборців країни. Країна складається з двох великих регіонів: Заходу і Сходу. Оскільки кожний з цих регіонів має власні історичні, релігійні та культурні традиції, соціологічна служба вирішила поділити дослідження на дві частини: опитування на Заході та Сході. Обсяг опитаної випадкової вибірки в західному регіоні становив 2500 респондентів, у східному – 2300. Кожний респондент вказував партію, прихильником якої був у 2000 р. та 2004 р. Дані дослідження наведені в таблиці.

| Політична партія | Кількість прихильників за роками та регіонами | | | |
|------------------|---|----------------|-----------------|----------------|
| | 2000 р. | | 2006 р. | |
| | Західний регіон | Східний регіон | Західний регіон | Східний регіон |
| Білі | 585 | 586 | 623 | 607 |
| Червоні | 639 | 521 | 604 | 537 |
| Зелені | 594 | 553 | 539 | 478 |
| Чорні | 682 | 640 | 734 | 678 |

Враховуючи, що співвідношення між кількостями виборців на Заході та Сході становить 46 % до 54 %, для кожної партії знайти точкову оцінку генеральної пропорції її прихильників:

- у 2000 р. та 2006 р. окремо в кожному регіоні;
- у 2000 р. та 2006 р. по країні разом.

2. Для кожної партії визначити 95 %-й довірчий інтервал для генеральної пропорції її прихильників:

- у 2000 р. та 2006 р. окремо у кожному регіоні;
- у 2000 р. та 2006 р. по країні разом.

3. Для кожної партії визначити 95 %-й довірчий інтервал для різниці генеральних пропорцій її прихильників на Заході та Сході:

- у 2000 р.;
- у 2006 р.

4. За допомогою χ^2 -критерію Пірсона перевірити на рівні значущості 5 %, чи узгоджений з рівномірним розподілом розподіл кількості прихильників між політичними партіями:

- на Заході у 2006 р.;
- на Сході у 2006 р.;
- по країні разом у 2006 р.

Розв'язання

1. Обчислимо точкові оцінки генеральних пропорцій прихильників партій за регіонами у 2000 р. як відношення кількості прихильників партій у вибірці до обсягу вибірки.

| Політична партія | Відношення кількості прихильників партії у вибірці до обсягу вибірки | |
|------------------|--|------------------------------------|
| | Захід | Схід |
| Білі | $\frac{585}{2500} = 23,40\%$ | $\frac{586}{2300} \approx 25,48\%$ |
| Червоні | $\frac{639}{2500} = 25,56\%$ | $\frac{586}{2300} \approx 25,48\%$ |
| Зелені | $\frac{594}{2500} = 23,76\%$ | $\frac{586}{2300} \approx 25,48\%$ |
| Чорні | $\frac{682}{2500} = 27,28\%$ | $\frac{586}{2300} \approx 25,48\%$ |

Для обчислення точкових оцінок генеральних пропорцій прихильників партій у 2000 р. по країні загалом потрібно врахувати співвідношення між кількістю виборців на Заході та Сході 46% : 54%. Для цього використаємо таку формулу:

$$P_{\text{зар}} = 46\% \cdot P_{\text{Захід}} + 54\% \cdot P_{\text{Схід}}$$

Точкові оцінки генеральних пропорцій прихильників партій у 2000 р. по країні загалом та за регіонами наведені в таблиці.

| Політична партія | Генеральна пропорція прихильників по країні |
|------------------|---|
| Білі | $46\% \cdot 23,40\% + 54\% \cdot 25,48\% \approx 24,52\%$ |
| Червоні | $46\% \cdot 25,56\% + 54\% \cdot 22,65\% \approx 23,99\%$ |
| Зелені | $46\% \cdot 23,76\% + 54\% \cdot 24,04\% \approx 23,91\%$ |
| Чорні | $46\% \cdot 27,28\% + 54\% \cdot 27,83\% \approx 27,57\%$ |

| Політична партія | Генеральна пропорція прихильників за регіонами | |
|------------------|--|--------------------------------|
| | Захід | Схід |
| Білі | $\frac{623}{2500} = 24,92\%$ | $\frac{607}{2300} \gg 26,39\%$ |
| Червоні | $\frac{604}{2500} = 24,16\%$ | $\frac{537}{2300} \gg 23,35\%$ |
| Зелені | $\frac{539}{2500} = 21,56\%$ | $\frac{478}{2300} \gg 20,78\%$ |
| Чорні | $\frac{734}{2500} = 29,36\%$ | $\frac{678}{2300} \gg 29,48\%$ |

Точкові оцінки генеральних пропорцій прихильників партій по країні загалом у 2006 р. наведені в таблиці.

| Політична партія | Генеральна пропорція прихильників по країні |
|------------------|---|
| Білі | $46\% \cdot 24,92\% + 54\% \cdot 26,39\% \approx 25,71\%$ |
| Червоні | $46\% \cdot 24,16\% + 54\% \cdot 23,35\% \approx 23,72\%$ |
| Зелені | $46\% \cdot 21,56\% + 54\% \cdot 20,78\% \approx 21,14\%$ |
| Чорні | $46\% \cdot 29,36\% + 54\% \cdot 29,48\% \approx 29,42\%$ |

2. Інтервальні оцінки генеральних пропорцій

$$P_{\text{точ}} - \Delta \leq P_{\text{ген}} \leq P_{\text{точ}} + \Delta.$$

Обчислюємо:

$$D = 1,64 \sqrt{\frac{P_{\text{точ}} (1 - P_{\text{точ}})}{N}},$$

де 1,64 – константа, що відповідає рівню довіри до інтервальної оцінки 95 %; N – обсяг відповідної вибірки.

При цьому в разі обчислення оцінки по країні загалом як N беремо не сумарний обсяг вибірок 4800, а число, яке враховує співвідношення виборців на Заході та Сході:

$$N = \frac{2300}{54} \gg 4260.$$

Обчислимо інтервальні оцінки генеральних пропорцій прихильників партій для 2000 р. у західному регіоні. Для цього обчислимо спочатку Δ у вигляді таблиці.

| Політична партія | Δ |
|------------------|---|
| Білі | $1,64 \times \sqrt{\frac{23,40\% \times (1 - 23,40\%)}{2500}} \approx 1,39\%$ |
| Червоні | $1,64 \times \sqrt{\frac{25,56\% \times (1 - 25,56\%)}{2500}} \approx 1,43\%$ |
| Зелені | $1,64 \times \sqrt{\frac{23,76\% \times (1 - 23,76\%)}{2500}} \approx 1,40\%$ |
| Чорні | $1,64 \times \sqrt{\frac{27,28\% \times (1 - 27,28\%)}{2500}} \approx 1,46\%$ |

Остаточні інтервальні оцінки генеральних пропорцій прихильників партій у 2000 р. у західному регіоні наведені в таблиці.

| Політична партія | Інтервальна оцінка генеральної пропорції прихильників | |
|------------------|---|------------------------------|
| | Нижня межа | Верхня межа |
| Білі | $23,40\% - 1,39\% = 22,01\%$ | $23,40\% + 1,39\% = 24,79\%$ |
| Червоні | $25,56\% - 1,43\% = 24,13\%$ | $25,56\% + 1,43\% = 26,99\%$ |
| Зелені | $23,76\% - 1,40\% = 22,36\%$ | $23,76\% + 1,40\% = 25,16\%$ |
| Чорні | $27,28\% - 1,46\% = 25,82\%$ | $27,28\% + 1,46\% = 28,74\%$ |

Виконавши такі самі дії для східного регіону, у 2000 р. матимемо значення Δ , наведені в таблиці.

| Політична партія | Δ |
|------------------|---|
| Білі | $1,64 \times \sqrt{\frac{25,48\% \times (1 - 25,48\%)}{2300}} \approx 1,49\%$ |
| Червоні | $1,64 \times \sqrt{\frac{22,65\% \times (1 - 22,65\%)}{2300}} \approx 1,46\%$ |
| Зелені | $1,64 \times \sqrt{\frac{24,04\% \times (1 - 24,04\%)}{2300}} \approx 1,41\%$ |
| Чорні | $1,64 \times \sqrt{\frac{27,83\% \times (1 - 27,83\%)}{2300}} \approx 1,56\%$ |

Інтервальні оцінки генеральних пропорцій прихильників партій наведені в таблиці.

| Політична партія | Інтервальна оцінка генеральної пропорції прихильників | |
|------------------|---|-------------------------|
| | Нижня межа | Верхня межа |
| Білі | 25,48% – 1,49% = 23,99% | 25,48% + 1,49% = 26,97% |
| Червоні | 22,65% – 1,46% = 21,19% | 22,65% + 1,46% = 24,12% |
| Зелені | 24,04% – 1,41% = 22,64% | 24,04% + 1,41% = 25,45% |
| Чорні | 27,83% – 1,56% = 26,27% | 27,83% + 1,56% = 29,38% |

Для країни загалом у 2000 р. визначимо Δ та інтервальні оцінки генеральних пропорцій прихильників партій у вигляді таблиць.

| Політична партія | Δ |
|------------------|---|
| Білі | $1,64 \times \sqrt{\frac{24,52\% \times (1 - 24,52\%)}{4260}} \approx 1,08\%$ |
| Червоні | $1,64 \times \sqrt{\frac{23,99\% \times (1 - 23,99\%)}{4260}} \approx 1,07\%$ |
| Зелені | $1,64 \times \sqrt{\frac{23,91\% \times (1 - 23,91\%)}{4260}} \approx 1,07\%$ |
| Чорні | $1,64 \times \sqrt{\frac{27,57\% \times (1 - 27,57\%)}{4260}} \approx 1,12\%$ |

| Політична партія | Інтервальна оцінка генеральної пропорції прихильників | |
|------------------|---|-------------------------|
| | Нижня межа | Верхня межа |
| Білі | 24,52% – 1,08% = 23,44% | 24,52% + 1,08% = 25,60% |
| Червоні | 23,99% – 1,07% = 22,92% | 23,99% + 1,07% = 25,06% |
| Зелені | 23,91% – 1,07% = 22,84% | 23,91% + 1,07% = 24,98% |
| Чорні | 27,57% – 1,12% = 26,45% | 27,57% + 1,12% = 28,70% |

Довірчі інтервали для генеральних пропорцій прихильників партій у 2006 р. обчислюються аналогічно, тому наведемо лише остаточні результати.

На Заході.

| Політична партія | Δ | Інтервальна оцінка генеральної пропорції прихильників | |
|------------------|----------|---|-------------|
| | | Нижня межа | Верхня межа |
| Білі | 1,42 % | 23,50 % | 26,34 % |
| Червоні | 1,40 % | 22,76 % | 25,56 % |
| Зелені | 1,35 % | 20,21 % | 22,91 % |
| Чорні | 1,49 % | 27,87 % | 30,85 % |

На Сході.

| Політична партія | Δ | Інтервальна оцінка генеральної пропорції прихильників | |
|------------------|----------|---|-------------|
| | | Нижня межа | Верхня межа |
| Білі | 1,51 % | 24,88 % | 27,90 % |
| Червоні | 1,45 % | 21,90 % | 24,79 % |
| Зелені | 1,39 % | 19,40 % | 22,17 % |
| Чорні | 1,56 % | 27,92 % | 31,04 % |

По країні загалом.

| Політична партія | Δ | Інтервальна оцінка генеральної пропорції прихильників | |
|------------------|----------|---|-------------|
| | | Нижня межа | Верхня межа |
| Білі | 1,10 % | 24,62 % | 26,81 % |
| Червоні | 1,07 % | 22,65 % | 24,79 % |
| Зелені | 1,03 % | 20,11 % | 22,17 % |
| Чорні | 1,15 % | 28,28 % | 30,57 % |

3. Для знаходження 95 % довірчих інтервалів для різниці генеральних пропорцій прихильників партій на Заході та Сході скористаємося такими формулами:

$$\Delta P_{\text{точ}} - \Delta \leq \Delta P_{\text{ген}} \leq \Delta P_{\text{точ}} + \Delta;$$

$$\Delta P_{\text{точ}} = P_{\text{схід}} - P_{\text{захід}};$$

$$D = 1,64 \sqrt{\frac{P_{\text{схід}} (1 - P_{\text{схід}})}{N_{\text{схід}}} + \frac{P_{\text{захід}} (1 - P_{\text{захід}})}{N_{\text{захід}}}},$$

де $N_{\text{схід}}$, $N_{\text{захід}}$ – обсяг вибірки відповідно східної та західної.

Для 2000 р. розрахунки $\Delta P_{\text{точ}}$ та Δ наведені в таблицях.

| Політична партія | $\Delta P_{\text{точ}}$ |
|------------------|--------------------------|
| Білі | 25,48% - 23,40% = 2,08% |
| Червоні | 22,65% - 25,56% = -2,91% |
| Зелені | 24,04% - 23,76% = 0,28% |
| Чорні | 27,83% - 27,28% = 0,55% |

| Політична партія | Δ |
|------------------|---|
| Білі | $1,64 \times \sqrt{\frac{25,48\% \times (1 - 25,48\%)}{2300} + \frac{23,40\% \times (1 - 23,40\%)}{2500}} \approx 2,04\%$ |
| Червоні | $1,64 \times \sqrt{\frac{22,65\% \times (1 - 22,65\%)}{2300} + \frac{25,56\% \times (1 - 25,56\%)}{2500}} \approx 2,02\%$ |
| Зелені | $1,64 \times \sqrt{\frac{24,04\% \times (1 - 24,04\%)}{2300} + \frac{23,76\% \times (1 - 23,76\%)}{2500}} \approx 2,02\%$ |
| Чорні | $1,64 \times \sqrt{\frac{27,83\% \times (1 - 27,83\%)}{2300} + \frac{27,28\% \times (1 - 27,28\%)}{2500}} \approx 2,12\%$ |

Остаточні 95 %-ві довірчі інтервали для різниці генеральних пропорцій прихильників партій на Заході та Сході у 2000 р. наведені в таблиці.

| Політична партія | Інтервальна оцінка генеральної пропорції прихильників | |
|------------------|---|------------------------------|
| | Нижня межа | Верхня межа |
| Білі | $2,08\% - 2,04\% = 0,04\%$ | $2,08\% + 2,04\% = 4,12\%$ |
| Червоні | $-2,91\% - 2,02\% = -4,93\%$ | $-2,91\% + 2,02\% = -0,88\%$ |
| Зелені | $0,28\% - 2,02\% = -1,74\%$ | $0,28\% + 2,02\% = 2,30\%$ |
| Чорні | $0,55\% - 2,12\% = -1,57\%$ | $0,55\% + 2,12\% = 2,66\%$ |

Аналогічно виконуються розрахунки для 2006 р. Наведемо лише остаточні результати.

| Політична партія | $\Delta P_{\text{точ}}$ | Δ | Інтервальна оцінка генеральної пропорції прихильників | |
|------------------|-------------------------|----------|---|-------------|
| | | | Нижня межа | Верхня межа |
| Білі | 1,5 % | 2,07 % | -0,60 % | 3,54 % |
| Червоні | -0,8 % | 2,02 % | -2,83 % | 1,20 % |
| Зелені | -0,8 % | 1,94 % | -2,71 % | 1,16 % |
| Чорні | 0,1 % | 2,16 % | -2,04 % | 2,28 % |

4. Емпіричне значення χ^2 -критерію Пірсона

$$\chi^2_{\text{емп}} = \frac{(n_{\text{емп}} - n_{\text{теор}})^2}{n_{\text{теор}}},$$

де сума береться за всіма можливими значеннями ознаки; $n_{\text{емп}}$ — частота конкретного значення ознаки у вибірці; $n_{\text{теор}}$ — теоретична частота значення.

Для порівняння емпіричного розподілу з рівномірним маємо

$$n_{\text{теор}} = \frac{N}{m},$$

де N – обсяг вибірки; m – кількість значень ознаки.

Критичне значення критерію Пірсона визначається за допомогою таблиць¹ згідно з кількістю ступенів вільності, яка дорівнює $m - 1$, та рівнем значущості.

У розглядуваному випадку кількість ступенів вільності $4 - 1 = 3$; рівень значущості за умовою дорівнює 5%. Тоді критичне значення критерію Пірсона дорівнює 7,82.

Якщо емпіричне значення критерію перевищить критичне (у розглядуваному випадку $> 7,82$), слід прийняти гіпотезу про відмінність досліджуваного розподілу від рівномірного (на рівні значущості 5%).

Насамперед зауважимо, що

$$n_{\text{теор}} = \frac{2500}{4} = 625.$$

Процедуру обчислення емпіричного значення χ^2 -критерію наведемо в таблиці.

| Політична партія | $n_{\text{емп}}$ | $n_{\text{емп}} - n_{\text{теор}}$ | $(n_{\text{емп}} - n_{\text{теор}})^2$ | $\frac{(n_{\text{емп}} - n_{\text{теор}})^2}{n_{\text{теор}}}$ |
|-----------------------|------------------|------------------------------------|--|--|
| Білі | 623 | -2 | 4 | 0,01 |
| Червоні | 604 | -21 | 441 | 0,71 |
| Зелені | 539 | -86 | 7396 | 11,83 |
| Чорні | 734 | 109 | 11881 | 19,01 |
| $\chi^2_{\text{емп}}$ | | | | 31,56 |

Оскільки емпіричне значення критерію перевищує критичне, приймаємо гіпотезу про відмінність розподілу кількості прихильників різних політичних партій на Заході у 2006 р. від рівномірного.

Для Сходу

$$n_{\text{теор}} = \frac{2300}{4} = 575.$$

Емпіричне значення χ^2 -критерію наведено в таблиці.

¹ Див., наприклад, [18] або [25].

| Політична партія | $n_{\text{емп}}$ | $n_{\text{емп}} - n_{\text{теор}}$ | $(n_{\text{емп}} - n_{\text{теор}})^2$ | $\frac{(n_{\text{емп}} - n_{\text{теор}})^2}{n_{\text{теор}}}$ |
|-----------------------|------------------|------------------------------------|--|--|
| Білі | 607 | 32 | 1024 | 1,78 |
| Червоні | 537 | -38 | 1444 | 2,51 |
| Зелені | 478 | -97 | 9409 | 16,36 |
| Чорні | 678 | 103 | 10609 | 18,45 |
| $\chi^2_{\text{емп}}$ | | | | 39,10 |

І в цьому разі (2006 р., Схід) розподіл кількості прихильників різних політичних партій відмінний від рівномірного.

Для країни загалом слід врахувати співвідношення виборців на Сході та Заході. Зокрема (див. п. 2), обсяг вибірки дорівнює 4260. Тоді для знаходження емпіричних частот нам знадобиться формула

$$n_{\text{емп}} = P_{\text{точ}} \cdot 4260,$$

де $P_{\text{точ}}$ — точкова оцінка пропорції прихильників партії по країні загалом у 2006 р. (див. п. 2).

Крім того,

$$n_{\text{теор}} = \frac{4260}{4} = 1065.$$

Емпіричне значення χ^2 -критерію наведено в таблиці.

| Політична партія | $n_{\text{емп}}$ | $n_{\text{емп}} - n_{\text{теор}}$ | $(n_{\text{емп}} - n_{\text{теор}})^2$ | $\frac{(n_{\text{емп}} - n_{\text{теор}})^2}{n_{\text{теор}}}$ |
|-----------------------|------------------|------------------------------------|--|--|
| Білі | 1095 | 30,44 | 926,46 | 0,87 |
| Червоні | 1011 | -54,47 | 2966,68 | 2,79 |
| Зелені | 901 | -164,43 | 27036,27 | 25,39 |
| Чорні | 1253 | 188,46 | 35515,84 | 33,35 |
| $\chi^2_{\text{емп}}$ | | | | 62,40 |

Отже, і по всій країні у 2004 р. розподіл кількості прихильників різних політичних партій відмінний від рівномірного.

Приклад 2

Література: [5, розд. 6, 7; 8, розд. 15]

Незалежна соціологічна служба провела опитування 400 випадкових респондентів працездатного віку в центральному регіоні країни про рівень їх щомісячних прибутків X у 2000 р. та 2006 р. (з точністю до 10 ум. од.). Дані опитування наведені в таблиці.

| Місячні прибутки X, ум. од. | Кількість опитаних | | Місячні прибутки X, ум. од. | Кількість опитаних | |
|-----------------------------------|--------------------|---------|-----------------------------------|--------------------|---------|
| | 2000 р. | 2006 р. | | 2000 р. | 2006 р. |
| 10 | 4 | 0 | 110 | 18 | 17 |
| 20 | 8 | 4 | 120 | 12 | 19 |
| 30 | 6 | 2 | 130 | 25 | 13 |
| 40 | 24 | 16 | 140 | 13 | 14 |
| 50 | 22 | 35 | 150 | 7 | 10 |
| 60 | 48 | 56 | 160 | 14 | 9 |
| 70 | 67 | 40 | 170 | 4 | 12 |
| 80 | 47 | 57 | 180 | 11 | 8 |
| 90 | 37 | 42 | 190 | 4 | 12 |
| 100 | 23 | 26 | 200 | 6 | 8 |

1. Здійснити статистичний аналіз отриманих даних окремо для 2000 р. та 2006 р. Для цього виконати такі дії:

- обчислити вибіркові середнє, моду та медіану місячних прибутків;
- обчислити вибіркові дисперсію, стандартне відхилення, середнє абсолютне відхилення, виправлене стандартне відхилення, коефіцієнт варіації та розмах варіації місячних прибутків;
- побудувати полігон, гістограму та кумуляту частот вибіркового розподілу місячних прибутків.

2. Окремо для 2000 р. та 2006 р. за допомогою χ^2 -критерію Пірсона з рівнем значущості 5 % перевірити, чи узгоджений вибірковий розподіл місячних прибутків з пуассоновим розподілом. За узгодженості обчислити точкову оцінку параметра λ -розподілу Пуассона.

3. Окремо для 2000 р. та 2006 р. знайти 95 %-й довірчий інтервал для середнього значення місячних прибутків.

4. Окремо для 2000 р. та 2006 р. за допомогою λ -критерію Колмогорова — Смірнова на рівні значущості 5 % перевірити, чи узгоджений вибірковий розподіл місячних прибутків з нормальним розподілом. За узгодженості обчислити:

- точкові оцінки параметрів (μ , σ) нормального розподілу;
- 95 %-й довірчий інтервал для стандартного відхилення місячних прибутків.

5. За допомогою кутового перетворення Фішера в поєднанні з процедурою розрахунку λ -критерію Колмогорова — Смірнова на рівні значущості 5 % перевірити наявність позитивного зсуву в місячних прибутках.

Розв'язання

1. Нагадаємо, що вибіркове середнє — це звичайне середнє арифметичне значення, мода — значення, яке зустрічається найчастіше, медіана — це число, яке в упорядкованому списку індивідуальних значень розміщується посередині.

Обчислимо середнє, моду та медіану місячних прибутків у 2000 р. за допомогою таблиці.

| Місячні прибутки X | Частота n | Добуток Xn | Накопичена частота cn |
|-------------------------|-------------|----------------------------|-------------------------|
| 10 | 4 | 40 | 4 |
| 20 | 8 | 160 | $4 + 8 = 12$ |
| 30 | 6 | 180 | $12 + 6 = 18$ |
| 40 | 24 | 960 | $18 + 24 = 42$ |
| 50 | 22 | 1100 | $42 + 22 = 64$ |
| 60 | 48 | 2880 | $64 + 48 = 112$ |
| 70 | 67 | 4690 | $112 + 67 = 179$ |
| 80 | 47 | 3760 | $179 + 47 = 226$ |
| 90 | 37 | 3330 | $226 + 37 = 263$ |
| 100 | 23 | 2300 | $263 + 23 = 286$ |
| 110 | 18 | 1980 | $286 + 18 = 304$ |
| 120 | 12 | 1440 | $304 + 12 = 316$ |
| 130 | 25 | 3250 | $316 + 25 = 341$ |
| 140 | 13 | 1820 | $341 + 13 = 354$ |
| 150 | 7 | 1050 | $354 + 7 = 361$ |
| 160 | 14 | 2240 | $361 + 14 = 375$ |
| 170 | 4 | 680 | $375 + 4 = 379$ |
| 180 | 11 | 1980 | $379 + 11 = 390$ |
| 190 | 4 | 760 | $390 + 4 = 394$ |
| 200 | 6 | 1200 | $394 + 6 = 400$ |
| Разом | 400 | 35800 | |
| | Середнє | $\frac{35800}{400} = 89,5$ | |

Оскільки значення 70 зустрічається 67 разів, мода дорівнює 70.

Для того щоб визначити медіану, потрібно розглянути значення в середині списку, тобто на 200-й та 201-й позиціях. З цією метою обчислюємо стовпець накопичених частот. Шукаємо місячні прибутки, для яких накопичена частота вперше перетинає число 200. Це — 80 ум. од.

Отже, медіана дорівнює 80.

Так само обчислюємо середнє, моду, медіану для 2006 р.

| Місячні прибутки X | Частота n | Добуток Xn | Накопичена частота cn |
|----------------------|-------------|------------------------------|-------------------------|
| 10 | 0 | 0 | 0 |
| 20 | 4 | 80 | $0 + 4 = 4$ |
| 30 | 2 | 60 | $4 + 2 = 6$ |
| 40 | 16 | 640 | $6 + 16 = 22$ |
| 50 | 35 | 1750 | $22 + 35 = 57$ |
| 60 | 56 | 3360 | $57 + 56 = 113$ |
| 70 | 40 | 2800 | $113 + 40 = 153$ |
| 80 | 57 | 4560 | $153 + 57 = 210$ |
| 90 | 42 | 3780 | $210 + 42 = 252$ |
| 100 | 26 | 2600 | $252 + 26 = 278$ |
| 110 | 17 | 1870 | $278 + 17 = 295$ |
| 120 | 19 | 2280 | $295 + 19 = 314$ |
| 130 | 13 | 1690 | $314 + 13 = 327$ |
| 140 | 14 | 1960 | $327 + 14 = 341$ |
| 150 | 10 | 1500 | $341 + 10 = 351$ |
| 160 | 9 | 1440 | $351 + 9 = 360$ |
| 170 | 12 | 2040 | $360 + 12 = 372$ |
| 180 | 8 | 1440 | $372 + 8 = 380$ |
| 190 | 12 | 2280 | $380 + 12 = 392$ |
| 200 | 8 | 1600 | $392 + 8 = 400$ |
| Разом | 400 | 37730 | |
| Середнє | | $\frac{37730}{400} = 94,325$ | |

Очевидно, що мода і медіана в цьому разі збігаються і дорівнюють 80.

Нагадаємо, що середнє абсолютне відхилення дорівнює середньому значенню модулів індивідуальних відхилень від вибіркового середнього; дисперсія D — це середнє значення квадратів індивідуальних відхилень; стандартне відхилення σ обчислюється як корінь з дисперсії; виправлене стандартне відхилення s

$$s = \sigma \sqrt{\frac{n}{n-1}},$$

де n — обсяг вибірки; коефіцієнт варіації — це відношення стандартного відхилення до середнього значення; розмах варіації дорівнює різниці найбільшого і найменшого індивідуального значення у вибірці.

Розглянемо дані для 2000 р. Виконаємо попередні обчислення в таблиці.

| X | n | Відхилення $x = X - \bar{X}$ | $ x n$ | x^2 | x^2n |
|------------------------------|-----|------------------------------|--------|--------------------------------|----------|
| 10 | 4 | -79,5 | 318 | 6320,25 | 25281,00 |
| 20 | 8 | -69,5 | 556 | 4830,25 | 38642,00 |
| 30 | 6 | -59,5 | 357 | 3540,25 | 21241,50 |
| 40 | 24 | -49,5 | 1188 | 2450,25 | 58806,00 |
| 50 | 22 | -39,5 | 869 | 1560,25 | 34325,50 |
| 60 | 48 | -29,5 | 1416 | 870,25 | 41772,00 |
| 70 | 67 | -19,5 | 1306,5 | 380,25 | 25476,75 |
| 80 | 47 | -9,5 | 446,5 | 90,25 | 4241,75 |
| 90 | 37 | 0,5 | 18,5 | 0,25 | 9,25 |
| 100 | 23 | 10,5 | 241,5 | 110,25 | 2535,75 |
| 110 | 18 | 20,5 | 369 | 420,25 | 7564,50 |
| 120 | 12 | 30,5 | 366 | 930,25 | 11163,00 |
| 130 | 25 | 40,5 | 1012,5 | 1640,25 | 41006,25 |
| 140 | 13 | 50,5 | 656,5 | 2550,25 | 33153,25 |
| 150 | 7 | 60,5 | 423,5 | 3660,25 | 25621,75 |
| 160 | 14 | 70,5 | 987 | 4970,25 | 69583,50 |
| 170 | 4 | 80,5 | 322 | 6480,25 | 25921,00 |
| 180 | 11 | 90,5 | 995,5 | 8190,25 | 90092,75 |
| 190 | 4 | 100,5 | 402 | 10100,25 | 40401,00 |
| 200 | 6 | 110,5 | 663 | 12210,25 | 73261,50 |
| Разом | 400 | | 12914 | | 670100 |
| Середнє абсолютне відхилення | | | | $\frac{12914}{400} = 32,285$ | |
| Дисперсія | | | | $\frac{670100}{400} = 1675,25$ | |
| Стандартне відхилення | | | | $\sqrt{1675,25} \approx 40,93$ | |

Виправлене стандартне відхилення

$$s = 40,93 \sqrt{\frac{400}{399}} \approx 40,98.$$

Коефіцієнт варіації

$$k = \frac{40,93}{89,5} \approx 0,457 = 45,7\%.$$

Найбільші місячні прибутки у вибірці 200, найменші – 10, тоді розмах варіації

$$R = 200 - 10 = 190.$$

Розглянемо тепер дані для 2006 р.

| X | n | Відхилення $x = X - \bar{X}$ | $ x n$ | x^2 | x^2n |
|------------------------------|-----|------------------------------|----------|---|-----------|
| 10 | 0 | -84,325 | 0 | 7110,71 | 0,00 |
| 20 | 4 | -74,325 | 297,3 | 5524,21 | 22096,82 |
| 30 | 2 | -64,325 | 128,65 | 4137,71 | 8275,41 |
| 40 | 16 | -54,325 | 869,2 | 2951,21 | 47219,29 |
| 50 | 35 | -44,325 | 1551,375 | 1964,71 | 68764,70 |
| 60 | 56 | -34,325 | 1922,2 | 1178,21 | 65979,52 |
| 70 | 40 | -24,325 | 973 | 591,71 | 23668,23 |
| 80 | 57 | -14,325 | 816,525 | 205,21 | 11696,72 |
| 90 | 42 | -4,325 | 181,65 | 18,71 | 785,64 |
| 100 | 26 | 5,675 | 147,55 | 32,21 | 837,35 |
| 110 | 17 | 15,675 | 266,475 | 245,71 | 4177,00 |
| 120 | 19 | 25,675 | 487,825 | 659,21 | 12524,91 |
| 130 | 13 | 35,675 | 463,775 | 1272,71 | 16545,17 |
| 140 | 14 | 45,675 | 639,45 | 2086,21 | 29206,88 |
| 150 | 10 | 55,675 | 556,75 | 3099,71 | 30997,06 |
| 160 | 9 | 65,675 | 591,075 | 4313,21 | 38818,85 |
| 170 | 12 | 75,675 | 908,1 | 5726,71 | 68720,47 |
| 180 | 8 | 85,675 | 685,4 | 7340,21 | 58721,65 |
| 190 | 12 | 95,675 | 1148,1 | 9153,71 | 109844,47 |
| 200 | 8 | 105,675 | 845,4 | 11167,21 | 89337,65 |
| Разом | 400 | | 13479,8 | | 708217,75 |
| Середнє абсолютне відхилення | | | | $\frac{13479,8}{400} = 33,7$ | |
| Дисперсія | | | | $\frac{708217,75}{400} \approx 1770,54$ | |
| Стандартне відхилення | | | | $\sqrt{1770,54} \approx 42,08$ | |

Виправлене стандартне відхилення, коефіцієнт варіації та розмах варіації відповідно

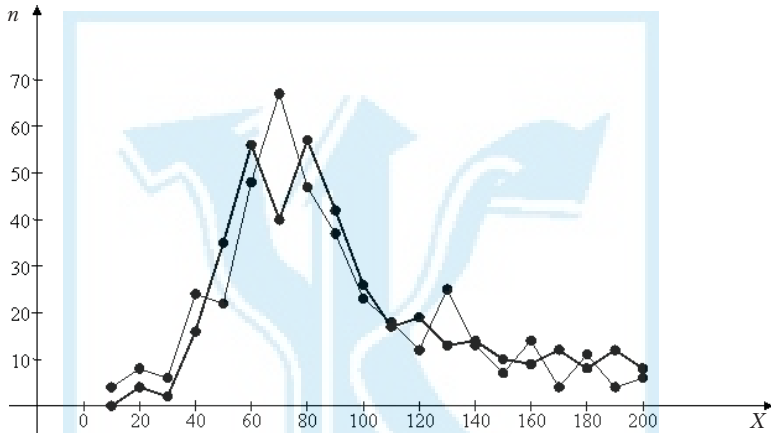
$$s = 42,08 \sqrt{\frac{400}{399}} \approx 42,13;$$

$$k = \frac{42,08}{94,325} \approx 0,446 = 44,6\%;$$

$$R = 200 - 20 = 180.$$

Нагадаємо, що лінія полігону частот з'єднує точки (x_i, n_i) , де x_i — значення i -го рівня ознаки; n_i — частота.

Побудуємо полігони частот розподілу місячних прибутків у 2000 р. (тонка лінія) та 2006 р. (жирна лінія).



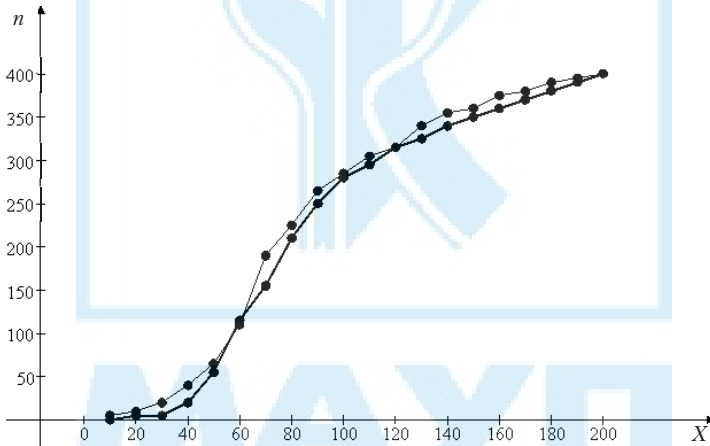
Гістограми легко побудувати за полігонами частот (див., наприклад, розв'язання прикладу 1) і тому не наводитимемо їх.

Для побудови кумулят частот потрібно обчислити попередньо накопичені частоти за роками (див. таблицю).

| X | 2000 р. | | 2006 р. | |
|----|-------------|-------------------------|-------------|-------------------------|
| | Частота n | Накопичена частота cn | Частота n | Накопичена частота cn |
| 1 | 2 | 3 | 4 | 5 |
| 10 | 4 | 4 | 0 | 0 |
| 20 | 8 | 4 + 8 = 12 | 4 | 0 + 4 = 4 |
| 30 | 6 | 12 + 6 = 18 | 2 | 4 + 2 = 6 |
| 40 | 24 | 18 + 24 = 42 | 16 | 6 + 16 = 22 |
| 50 | 22 | 42 + 22 = 64 | 35 | 22 + 35 = 57 |
| 60 | 48 | 64 + 48 = 112 | 56 | 57 + 56 = 113 |
| 70 | 67 | 112 + 67 = 179 | 40 | 113 + 40 = 153 |

| 1 | 2 | 3 | 4 | 5 |
|-----|----|------------------|----|------------------|
| 80 | 47 | $179 + 47 = 226$ | 57 | $153 + 57 = 210$ |
| 90 | 37 | $226 + 37 = 263$ | 42 | $210 + 42 = 252$ |
| 100 | 23 | $263 + 23 = 286$ | 26 | $252 + 26 = 278$ |
| 110 | 18 | $286 + 18 = 304$ | 17 | $278 + 17 = 295$ |
| 120 | 12 | $304 + 12 = 316$ | 19 | $295 + 19 = 314$ |
| 130 | 25 | $316 + 25 = 341$ | 13 | $314 + 13 = 327$ |
| 140 | 13 | $341 + 13 = 354$ | 14 | $327 + 14 = 341$ |
| 150 | 7 | $354 + 7 = 361$ | 10 | $341 + 10 = 351$ |
| 160 | 14 | $361 + 14 = 375$ | 9 | $351 + 9 = 360$ |
| 170 | 4 | $375 + 4 = 379$ | 12 | $360 + 12 = 372$ |
| 180 | 11 | $379 + 11 = 390$ | 8 | $372 + 8 = 380$ |
| 190 | 4 | $390 + 4 = 394$ | 12 | $380 + 12 = 392$ |
| 200 | 6 | $394 + 6 = 400$ | 8 | $392 + 8 = 400$ |

Побудуємо кумуляти частот для 2000 р. (тонка лінія) та 2006 р. (жирна лінія).



2. Аналізуючи полігони частот розподілів місячних прибутків X , бачимо, що у вибірці переважають малі значення ознаки X , а частота швидко зменшується зі збільшенням X . Такими самими властивостями характеризується розподіл Пуассона¹.

¹ Див., наприклад, [18].

Відомо, що середнє значення (точніше, математичне сподівання) та дисперсія розподілу Пуассона збігаються. Проте в розглядуваному випадку середнє та дисперсія відрізняються на порядок. Наприклад, у 2000 р. середнє становило 89,5, дисперсія — 1675,25. Отже, досліджувані розподіли з пуассоновими не збігаються.

Але, поділивши X на деяке число α , отримаємо, що середнє значення зменшиться у α разів, а дисперсія — у α^2 разів. Іншими словами, змінюючи масштаб X , можна досягти рівності середнього значення та дисперсії. При цьому масштабування X не змінює вигляд кривої його розподілу і означає фактично лише перехід до іншої грошової одиниці.

Обчислити число α можна, використовуючи рівняння

$$\frac{\bar{X}}{\alpha} = \frac{D(X)}{\alpha^2},$$

де \bar{X} — середнє; а $D(X)$ — дисперсія розподілу X .

Розв'язуючи його, маємо

$$\alpha = \frac{D(X)}{\bar{X}}.$$

Для 2000 р.

$$\alpha = \frac{1675,25}{89,5} \gg 18,7.$$

Для 2006 р.

$$\alpha = \frac{1770,54}{94,33} \gg 18,8.$$

Оскільки розподіл Пуассона — це розподіл цілочисельної величини, для спрощення подальших обчислень візьмемо коефіцієнт масштабування $\alpha = 20^1$.

При $\alpha = 20$ для 2000 р.

$$\bar{X} = \frac{89,5}{20} = 4,475;$$

$$D(X) = \frac{1675,25}{20} \gg 4,19,$$

¹ Якби дані стосувалися України, а під ум. од. розуміти, наприклад, долари США, то, враховуючи приблизне співвідношення між гривнею і долларом, як 5 до 1, ми б отримали опис місячних прибутків X у масштабі 100 грн.

для 2006 р.

$$\bar{X} = \frac{94,33}{20} \gg 4,72;$$

$$D(X) = \frac{1770,54}{20} \gg 4,43.$$

Як бачимо, після зменшення X у 20 разів, середнє та дисперсія різняться неістотно.

Після масштабування місячні прибутки X , які змінювались у вибірці від 10 до 200 з кроком 10, змінюватимуться від 0,5 до 10 з кроком 0,5. Але, як вже згадувалось, розподіл Пуассона — це розподіл цілочисельної величини. Тому потрібно мати розподіл ознаки X , яка може приймати значення 0, 1, 2 і т. д.

Для того щоб побудувати такий розподіл за відомими даними, вважатимемо, що серед людей, які мають таке значення ознаки X , половина насправді має дещо менше значення, а половина — дещо більше. Тоді для рівня t у шуканому розподілі його частоту обчислимо як суму всіх частот, що потрапляють в інтервал $(t - 0,5; t + 0,5)$.

Так, для 2000 р. отримаємо такий розподіл (він є моделлю вихідного і вже може порівнюватися з пуассоновим).

| X | Частота n |
|-----|---|
| 0 | $\frac{4}{2} = 2$ |
| 1 | $\frac{4}{2} + 8 + \frac{6}{2} = 13$ |
| 2 | $\frac{6}{2} + 24 + \frac{22}{2} = 38$ |
| 3 | $\frac{22}{2} + 48 + \frac{67}{2} = 92,5$ |
| 4 | $\frac{67}{2} + 47 + \frac{37}{2} = 99$ |
| 5 | $\frac{37}{2} + 23 + \frac{18}{2} = 50,5$ |
| 6 | $\frac{18}{2} + 12 + \frac{25}{2} = 33,5$ |

| X | Частота n |
|-------------|---|
| 7 | $\frac{25}{2} + 13 + \frac{7}{2} = 29$ |
| 8 | $\frac{7}{2} + 14 + \frac{4}{2} = 19,5$ |
| 9 | $\frac{4}{2} + 11 + \frac{4}{2} = 15$ |
| 10 і більше | $\frac{4}{2} + 6 = 8$ |
| Разом | 400 |

Частоту для $X = 5$ у цій таблиці отримано, наприклад, як суму половини частоти для $X = 4,5$ (вихідного 90), частоти для $X = 5$ (вихідного 100) та половини частоти для $X = 5,5$ (вихідного 110).

Діючи так само, для 2006 року отримаємо такий розподіл:

| X | Частота n |
|-----|---|
| 0 | $\frac{0}{2} = 0$ |
| 1 | $\frac{0}{2} + 4 + \frac{2}{2} = 5$ |
| 2 | $\frac{2}{2} + 16 + \frac{35}{2} = 34,5$ |
| 3 | $\frac{35}{2} + 56 + \frac{40}{2} = 93,5$ |
| 4 | $\frac{40}{2} + 57 + \frac{42}{2} = 98$ |
| 5 | $\frac{42}{2} + 26 + \frac{17}{2} = 55,5$ |
| 6 | $\frac{17}{2} + 19 + \frac{13}{2} = 34$ |
| 7 | $\frac{13}{2} + 14 + \frac{10}{2} = 25,5$ |
| 8 | $\frac{10}{2} + 9 + \frac{12}{2} = 20$ |

| | |
|-------------|--|
| X | Частота n |
| 9 | $\frac{12}{2} + 8 + \frac{12}{2} = 20$ |
| 10 і більше | $\frac{12}{2} + 8 = 14$ |
| Разом | 400 |

Отримані розподіли потрібно порівняти з теоретичними розподілами Пуассона. При цьому як параметр λ розподілу Пуассона братимемо точкову оцінку середнього значення. А саме для 2000 р. $\lambda = 4,475$, для 2006 р. — $\lambda = 4,72$.

Побудувати шукані теоретичні розподіли можна кількома шляхами: по-перше, можна скористатися формулою для розрахунку ймовірності, прийняти значення m у разі розподілу Пуассона з параметром λ :

$$P(m) = \frac{\lambda^m}{m!} e^{-\lambda}$$

Тут під $m!$ розуміємо m -факторіал:

$$m! = 1 \cdot 2 \cdot \dots \cdot m$$

По-друге, можна використати можливості деяких прикладних програм. Наприклад, у Microsoft Excel є функція ПУАССОН(), значенням якої є ймовірність прийняти це значення.

І, по-третє, можна скористатися таблицями функції Пуассона¹. Щоправда, у цих таблицях, як правило, наведено дані лише для певних значень λ . Зокрема, у цьому разі довелося б будувати розподіл Пуассона для $\lambda = 4,5$ і цей розподіл дещо відрізнявся б від шуканих.

Скористаємося другим підходом. Отримані теоретичні розподіли наведені в таблиці.

| X | 2000 р., $\lambda = 4,475$ | | 2006 р., $\lambda = 4,72$ | |
|---|--------------------------------|----------------------------------|--------------------------------|----------------------------------|
| | Відносна частота ωn | Частота, $\omega n \cdot 400$ | Відносна частота ωn | Частота, $\omega n \cdot 400$ |
| 1 | 2 | 3 | 4 | 5 |
| 0 | 0,01 | 4,6 | 0,01 | 3,6 |
| 1 | 0,05 | 20,4 | 0,04 | 16,9 |
| 2 | 0,11 | 45,6 | 0,10 | 39,8 |

¹ Див., наприклад, [18] або [25].

| | | | | |
|-------------|------|------|------|------|
| 1 | 2 | 3 | 4 | 5 |
| 3 | 0,17 | 68,0 | 0,16 | 62,6 |
| 4 | 0,19 | 76,1 | 0,18 | 73,8 |
| 5 | 0,17 | 68,1 | 0,17 | 69,6 |
| 6 | 0,13 | 50,8 | 0,14 | 54,7 |
| 7 | 0,08 | 32,5 | 0,09 | 36,9 |
| 8 | 0,05 | 18,2 | 0,05 | 21,7 |
| 9 | 0,02 | 9,0 | 0,03 | 11,4 |
| 10 і більше | 0,02 | 6,6 | 0,02 | 9,1 |
| Разом | 1 | 400 | 1 | 400 |

Перевіримо за допомогою критерію Пірсона узгодженість емпіричного та теоретичного розподілів для кожного року.

В обох випадках (2000 р. та 2006 р.) маємо такі гіпотези:

H_0 : Відмінності між емпіричним та теоретичним розподілами незначущі;

H_1 : Відмінності між емпіричним та теоретичним розподілами значущі.

В обох випадках кількість ступенів вільності дорівнює 10 (кількість рядків у таблиці розподілу мінус 1). Критичне значення критерію Пірсона для 10 ступенів вільності та рівня значущості 5 % дорівнює 18,31 (див. таблиці критерію Пірсона).

Знайдемо емпіричні значення критерію і порівняємо їх з теоретичними. Нагадаємо (див. розв'язання прикладу 1), що емпіричне значення критерію шукається як сума виразів, обчислених для всіх X :

$$D_{в.г}^2 f = \frac{(n_{\text{емп}} - n_{\text{теор}})^2}{n_{\text{теор}}},$$

де $n_{\text{емп}}$, $n_{\text{теор}}$ – частота розподілу відповідно до емпіричного та теоретичного.

Результати обчислень наведемо в таблицях.

Для 2000 р.

| X | $n_{\text{емп}}$ | $n_{\text{теор}}$ | $n_{\text{емп}} - n_{\text{теор}}$ | $(n_{\text{емп}} - n_{\text{теор}})^2$ | $\frac{(n_{\text{емп}} - n_{\text{теор}})^2}{n_{\text{теор}}}$ |
|-----|------------------|-------------------|------------------------------------|--|--|
| 1 | 2 | 3 | 4 | 5 | 6 |
| 0 | 2,0 | 4,6 | 2,6 | 6,53 | 1,43 |
| 1 | 13,0 | 20,4 | 7,4 | 54,59 | 2,68 |

| | | | | | |
|-----------------------|------|------|-------|--------|--------------|
| 1 | | | | | |
| 2 | 38,0 | 45,6 | 7,6 | 58,05 | 1,27 |
| 3 | 92,5 | 68,0 | -24,5 | 597,86 | 8,79 |
| 4 | 99,0 | 76,1 | -22,9 | 523,06 | 6,87 |
| 5 | 50,5 | 68,1 | 17,6 | 311,03 | 4,56 |
| 6 | 33,5 | 50,8 | 17,3 | 299,91 | 5,90 |
| 7 | 29,0 | 32,5 | 3,5 | 12,16 | 0,37 |
| 8 | 19,5 | 18,2 | -1,3 | 1,76 | 0,10 |
| 9 | 15,0 | 9,0 | -6,0 | 35,57 | 3,94 |
| 10 і більше | 8,0 | 6,6 | -1,4 | 1,94 | 0,29 |
| $\chi^2_{\text{емп}}$ | | | | | 36,21 |

Для 2006 р.

| Рівень X | $n_{\text{емп}}$ | $n_{\text{теор}}$ | $n_{\text{емп}} - n_{\text{теор}}$ | $(n_{\text{емп}} - n_{\text{теор}})^2$ | $\frac{(n_{\text{емп}} - n_{\text{теор}})^2}{n_{\text{теор}}}$ |
|-----------------------|------------------|-------------------|------------------------------------|--|--|
| 0 | 0,0 | 3,6 | 3,6 | 12,81 | 3,58 |
| 1 | 5,0 | 16,9 | 11,9 | 141,17 | 8,36 |
| 2 | 34,5 | 39,8 | 5,3 | 28,19 | 0,71 |
| 3 | 93,5 | 62,6 | -30,9 | 955,85 | 15,27 |
| 4 | 98,0 | 73,8 | -24,2 | 586,15 | 7,94 |
| 5 | 55,5 | 69,6 | 14,1 | 198,87 | 2,86 |
| 6 | 34,0 | 54,7 | 20,7 | 428,91 | 7,84 |
| 7 | 25,5 | 36,9 | 11,4 | 129,07 | 3,50 |
| 8 | 20,0 | 21,7 | 1,7 | 3,00 | 0,14 |
| 9 | 20,0 | 11,4 | -8,6 | 74,18 | 6,51 |
| 10 і більше | 14,0 | 9,1 | -4,9 | 24,34 | 2,69 |
| $\chi^2_{\text{емп}}$ | | | | | 59,40 |

Бачимо, що в обох випадках емпіричні значення критерію Пірсона (36,21 та 59,40) перевищують критичне значення 18,31. Отже, в обох випадках слід прийняти гіпотезу H_1 про відмінність емпіричних розподілів від теоретичних пуассонових¹.

¹ Зауважимо, що якщо досліджувалася узгодженість розподілів за допомогою критерію Колмогорова – Смірнова (див., наприклад, п. 4), слід було б відкинути гіпотезу H_1 на рівні значущості 22 % для 2000 р. та на рівні значущості 11 % – для 2006 р.

Іншими словами, в обох випадках слід було б прийняти гіпотезу про подібність емпіричних розподілів з теоретичними пуассоновими.

До речі, у цьому разі як точкову оцінку параметрів λ потрібно було б брати середні значення емпіричних розподілів. Як правило, якщо хоча б один критерій показує значущість відмінностей, цього достатньо.

3. Щоб обчислити довірчий інтервал для генерального середнього μ місячних прибутків, скористаємося формулою

$$\bar{X} - D \leq \mu \leq \bar{X} + D.$$

Обчислюємо:

$$D = z \frac{s}{\sqrt{n}},$$

де значення z залежить від заданого рівня довіри і в розглядуваному випадку (95 %) приблизно дорівнює 1,64; s — виправлене стандартне відхилення (див. п. 2 прикладу 1); $n = 400$ — обсяг вибірки.

Маємо

$$D_{2000} = 1,64 \frac{40,98}{\sqrt{400}} \approx 3,36;$$

$$D_{2006} = 1,64 \frac{42,13}{\sqrt{400}} \approx 3,45.$$

Тому з рівнем довіри 95 %

$$\mu_{2000} \in (89,5 - 3,36; 89,5 + 3,36) = (86,14; 92,86);$$

$$\mu_{2006} \in (94,325 - 3,45; 94,325 + 3,45) \approx (90,87; 97,78).$$

4. Для порівняння емпіричних розподілів з теоретичними нормальними попередньо потрібно визначити їх.

Зауважимо, що незміщеними точковими оцінками генерального середнього та генерального стандартного відхилення є відповідно середнє значення вибірки \bar{X} та виправлене стандартне відхилення s . Тому як теоретичні природно брати нормальні розподіли з параметрами (\bar{X}, s) .

На відміну від пуассонового нормальний розподіл — це розподіл неперервної випадкової величини. Тому потрібно знати не точкові частоти розподілів, а інтервальні. Зокрема, емпіричну частоту $X = 10$ природно вважати частотою для інтервалу $(-\infty; 15)$, $X = 20$ — частотою для інтервалу $(15; 25)$.

Іншими словами, емпіричні розподіли в інтервальній формі наведені в таблиці (оскільки для критерію Колмогорова — Смірнова використовуються відносні частоти розподілів, одразу обчислимо їх як відношення частот до обсягу вибірки 400).

З іншого боку, критерій Персона зазвичай застосовують для номінальних змінних. У розглядуваному випадку дані мають числову природу і використання критерію Колмогорова — Смірнова природніше.

| Інтервал | | 2000 р. | | 2006 р. | |
|------------|-------------|---------|------------------|---------|------------------|
| Нижня межа | Верхня межа | Частота | Відносна частота | Частота | Відносна частота |
| – ∞ | 15 | 4 | 0,01 | 0 | 0,00 |
| 15 | 25 | 8 | 0,02 | 4 | 0,01 |
| 25 | 35 | 6 | 0,02 | 2 | 0,01 |
| 35 | 45 | 24 | 0,06 | 16 | 0,04 |
| 45 | 55 | 22 | 0,06 | 35 | 0,09 |
| 55 | 65 | 48 | 0,12 | 56 | 0,14 |
| 65 | 75 | 67 | 0,17 | 40 | 0,10 |
| 75 | 85 | 47 | 0,12 | 57 | 0,14 |
| 85 | 95 | 37 | 0,09 | 42 | 0,11 |
| 95 | 105 | 23 | 0,06 | 26 | 0,07 |
| 105 | 115 | 18 | 0,05 | 17 | 0,04 |
| 115 | 125 | 12 | 0,03 | 19 | 0,05 |
| 125 | 135 | 25 | 0,06 | 13 | 0,03 |
| 135 | 145 | 13 | 0,03 | 14 | 0,04 |
| 145 | 155 | 7 | 0,02 | 10 | 0,03 |
| 155 | 165 | 14 | 0,04 | 9 | 0,02 |
| 165 | 175 | 4 | 0,01 | 12 | 0,03 |
| 175 | 185 | 11 | 0,03 | 8 | 0,02 |
| 185 | 195 | 4 | 0,01 | 12 | 0,03 |
| 195 | ∞ | 6 | 0,02 | 8 | 0,02 |
| Разом | | 400 | 1,00 | 400 | 1,00 |

Відносні інтервальні частоти теоретичних нормальних розподілів можна обчислити, наприклад, використовуючи таблицю значень інтегральної функції Лапласа Φ^1 , або інтегральну версію функції НОРМРАСП() у Microsoft Excel.

У першому випадку частоту f для інтервалу (a, b) можна обчислити за формулою

$$f = \begin{cases} F_{\frac{\bar{X}-a}{s}} - F_{\frac{\bar{X}-b}{s}}, & \text{якщо } b < \bar{X}; \\ F_{\frac{\bar{X}-a}{s}} + F_{\frac{b-\bar{X}}{s}}, & \text{якщо } a < \bar{X} < b; \\ F_{\frac{b-\bar{X}}{s}} - F_{\frac{a-\bar{X}}{s}}, & \text{якщо } \bar{X} < a, \end{cases}$$

вважаючи, що $\Phi(-x) = -\Phi(x)$ і $\Phi(x) \approx 0,5$ при $x > 5$.

¹ Див., наприклад, [18] або [25].

Скористаємося другим підходом і отримаємо відносні інтервальні частоти f для теоретичних розподілів.

| Інтервал | | $f_{\text{теор}}, 2000 \text{ р.}$ $\mu = 89,5; \sigma = 40,98$ | $f_{\text{теор}}, 2006 \text{ р.}$ $\mu = 94,325; \sigma = 42,13$ |
|---------------|----------------|--|--|
| Нижня межа | Верхня межа | | |
| –∞ | 15 | 0,03 | 0,03 |
| 15 | 25 | 0,02 | 0,02 |
| 25 | 35 | 0,03 | 0,03 |
| 35 | 45 | 0,05 | 0,04 |
| 45 | 55 | 0,06 | 0,05 |
| 55 | 65 | 0,08 | 0,07 |
| 65 | 75 | 0,09 | 0,08 |
| 75 | 85 | 0,09 | 0,09 |
| 85 | 95 | 0,10 | 0,09 |
| 95 | 105 | 0,09 | 0,09 |
| 105 | 115 | 0,09 | 0,09 |
| 115 | 125 | 0,07 | 0,08 |
| 125 | 135 | 0,06 | 0,07 |
| 135 | 145 | 0,05 | 0,05 |
| 145 | 155 | 0,03 | 0,04 |
| 155 | 165 | 0,02 | 0,03 |
| 165 | 175 | 0,01 | 0,02 |
| 175 | 185 | 0,01 | 0,01 |
| 185 | 195 | 0,00 | 0,01 |
| 195 | ∞ | 0,01 | 0,01 |
| Разом | | 1,00 | 1,00 |

Сформулюємо гіпотези (вони однакові для 2000 р. та 2006 р.):

H_0 : Відмінності між емпіричним та теоретичним розподілами незначущі;

H_1 : Відмінності між емпіричним та теоретичним розподілами значущі.

Нагадаємо, що для знаходження емпіричного значення критерію Колмогорова – Смірнова попередньо потрібно обчислити максимальну за модулем різницю d_{max} між накопиченими частотами cf порівнюваних розподілів.

Отже (опустимо в таблиці стовпець інтервалів для X), для 2000 р. отримуємо такі дані.

| $f_{\text{эмп}}$ | $cf_{\text{эмп}}$ | $f_{\text{теор}}$ | $cf_{\text{теор}}$ | $cf_{\text{теор}} - cf_{\text{эмп}}$ |
|------------------|-------------------|-------------------|--------------------|--------------------------------------|
| 0,01 | 0,01 | 0,03 | 0,03 | 0,02 |
| 0,02 | 0,03 | 0,02 | 0,06 | 0,03 |
| 0,02 | 0,05 | 0,03 | 0,09 | 0,05 |
| 0,06 | 0,11 | 0,05 | 0,14 | 0,03 |
| 0,06 | 0,16 | 0,06 | 0,20 | 0,04 |
| 0,12 | 0,28 | 0,08 | 0,27 | -0,01 |
| 0,17 | 0,45 | 0,09 | 0,36 | -0,09 |
| 0,12 | 0,57 | 0,09 | 0,46 | -0,11 |
| 0,09 | 0,66 | 0,10 | 0,55 | -0,10 |
| 0,06 | 0,72 | 0,09 | 0,65 | -0,07 |
| 0,05 | 0,76 | 0,09 | 0,73 | -0,03 |
| 0,03 | 0,79 | 0,07 | 0,81 | 0,02 |
| 0,06 | 0,85 | 0,06 | 0,87 | 0,01 |
| 0,03 | 0,89 | 0,05 | 0,91 | 0,03 |
| 0,02 | 0,90 | 0,03 | 0,95 | 0,04 |
| 0,04 | 0,94 | 0,02 | 0,97 | 0,03 |
| 0,01 | 0,95 | 0,01 | 0,98 | 0,03 |
| 0,03 | 0,98 | 0,01 | 0,99 | 0,02 |
| 0,01 | 0,99 | 0,00 | 0,99 | 0,01 |
| 0,02 | 1,00 | 0,01 | 1,00 | 0,00 |

Для 2006 р.

| $f_{\text{эмп}}$ | $cf_{\text{эмп}}$ | $f_{\text{теор}}$ | $cf_{\text{теор}}$ | $cf_{\text{теор}} - cf_{\text{эмп}}$ |
|------------------|-------------------|-------------------|--------------------|--------------------------------------|
| 1 | 2 | 3 | 4 | 5 |
| 0,00 | 0,00 | 0,03 | 0,03 | 0,03 |
| 0,01 | 0,01 | 0,02 | 0,05 | 0,04 |
| 0,01 | 0,02 | 0,03 | 0,08 | 0,06 |
| 0,04 | 0,06 | 0,04 | 0,12 | 0,07 |
| 0,09 | 0,14 | 0,05 | 0,18 | 0,03 |
| 0,14 | 0,28 | 0,07 | 0,24 | -0,04 |
| 0,10 | 0,38 | 0,08 | 0,32 | -0,06 |
| 0,14 | 0,53 | 0,09 | 0,41 | -0,11 |
| 0,11 | 0,63 | 0,09 | 0,51 | -0,12 |
| 0,07 | 0,70 | 0,09 | 0,60 | -0,09 |
| 0,04 | 0,74 | 0,09 | 0,69 | -0,05 |
| 0,05 | 0,79 | 0,08 | 0,77 | -0,02 |
| 0,03 | 0,82 | 0,07 | 0,83 | 0,02 |

| 1 | 2 | 3 | 4 | 5 |
|------|------|------|------|------|
| 0,04 | 0,85 | 0,05 | 0,89 | 0,03 |
| 0,03 | 0,88 | 0,04 | 0,93 | 0,05 |
| 0,02 | 0,90 | 0,03 | 0,95 | 0,05 |
| 0,03 | 0,93 | 0,02 | 0,97 | 0,04 |
| 0,02 | 0,95 | 0,01 | 0,98 | 0,03 |
| 0,03 | 0,98 | 0,01 | 0,99 | 0,01 |
| 0,02 | 1,00 | 0,01 | 1,00 | 0,00 |

Отже, $d_{\max} = 0,11$ для 2000 р. і $d_{\max} = 0,12$ для 2006 р. За цими даними знаходимо емпіричні значення критерію Колмогорова – Смірнова, використовуючи формулу

Маємо

$$\lambda_{\text{емп}} = d_{\max} \sqrt{\frac{n}{2}}.$$

$$\lambda_{2000} = 0,11 \times \sqrt{200} \approx 1,56;$$

$$\lambda_{2006} = 0,12 \times \sqrt{200} \approx 1,70.$$

Рівень значущості, який відповідає цим значенням, шукаємо в таблицях критерію Колмогорова – Смірнова (див., наприклад, [18] або [25]).

Для 2000 р. критерій Колмогорова – Смірнова дорівнює 1,5 %, для 2006 р. – 0,6 %. В обох випадках рівень значущості менший від 5 %, і тому приймаємо гіпотезу H_1 про відмінність емпіричного та теоретичного нормального розподілів.

Щодо двох завдань за узгодженості теоретичного нормального розподілу з емпіричним зауважимо, що найкращими точковими оцінками параметрів теоретичного розподілу є вибіркове середнє та виправлене стандартне відхилення, а задача знаходження довірчого інтервалу для стандартного відхилення детально розглянута у [18; 25].

СПИСОК ЛІТЕРАТУРИ

Основна

1. *Захаров В. П.* Применение математических методов в социально-психологических исследованиях. — Л.: Изд-во ЛГУ, 1985.
2. *Кендалл М. Дж., Стюарт А.* Статистические алгоритмы в социологических исследованиях. — Новосибирск: Наука, 1985.

3. Максименко В. С., Паниотто В. І., Харченко Н. М. Статистичний аналіз соціологічних даних. — К.: Академія, 2004.
4. Паниотто В. І., Максименко В. С. Количественные методы в социологии. — К.: Наук. думка, 1982.
5. Телейко А. Б., Чорней Р. К. Математико-статистичні методи в соціології та психології. — К.: МАУП, 2007.
6. Толстова Ю. Н. Анализ социологических данных. — М.: Науч. мир, 2000.
7. Фелингер А. Ф. Статистические алгоритмы в социологических исследованиях. — Новосибирск: Наука, 1985.
8. *Практикум з теорії ймовірності та математичної статистики*: Навч. посіб. / Р. К. Чорней, О. Ю. Дюженкова, О. Б. Жильцов та ін.; За ред. Р. К. Чорнея. — К.: МАУП, 2003.
9. Эренберг А. Анализ и интерпретация статистических данных. — М.: Финансы и статистика, 1981.

Додаткова

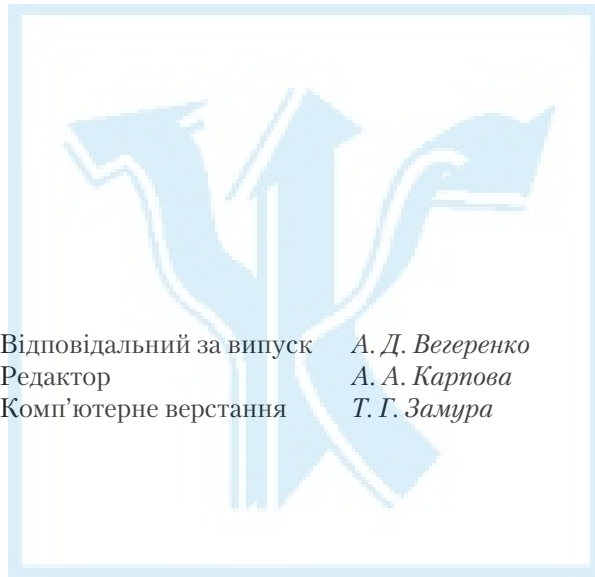
10. Бешелев С. Д., Гурвич Ф. Г. Математико-статистические методы экспертных оценок. — М.: Статистика, 1980.
11. Вайнберг Дж., Шуменер Дж. Статистика: Пер. с англ. — М.: Статистика, 1979.
12. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии: Пер. с англ. — М.: Прогресс, 1976.
13. Громыко Г. Л. Статистика. — М.: Изд-во МГУ, 1981.
14. Кимбл Г. Как правильно пользоваться статистикой: Пер. с англ. — М.: Финансы и статистика, 1982.
15. Лбов Г. С. Методы обработки разнотипных экспериментальных данных. — Новосибирск: Наука, 1981.
16. *Математическая статистика*: Учебник. — М.: Высш. шк., 1981.
17. Мюллер П., Нойман, Шторм Р. Таблицы по математической статистике: Пер. с нем. — М.: Финансы и статистика, 1982.
18. Окунь Я. Факторный анализ: Пер. с польск. — М.: Статистика, 1974.
19. Паповян С. С. Математические методы в социальной психологии. — М.: Наука, 1983.
20. Рунион Р. Справочник по непараметрической статистике. — М.: Финансы и статистика, 1982.
21. Сидоренко Е. В. Методы математической обработки в психологии. — СПб.: Речь, 2001.

22. *Суходольский Г. В.* Основы математической статистики для психологов. — Л.: Изд-во ЛГУ, 1972.
23. *Турчин В. М.* Математична статистика: Навч. посіб. — К.: Академія, 1999.
24. *Феллер В.* Введение в теорию вероятностей и ее приложения: В 2 т. — М.: Мир, 1964, 1967.
25. *Хастинг Н., Пикон Дж.* Справочник по статистическим распределениям: Пер. с англ. — М.: Статистика, 1980.
26. *Холлендер М., Вулф Д. А.* Непараметрические методы статистики: Пер. с англ. — М.: Финансы и статистика, 1983.



ЗМІСТ

| | |
|--|----|
| Пояснювальна записка | 3 |
| Теоретичні питання для контрольної роботи..... | 3 |
| Задачі для контрольної роботи..... | 4 |
| Приклади розв'язання задач..... | 13 |
| Список літератури..... | 39 |



| | |
|--------------------------|------------------------|
| Відповідальний за випуск | <i>А. Д. Вегеренко</i> |
| Редактор | <i>А. А. Карпова</i> |
| Комп'ютерне верстання | <i>Т. Г. Замура</i> |

МАУП

Зам. № ВКЦ-3180

Міжрегіональна Академія управління персоналом (МАУП)
03039 Київ-39, вул. Фрометівська, 2, МАУП