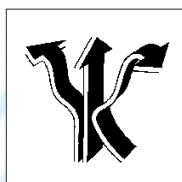


**МІЖРЕГІОНАЛЬНА АКАДЕМІЯ
УПРАВЛІННЯ ПЕРСОНАЛОМ**



МАУП

**МЕТОДИЧНІ РЕКОМЕНДАЦІЇ
ЩОДО ЗАБЕСПЕЧЕННЯ САМОСТІЙНОЇ РОБОТИ СТУДЕНТІВ
З ДИСЦИПЛІНИ
„Теорія причинності в соціології”
(для бакалаврів)**

МАУП

КИЇВ — 2016

Підготовлено професором кафедри вищої та прикладної математики
Степахно І.В.

Затверджено на засіданні кафедри вищої та прикладної математики
(протокол № 10 від 20.05. 2016 р.)

Схвалено Вченою радою Інституту комп'ютерно-інформаційних тех-
нологій Міжрегіональної Академії управління персоналом
(протокол № _____ від _____ 2016 р.)

Степахно І.В. Методичні рекомендації щодо забезпечення самостійної
роботи студентів з дисципліни „Теорія причинності в соціології”. — К.:
МАУП, 2016. — 37 с.

Методичні вказівки містять пояснювальну записку; питання з декіль-
кох розділів дисципліни „ Теорія причинності в соціології ”, які потребують
додаткових пояснень; приклади роз'яснювальних завдань та умови задач,
список літератури.

ПОЯСНЮВАЛЬНА ЗАПИСКА

Освоєння навчального курсу „Теорія причинності в соціології ” надає можливість підготувати фахівців вищої кваліфікації — соціологів, соціальних працівників, які змогли б організувати проведення кількісного соціологічного дослідження та здійснити математичний аналіз такого дослідження. Одним з видів такого дослідження є опитування членів суспільства з метою отримання достовірної незалежної інформації про ті чи інші чинники суспільного життя.

При виконанні завдань контрольної роботи студент опановує основні етапи та методи вибіркового дослідження під час аналізу соціальних явищ і показників, а також методи статистичного аналізу отриманої під час вибіркового дослідження інформації.

Завдання для контрольної роботи розраховані на студентів всіх форм навчання, як засіб перевірки знань при проведенні підсумкової атестації. Студенти заочної форми навчання використовують ці завдання як контрольну роботу згідно навчального плану.

Контрольна робота складається з двох теоретичних питань і двох задач. Номери теоретичних питань студент обирає 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 студент обирає за останньою цифрою номера своєї залікової книжки.

Контрольну роботу виконують у зошиті або на аркушах формату А4 з полями для зауважень викладача. При виконанні кожного завдання необхідно вказати його номер і переписати умову. Розв'язок завдання обов'язково потрібно пояснювати. У розрахунках слід використовувати правила наближених обчислень.

ТЕОРЕТИЧНІ ПИТАННЯ ДЛЯ КОНТРОЛЬНИХ РОБІТ

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, незміщенність, ефективність, BLUE-оцінки, змістовність. Найуживаніші точкові оцінки.
12. Інтервальні оцінки генерального середнього при відомому та невідомому стандартному відхиленні, 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.
21. Причинність у соціології та його походження.
22. Вплив причинності на якість соціологічних досліджень.
23. Математичні корені причинності.

ЗАДАЧІ ДЛЯ КОНТРОЛЬНИХ РОБІТ

Задача 1

В 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 свої історичні, релігійні та культурні традиції, соціологічна служба вирішила розбити своє дослідження на два: опитування на Заході та опитування на Сході. Обсяг опитаної випадкової вибірки в західно-

му регіоні становив 2500 респондентів, а в східному --- 2300. Кожний респондент вказував партію, прихильником якої він був у 2000 та 2004 році. Дані дослідження наведено у таблиці.

Варіант 1

Політична партія	2000 рік		2006 рік	
	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	2000 рік		2006 рік	
	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	2000 рік		2006 рік	
	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	2000 рік		2006 рік	
	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	2000 рік		2006 рік	
	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	2000 рік		2006 рік	
	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	2000 рік		2006 рік	
	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	2000 рік		2006 рік	
	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	2000 рік		2006 рік	
	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	2000 рік		2006 рік	
	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 і Сході 46% до 54%, для кожної партії знайдіть точкову оцінку генеральної пропорції її прихильників:

- у 2000 році окремо у кожному регіоні;

- у 2000 році разом по країні;
 - у 2006 році окремо у кожному регіоні;
 - у 2006 році разом по країні.
2. Для кожної партії визначте 95% довірчий інтервал для генеральної пропорції її прихильників:
- у 2000 році окремо у кожному регіоні;
 - у 2000 році разом по країні;
 - у 2006 році окремо у кожному регіоні;
 - у 2006 році разом по країні.
3. Для кожної партії визначте 95%-довірчий інтервал для різниці генеральних пропорцій її прихильників на Заході і Сході:
- у 2000 році;
 - у 2006 році.
4. За допомогою критерію χ^2 Пірсона перевірте на рівні значущості 5%, чи узгоджений з рівномірним розподілом розподіл кількості прихильників між політичними партіями:
- на Заході у 2006 році;
 - на Сході у 2006 році;
 - разом по країні у 2006 році.

Задача 2

Незалежна соціологічна служба провела опитування 400 випадкових респондентів працездатного віку у центральному регіоні країни про рівень їх щомісячних прибутків X у 2009 та 2015 роках (з точністю до 10 ум. од.). Дані опитування наведені в таблиці.

Варіант 1

Рівень щомісячних прибутків (ум. од.)	Кількість опитаних з даним рівнем прибутків		Рівень щомісячних прибутків (ум. од.)	Кількість опитаних з даним рівнем прибутків	
	2009 р.	2015 р.		2009 р.	2015 р.
010	04	00	110	14	21
020	05	04	120	23	18
030	08	08	130	14	12
040	19	17	140	10	25
050	25	23	150	18	13
060	56	39	160	10	07
070	52	40	170	10	16
080	51	63	180	11	06
090	25	46	190	08	09
100	33	25	200	04	08

Варіант 2

Рівень щомісячних прибутків (ум. од.)	Кількість опитаних з даним рівнем прибутків		Рівень щомісячних прибутків (ум. од.)	Кількість опитаних з даним рівнем прибутків	
	2009 р.	2015 р.		2009 р.	2015 р.
010	06	00	110	14	17
020	04	04	120	09	19
030	10	02	130	21	13
040	22	16	140	10	14
050	41	35	150	12	10
060	63	56	160	03	09
070	51	40	170	04	12
080	51	57	180	15	08
090	32	42	190	06	12
100	21	26	200	05	08

Варіант 3

Рівень щомісячних прибутків (ум. од.)	Кількість опитаних з даним рівнем прибутків		Рівень щомісячних прибутків (ум. од.)	Кількість опитаних з даним рівнем прибутків	
	2009 р.	2015 р.		2009 р.	2015 р.
010	04	00	110	18	25
020	05	01	120	11	16
030	11	04	130	19	16
040	19	08	140	09	09
050	18	22	150	08	11
060	72	50	160	07	13
070	57	56	170	12	21
080	53	57	180	09	13
090	39	44	190	05	03
100	22	26	200	02	05

Варіант 4

Рівень щомісячних прибутків (ум. од.)	Кількість опитаних з даним рівнем прибутків		Рівень щомісячних прибутків (ум. од.)	Кількість опитаних з даним рівнем прибутків	
	2009 р.	2015 р.		2009 р.	2015 р.
010	05	00	110	15	13
020	07	04	120	12	24
030	12	06	130	15	14
040	19	19	140	10	20
050	35	17	150	13	12
060	57	39	160	07	25
070	55	67	170	08	14
080	45	46	180	09	09
090	44	36	190	08	07
100	19	15	200	05	13

Варіант 5

Рівень щомісячних прибутків (ум. од.)	Кількість опитаних з даним рівнем прибутків		Рівень щомісячних прибутків (ум. од.)	Кількість опитаних з даним рівнем прибутків	
	2009 р.	2015 р.		2009 р.	2015 р.
010	03	00	110	14	19
020	05	03	120	16	20
030	07	02	130	12	15
040	19	16	140	11	12
050	31	27	150	13	16
060	54	45	160	11	08
070	47	56	170	07	14
080	60	50	180	08	06
090	38	50	190	04	13
100	33	18	200	07	10

Варіант 6

Рівень щомісячних прибутків (ум. од.)	Кількість опитаних з даним рівнем прибутків		Рівень щомісячних прибутків (ум. од.)	Кількість опитаних з даним рівнем прибутків	
	2009 р.	2015 р.		2009 р.	2015 р.
010	04	00	110	14	24
020	04	08	120	19	18
030	08	02	130	23	15
040	20	14	140	14	19
050	25	19	150	09	10
060	50	34	160	04	11
070	64	64	170	10	16
080	45	53	180	09	08
090	41	44	190	10	05
100	24	27	200	03	09

Варіант 7

Рівень щомісячних прибутків (ум. од.)	Кількість опитаних з даним рівнем прибутків		Рівень щомісячних прибутків (ум. од.)	Кількість опитаних з даним рівнем прибутків	
	2009 р.	2015 р.		2009 р.	2015 р.
010	05	00	110	15	24
020	05	07	120	13	16
030	04	07	130	24	14
040	24	15	140	12	21
050	37	23	150	06	10
060	61	37	160	07	14
070	66	55	170	03	10
080	42	65	180	05	09
090	46	35	190	05	05
100	19	23	200	01	10

Варіант 8

Рівень щомісячних прибутків (ум. од.)	Кількість опитаних з даним рівнем прибутків		Рівень щомісячних прибутків (ум. од.)	Кількість опитаних з даним рівнем прибутків	
	2009 р.	2015 р.		2009 р.	2015 р.
010	03	00	110	22	24
020	06	03	120	12	09
030	07	03	130	23	19
040	21	17	140	08	21
050	29	28	150	11	17
060	54	57	160	04	11
070	65	46	170	14	11
080	49	61	180	05	07
090	27	39	190	07	06
100	29	16	200	04	05

Варіант 9

Рівень щомісячних прибутків (ум. од.)	Кількість опитаних з даним рівнем прибутків		Рівень щомісячних прибутків (ум. од.)	Кількість опитаних з даним рівнем прибутків	
	2009 р.	2015 р.		2009 р.	2015 р.
010	05	00	110	22	23
020	03	03	120	09	11
030	07	04	130	15	17
040	27	12	140	21	15
050	24	21	150	10	21
060	58	39	160	07	10
070	66	57	170	05	13
080	43	57	180	08	11
090	44	43	190	11	12
100	13	26	200	02	05

Варіант 10

Рівень щомісячних прибутків (ум. од.)	Кількість опитаних з даним рівнем прибутків		Рівень щомісячних прибутків (ум. од.)	Кількість опитаних з даним рівнем прибутків	
	2009 р.	2015 р.		2009 р.	2015 р.
010	01	00	110	15	26
020	04	05	120	16	20
030	09	04	130	16	13
040	16	15	140	14	14
050	32	26	150	12	18
060	50	43	160	14	13
070	77	50	170	08	10
080	39	57	180	04	08
090	33	44	190	10	08
100	23	20	200	07	06

1. Проведіть статистичний аналіз отриманих даних окремо для 2009 та 2015 років. Для цього знайдіть:

- вибірккові середнє, моду та медіану рівня щомісячних прибутків;
 - вибірккові дисперсію, стандартне відхилення, середнє абсолютне відхилення, виправлене стандартне відхилення, коефіцієнт варіації та розмах варіації рівня щомісячних прибутків;
 - побудуйте полігон, гістограму та кумуляту частот вибіркового розподілу рівня щомісячних прибутків.
2. Окремо для 2009 та 2015 років за допомогою критерію χ^2 Пірсона з рівнем значущості 5% перевірте, чи вибіркковий розподіл рівня щомісячних прибутків є узгодженим з пуассонівським розподілом. У випадку узгодженості знайдіть точкову оцінку параметра λ розподілу Пуассона.
 3. Окремо для 2009 та 2015 років знайдіть 95% довірчий інтервал для середнього значення рівня щомісячних прибутків.
 4. Окремо для 2009 та 2015 років за допомогою критерію λ Колмогорова — Смірнова на рівні значущості 5% перевірте, чи вибіркковий розподіл рівня щомісячних прибутків є узгодженим з нормальним розподілом. У випадку узгодженості:
 - знайдіть точкові оцінки параметрів (μ, σ) нормального розподілу;
 - знайдіть 95% довірчий інтервал для стандартного відхилення рівня щомісячних прибутків.
 5. За допомогою кутового перетворення Фішера в поєднанні з процедурою підрахунку критерію λ Колмогорова — Смірнова на рівні значущості 5% перевірте наявність позитивного зсуву в рівні щомісячних прибутків.

ЗРАЗКИ РОЗВ'ЯЗУВАННЯ ЗАДАЧ

Приклад 1

Література: [18, Розділи 2, 5, 6]; [25, Розділи 12, 13].

В 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 свої історичні, релігійні та культурні традиції, соціологічна служба вирішила розбити своє дослідження на два: опитування на Заході та опитування на Сході. Обсяг опитаної випадкової вибірки в західному регіоні становив 2500 респондентів, а в східному --- 2300. Кожний респондент вказував партію, прихильником якої він був у 2009 та 2015 році. Дані дослідження наведено у таблиці.

Політична партія	2009 рік		2015 рік	
	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

1. Враховуючи, що співвідношення між кількостями виборців на Заході і Сході 46% до 54%, для кожної партії знайдіть точкову оцінку генеральної пропорції її прихильників:
 - у 2009 році окремо у кожному регіоні;
 - у 2009 році разом по країні;
 - у 2015 році окремо у кожному регіоні;
 - у 2015 році разом по країні.
2. Для кожної партії визначте 95% довірчий інтервал для генеральної пропорції її прихильників:
 - у 2009 році окремо у кожному регіоні;
 - у 2009 році разом по країні;
 - у 2015 році окремо у кожному регіоні;
 - у 2015 році разом по країні.
3. Для кожної партії визначте 95%-довірчий інтервал для різниці генеральних пропорцій її прихильників на Заході і Сході:
 - у 2009 році;
 - у 2015 році.
4. За допомогою критерію χ^2 Пірсона перевірте на рівні значущості 5%, чи узгоджений з рівномірним розподілом розподіл кількості прихильників між політичними партіями:
 - на Заході у 2015 році;
 - на Сході у 2015 році;
 - разом по країні у 2015 році.

Розв'язок

1. Обчислимо точкові оцінки генеральних пропорцій прихильників партій окремо по регіонам у 2009 році як відношення кількості прихильників партій у вибірці до обсягу вибірки. Маємо:

Політична партія	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	$\frac{682}{2500} = 27,28\%$	$\frac{586}{2300} \approx 25,48\%$

При знаходженні точкових оцінок генеральних пропорцій прихильників партій у 2009 році в цілому по країні потрібно врахувати співвідношення 46% : 54% між кількістю виборців на заході і сході. Для цього використаємо формулу

$$P_{\text{загальне}} = 46\% \cdot P_{\text{захід}} + 54\% \cdot P_{\text{схід}}.$$

Маємо такі точкові оцінки генеральних пропорцій прихильників партій у 2009 році в цілому по країні:

Політична партія	Генеральна пропорція 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	$46\% \cdot 27,28\% + 54\% \cdot 27,83\% \approx 27,57\%$

Так само отримаємо точкові оцінки генеральних пропорцій прихильників партій окремо по регіонам у 2015 році:

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

Точкові оцінки генеральних пропорцій прихильників партій в цілому по країні у 2015 році будуть такими:

Політична партія	Генеральна пропорція 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 шукатимемо за формулою

$$P_{\text{точкове}} - \Delta \leq P_{\text{генеральне}} \leq P_{\text{точкове}} + \Delta,$$

де Δ обчислюємо за формулою

$$\Delta = 1,64 \cdot \sqrt{\frac{P_{\text{точкове}} (1 - P_{\text{точкове}})}{N}},$$

де константа 1,64 відповідає рівню довіри до інтервальної оцінки 95%, а N — обсяг відповідної вибірки.

При цьому у випадках знаходження оцінки в цілому по країні як N беремо не сумарний обсяг вибірок 4800, а число, яке враховує співвідношення виборців на заході і сході:

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

Обчислимо інтервальні оцінки генеральних пропорцій прихильників партій для 2009 року у західному регіоні. Для цього знайдемо спочатку Δ .

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

Остаточно маємо такі інтервальні оцінки генеральних пропорцій прихильників партій для 2009 року у західному регіоні:

Політична партія	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 2009 році матимемо такі

Δ:

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

та інтервальні оцінки генеральних пропорцій прихильників партій:

Політична партія	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 в цілому у 2009 році маємо такі Δ

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

та інтервальні оцінки генеральних пропорцій прихильників партій

Політична партія	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 році знаходяться аналогічно, тому наведемо лише остаточні результати.

Маємо такі 95% довірчі інтервали для генеральних пропорцій прихильників партій у 2015 році на заході:

Політична партія	Δ	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	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%

та в цілому по країні

Політична партія	Δ	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 і сході ми скористаємося формулами

$$\Delta P_{\text{точкове}} - \Delta \leq \Delta P_{\text{генеральне}} \leq \Delta P_{\text{точкове}} + \Delta,$$

де

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

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

де $N_{\text{схід}}$ та $N_{\text{захід}}$ — обсяги відповідно східної та західної вибірок.

Для 2000 року маємо такі значення $\Delta P_{\text{точкове}}$:

Політична партія	$\Delta P_{\text{точкове}}$
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%

Далі знайдемо Δ :

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

Остаточного маємо такі 95% довірчі інтервали для різниці генеральних пропорцій прихильників партій на заході і сході у 2009 році:

Політична партія	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 ведуться розрахунки для випадку 2006 року. Наведемо лише остаточні результати.

Отримаємо такі 95% довірчі інтервали для різниці генеральних пропорцій прихильників партій на заході і сході у 2015 році:

Політична партія	$\Delta P_{\text{точкове}}$	Δ	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. Нагадаємо, що емпіричне значення критерію χ^2 Пірсона обчислюється за формулою

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

де сума береться по всім можливим значенням ознаки, $n_{\text{емп}}$ — частота конкретного значення ознаки у вибірці, $n_{\text{теор}}$ — теоретична частота значення.

У випадку порівняння емпіричного розподілу з рівномірним маємо

$$n_{\text{теор}} = \frac{N}{m},$$

де N — обсяг вибірки, а m — кількість значень ознаки.

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

У нашому випадку кількість ступенів вільності дорівнює $4 - 1 = 3$, а рівень значущості за умовою дорівнює 5%. Тоді критичне значення критерію Пірсона для нашої задачі дорівнює 7,82.

Якщо емпіричне значення критерію буде більшим від критичного (в нашому випадку більшим за 7,82), то ми повинні прийняти гіпотезу про відмінність досліджуваного розподілу від рівномірного (на рівні значущості 5%).

Перш за все зауважимо, що

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

Процедуру обчислення емпіричного значення критерію χ^2 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
$\chi_{\text{емп}}^2$				31,56

Оскільки емпіричне значення критерію більше, ніж критичне, приймаємо гіпотезу про відмінність розподілу числа прихильників різних політичних партій на заході у 2015 році від рівномірного.

Для сходу маємо:

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

Підрахуємо емпіричне значення критерію χ^2 :

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
$\chi_{\text{емп}}^2$				39,10

¹ Див., наприклад [18] або [25].

І в цьому разі (2015 рік, схід) розподіл кількості прихильників різних політичних партій відмінний від рівномірного.

Для країни вцілому ми повинні врахувати співвідношення виборців на сході і заході. Зокрема (див. п. 2), обсяг вибірки вважатимемо рівним 4260. Тоді для знаходження емпіричних частот нам знадобиться формула

$$n_{\text{емп}} = P_{\text{точкове}} \cdot 4260.$$

де $P_{\text{точкове}}$ — точкова оцінка пропорції прихильників партії вцілому по країні у 2006 році (див. п. 2).

Крім того,

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

Підрахуємо емпіричне значення критерію χ^2 :

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
$\chi_{\text{емп}}^2$				62,40

Остаточню робимо висновок, що і по всій країні у 2015 році розподіл числа прихильників різних політичних партій відмінний від рівномірного.

Приклад 2

Література: [18, Розділи 6, 7]; [25, Розділ 15].

Незалежна соціологічна служба провела опитування 400 випадкових респондентів працездатного віку у центральному регіоні країни про рівень їх щомісячних прибутків X у 2009 та 2015 роках (з точністю до 10 ум. од.). Дані опитування наведені в таблиці.

МАУП

Рівень щомісячних прибутків (ум. од.)	Кількість опитаних з даним рівнем прибутків		Рівень щомісячних прибутків (ум. од.)	Кількість опитаних з даним рівнем прибутків	
	2009 р.	2015 р.		2009 р.	2015 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

- Проведіть статистичний аналіз отриманих даних окремо для 2009 та 2015 років. Для цього знайдіть:
 - вибірккові середнє, моду та медіану рівня щомісячних прибутків;
 - вибірккові дисперсію, стандартне відхилення, середнє абсолютне відхилення, виправлене стандартне відхилення, коефіцієнт варіації та розмах варіації рівня щомісячних прибутків;
 - побудуйте полігон, гістограму та кумуляту частот вибіркового розподілу рівня щомісячних прибутків.
- Окремо для 2009 та 2015 років за допомогою критерію χ^2 Пірсона з рівнем значущості 5% перевірте, чи вибіркковий розподіл рівня щомісячних прибутків є узгодженим з пуассонівським розподілом. У випадку узгодженості знайдіть точкову оцінку параметра λ розподілу Пуассона.
- Окремо для 2009 та 2015 років знайдіть 95% довірчий інтервал для середнього значення рівня щомісячних прибутків.
- Окремо для 2009 та 2015 років за допомогою критерію λ Колмогорова — Смірнова на рівні значущості 5% перевірте, чи вибіркковий розподіл рівня щомісячних прибутків є узгодженим з нормальним розподілом. У випадку узгодженості:
 - знайдіть точкові оцінки параметрів (μ, σ) нормального розподілу;
 - знайдіть 95% довірчий інтервал для стандартного відхилення рівня щомісячних прибутків.
- За допомогою кутового перетворення Фішера в поєднанні з процедурою підрахунку критерію λ Колмогорова — Смірнова на рівні значущості 5% перевірте наявність позитивного зсуву в рівні щомісячних прибутків.

Розв'язок

1. Нагадаємо, що вибірккове середнє — це звичайне середнє арифметичне значення, мода — значення, яке зустрічається найчастіше, а медіана — це число, яке в упорядкованому списку індивідуальних значень розташоване

посередині.

Обчислимо середнє, моду та медіану рівня щомісячних прибутків у 2009 році за допомогою таблиці.

Рівень прибутків 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 найчастіше, а саме 67 разів, зустрічається значення 70, то мода дорівнює 70.

Для того, щоб знайти медіану, потрібно подивитись, які значення стоять в середині списку, тобто на 200-ій та 201-ій позиціях. З цією метою ми обчислили стовпчик накопичених частот. Шукаємо рівень прибутків, для якого накопичена частота вперше перетинає число 200. Це — рівень у 80 ум. од.

Отже, медіана дорівнює 80.

Так само обчислюємо середнє, моду, медіану для випадку 2006 року. Маємо:

Рівень прибутків 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

Рівень прибутків X	Частота n	Добуток Xn	Накопичена частота cn
60	56	3360	$57 + 56 = 113$
70	40	2800	$113 + 40 = 153$
80	57	4560	$153 + 57 = \mathbf{210}$
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 випадку співпадають 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 випадок 2000 року. Проведемо попередні обчислення у таблиці.

X	n	Відхилення $x = X - \bar{X}$	$ x \cdot n$	x^2	$x^2 \cdot n$
10	4	-79,5	318,5	6320,25	25281,00
20	8	-69,5	556,5	4830,25	38642,00
30	6	-59,5	357,5	3540,25	21241,50
40	24	-49,5	1188,5	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

X	n	Відхилення $x = X - \bar{X}$	$ x \cdot n$	x^2	$x^2 \cdot n$
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				$\sqrt{1675,25} \approx 40,93$	

Обчислимо тепер виправлене стандартне відхилення:

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

та коефіцієнт варіації

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

Найбільший рівень прибутків у вибірці 200, найменший — 10, тому розмах варіації

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

Розглянемо тепер випадок 2015 року.

X	n	Відхилення $x = X - \bar{X}$	$ x \cdot n$	x^2	$x^2 \cdot n$
10	0	-84,325	0,000	7110,71	0,00
20	4	-74,325	297,300	5524,21	22096,82
30	2	-64,325	128,650	4137,71	8275,41
40	16	-54,325	869,200	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

X	n	Відхилення $x = X - \bar{X}$	$ x \cdot n$	x^2	$x^2 \cdot n$
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				$\sqrt{1770,54} \approx 42,08$	

Для виправленого стандартного відхилення, коефіцієнта варіації та розмаху варіації маємо:

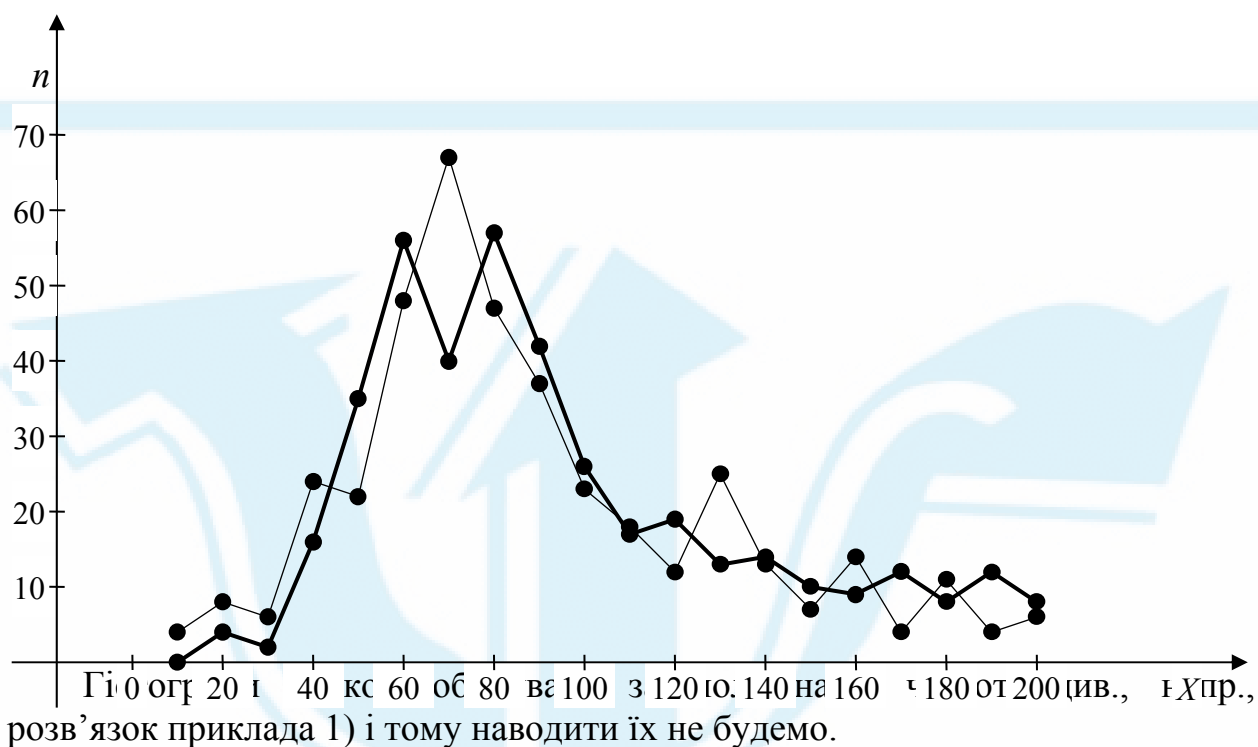
$$s = 42,08 \cdot \sqrt{\frac{400}{399}} \approx 42,13;$$

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

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

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

Наведемо на рисунку полігони частот розподілу рівня щомісячних прибутків у 2000 (тонка лінія) та 2006 (товста лінія) роках.



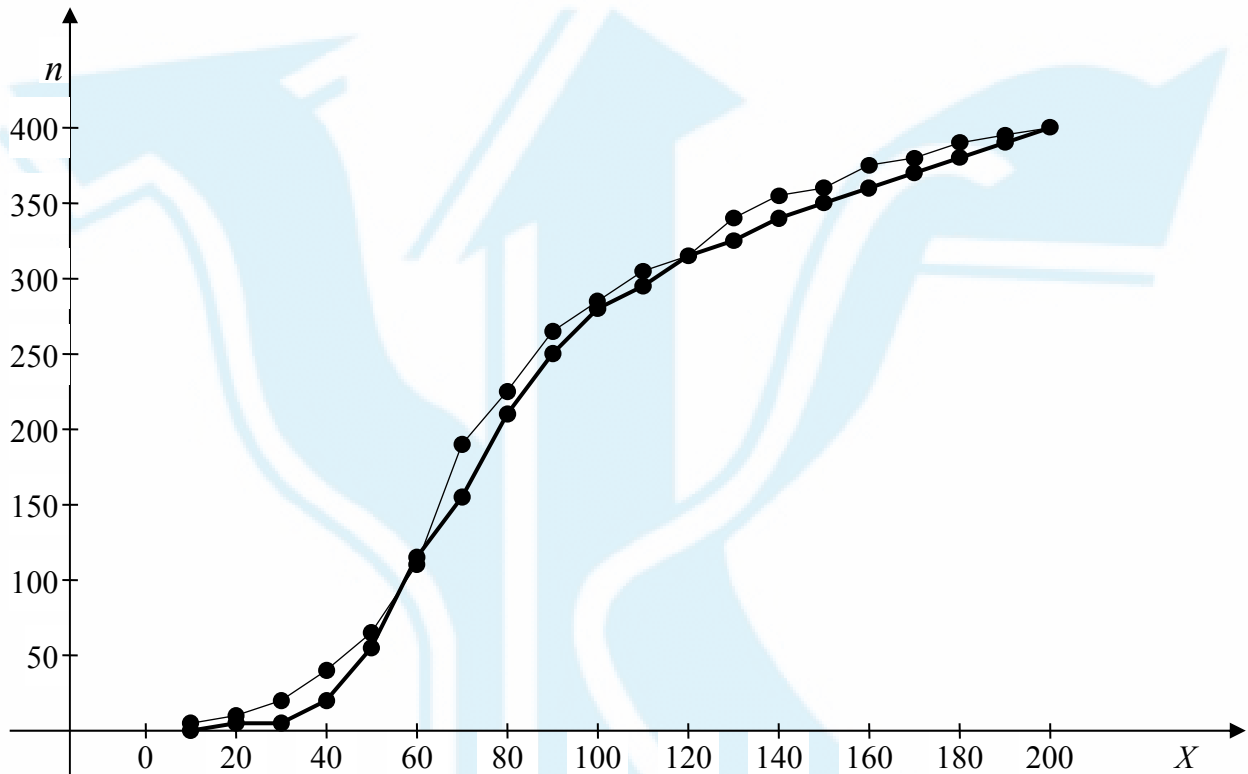
розв'язок приклада 1) і тому наводити їх не будемо.

Для побудови кумулят частот потрібно обчислити попередньо накопичені частоти. Маємо:

Рівень, X	2009 рік		2015 рік	
	Частота n	Накопичена частота cn	Частота n	Накопичена частота cn
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
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

Рівень, X	2009 рік		2015 рік	
	Частота n	Накопичена частота cn	Частота n	Накопичена частота cn
200	6	$394 + 6 = 400$	8	$392 + 8 = 400$

На рисунку зображено кумуляти частот для 2009 (тонка лінія) та 2015 (товста лінія) років.



2. Аналізуючи полігони частот розподілів рівня прибутків X , бачимо, що у вибірці переважають малі значення ознаки X , а частота швидко зменшується з ростом X . Подібними властивостями характеризується розподіл Пуассона¹.

Проте відомо, що середнє значення (точніше, математичне сподівання) та дисперсія розподілу Пуассона однакові. У нашому ж випадку середнє та дисперсія відрізняються на порядок. Так наприклад, у 2009 році середнє було 89,5, а дисперсія — аж 1675,25. Тому говорити про співпадіння досліджуваних розподілів з деякими пуассонівськими не можна.

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

Обчислити число α можна, використовуючи рівняння

¹ Див., напр., [18].

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

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

Розв'язуючи його, маємо

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

Для 2009 року маємо

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

а для 2015 —

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

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

При $\alpha = 20$ ми отримаємо для 2009 року

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

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

та для 2015 року —

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

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

Бачимо, що після зменшення рівня X у 20 разів середнє та дисперсія наших даних мало відрізняються.

Після масштабування рівень прибутків X , який змінювався у вибірці від 10 до 200 з кроком 10, буде змінюватись від 0,5 до 10 з кроком 0,5. Але, як ми вже згадували, розподіл Пуассона — це розподіл цілочисельної величини. Тому нам потрібно мати розподіл ознаки X , яка може приймати значення 0, 1, 2, і т.д.

Для того щоб побудувати такий розподіл за нашими даними, вважатимемо, що для серед тих людей, які мають дане значення ознаки X , половина,

¹ Якби наші дані стосувалися України, а під ум. од. ми розуміли, напр., долари США, то, враховуючи приблизне співвідношення між гривнею та доларом, як 5 до 1, ми б отримали опис рівня прибутків X у масштабі ста гривень.

насправді, має дещо менше значення, а половина — дещо більше. Тоді для рівня t в шуканому розподілі його частоту обчислимо, як суму всіх частот, що попадають в інтервал $(t - 0,5; t + 0,5)$.

Так, для 2009 року отримаємо такий розподіл (він є моделлю вихідного і вже може порівнюватися з пуассонівським).

Рівень 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	$\frac{18}{2} + 12 + \frac{25}{2} = 33,5$
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 рівня 5 в цій таблиці отримано, наприклад, як суму половини частоти для рівня 4,5 (вихідного 90), частоти для рівня 5 (вихідного 100) та половини частоти для рівня 5,5 (вихідного 110).

Діючи так само, для 2015 року отримаємо такий розподіл:

Рівень X	Частота n
0	$\frac{0}{2} = 0$
1	$\frac{0}{2} + 4 + \frac{2}{2} = 5$
2	$\frac{2}{2} + 16 + \frac{35}{2} = 34,5$

Рівень X	Частота n
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	$\frac{10}{2} + 9 + \frac{12}{2} = 20$
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 2009 року $\lambda = 4,475$, а для 2015 — $\lambda = 4,72$.

Для того, щоб побудувати шукані теоретичні розподіли, можна вибрати декілька шляхів. По-перше, можна скористатися формулою для підрахунку ймовірності прийняти значення m у випадку 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 дане значення.

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

Ми скористаємося другим підходом. Отримаємо такі теоретичні розподіли.

¹ Див., напр., [18] або [25].

Рівень X	2009 рік, $\lambda = 4,475$		2015 рік, $\lambda = 4,72$	
	Відносна частота wn	Частота, $wn \cdot 400$	Відносна частота wn	Частота, $wn \cdot 400$
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
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 (2009 та 2015 рік) 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 теоретичним. Нагадаємо (див. розв'язок прикладу 1), що емпіричне значення критерію шукається як сума виразів

$$\Delta_{\text{В}}^2 f = \frac{(n_{\text{емп}} - n_{\text{теор}})^2}{n_{\text{теор}}},$$

обчислених для всіх рівнів X . Тут $n_{\text{емп}}$ та $n_{\text{теор}}$ — частоти відповідно емпіричного та теоретичного розподілів.

Результати обчислень наведемо в таблицях. Маємо для 2009 року:

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	2,0	4,6	2,6	6,53	1,43
1	13,0	20,4	7,4	54,59	2,68
2	38,0	45,6	7,6	58,05	1,27
3	92,5	68,0	-24,5	597,86	8,79

Рівень X	$n_{\text{емп}}$	$n_{\text{теор}}$	$n_{\text{емп}} - n_{\text{теор}}$	$(n_{\text{емп}} - n_{\text{теор}})^2$	$\frac{(n_{\text{емп}} - n_{\text{теор}})^2}{n_{\text{теор}}}$
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
$\chi_{\text{емп}}^2$					36,21

та 2015 року —

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
$\chi_{\text{емп}}^2$					59,40

Бачимо, що в обох випадках емпіричні значення критерію Пірсона (36,21 та 59,40) більші за критичне 18,31. Отже, в обох випадках ми повинні прийняти гіпотезу H_1 про відмінність емпіричних розподілів від теоретичних пуассонівських¹.

¹ Зауважимо, що якби ми досліджували узгодженість даних розподілів за допомогою критерію Колмогорова — Смірнова (див., напр., п. 4), ми мали б відкинути гіпотезу H_1 на рівні значущості 22% для випадку 2009 року та на рівні значущості 11% для випадку 2015 року.

Іншими словами, в обох випадках ми б прийняли гіпотезу про подібність емпіричних розподілів з теоретичними пуассонівськими.

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

З іншого боку, критерій Пірсона, як правило, застосовують для номінальних змінних. У нашому випадку дані мають числову природу і використання критерію Колмогорова

3. Для знаходження довірчого інтервалу для генерального середнього μ рівня щомісячних прибутків скористаємося формулою

$$\bar{X} - \Delta \leq \mu \leq \bar{X} + \Delta,$$

де

$$\Delta = z \frac{s}{\sqrt{n}}.$$

В останній формулі значення z залежить від заданого рівня довіри і в нашому випадку (95%) приблизно дорівнює 1,64, s — це виправлене стандартне відхилення (див. п. 2 приклада 1), $n = 400$ — обсяг вибірки.

Маємо

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

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

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 нам попередньо потрібно знайти ці теоретичні розподіли.

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

На відміну від пуассонівського нормальний розподіл — це розподіл неперервної випадкової величини. Тому нас цікавитимуть не точкові частоти розподілів, а інтервальні. Зокрема, емпіричну частоту рівня 10 природно вважати частотою для інтервалу $(-\infty; 15)$, рівня 20 — частотою для інтервалу $(15; 25)$ і т.д.

Іншими словами, емпіричні розподіли у інтервальній формі матимуть вигляд (оскільки в критерії Колмогорова — Смірнова використовуються відносні частоти розподілів, відразу обчислимо їх як відношення частот до обсягу вибірки 400):

ва — Смірнова виглядає природнішим.

Інтервал		2009 рік		2015 рік	
Нижня межа	Верхня межа	Частота	Відносна частота	Частота	Відносна частота
$-\infty$	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	400	1

Відносні інтервальні частоти теоретичних нормальних розподілів можна обчислити, наприклад, використовуючи таблицю значень інтегральної функції Лапласа Φ ¹, або скориставшись інтегральною версією функції НОРМРАСП() в Microsoft Excel.

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

$$f = \begin{cases} \Phi\left(\frac{\bar{X} - a}{s}\right) - \Phi\left(\frac{\bar{X} - b}{s}\right), & \text{якщо } b < \bar{X}; \\ \Phi\left(\frac{\bar{X} - a}{s}\right) + \Phi\left(\frac{b - \bar{X}}{s}\right), & \text{якщо } a < \bar{X} < b; \\ \Phi\left(\frac{b - \bar{X}}{s}\right) - \Phi\left(\frac{a - \bar{X}}{s}\right), & \text{якщо } \bar{X} < a, \end{cases}$$

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

Ми скористалися другим підходом і отримали такі відносні інтервальні

¹ Див., напр., [18] або [25].

частоти f для теоретичних розподілів.

Інтервал		$f_{\text{теор}}, 2009 \text{ рік}$ $\mu = 89,5, \sigma = 40,98$	$f_{\text{теор}}, 2015 \text{ рік}$ $\mu = 94,325, \sigma = 42,13$
Нижня межа	Верхня межа		
$-\infty$	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	1

Сформулюємо гіпотези (вони однакові для 2009 та 2015 року):

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 знайти максимальну за модулем різницю d_{max} між накопиченими частотами cf порівнюваних розподілів.

Отже (опустимо в таблиці стовпчик інтервалів для X), для 2009-го року:

$f_{\text{емп}}$	$cf_{\text{емп}}$	$f_{\text{теор}}$	$cf_{\text{теор}}$	$cf_{\text{теор}} - cf_{\text{емп}}$
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

$f_{\text{емп}}$	$cf_{\text{емп}}$	$f_{\text{теор}}$	$cf_{\text{теор}}$	$cf_{\text{теор}} - cf_{\text{емп}}$
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 2015 року:

$f_{\text{емп}}$	$cf_{\text{емп}}$	$f_{\text{теор}}$	$cf_{\text{теор}}$	$cf_{\text{теор}} - cf_{\text{емп}}$
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
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} дорівнює 0,11 у 2009 і 0,12 у 2015 році. За цими даними знаходимо емпіричні значення критерію Колмогорова — Смірнова, використовуючи формулу

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

Маємо:

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

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

Рівень значущості, який відповідає цим значенням, шукаємо у таблицях критерію Колмогорова — Смірнова (див., напр., [18] або [25]).

Для 2009 року він дорівнює 1,5%, а для 2015— 0,6%. В обох випадках рівень значущості менший за 5%, і тому ми приймаємо гіпотезу H_1 про відмінність емпіричного та теоретичного нормального розподілів.

Щодо двох завдань для випадку 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 в [18] та [25].

МАУП

СПИСОК ЛІТЕРАТУРИ

Основна

1. Кендалл М. Дж., Стюарт А. Статистические алгоритмы в социологических исследованиях. — Новосибирск: Наука, 1985.
2. Кимбл Г. Как правильно пользоваться статистикой: пер. с англ. — М.: Финансы и статистика, 1982.
3. Максименко В. С., Паниотто В. И., Харченко Н. М. Статистичний аналіз соціологічних даних. — К.: Видавничий дім „КМ Академія”, 2004.
4. Окунь Я. Факторный анализ: пер. с польск. — М.: Статистика, 1974.
5. Паниотто В. И., Максименко В. С. Количественные методы в социологии. — К.: Наукова думка, 1982.
6. Телейко А. Б., Чорней Р. К. Математико-статистичні методи в соціології та психології. — К.: МАУП, 2007.
7. Толстова Ю. Н. Анализ социологических данных. — М.: Научный мир, 2000.
8. Чорней Р. К., Дюженкова О. Ю., Жильцов О. Б. та ін. Практикум з теорії ймовірності та математичної статистики: Навч. посіб. / За ред. Р. К. Чорнея. — К.: МАУП, 2003.

Додаткова

9. Бешелев С. Д., Гурвич Ф. Г. Математико-статистические методы экспертных оценок. — М.: Статистика, 1980.
10. Вайнберг Дж., Шуменер Дж. Статистика: пер. с англ. — М.: Статистика, 1979.
11. Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии: пер. с англ. — М.: Прогрес, 1976.
12. Громько Г. Л. Статистика. — М.: Изд-во МГУ, 1981.
13. Захаров В. П. Применение математических методов в социально-психологических исследованиях. — Л.: ЛГУ, 1985.
14. Лбов Г. С. Методы обработки разнотипных экспериментальных данных. — Новосибирск: Наука, 1981.
15. Математическая статистика: Учебник. — М.: Высш. шк., 1981.
16. Мюллер П., Нойман, Шторм Р. Таблицы по математической статистике: пер. с нем. — М.: Финансы и статистика, 1982.
17. Паповян С. С. Математические методы в социальной психологии. — М.: Наука, 1983.
18. Рунион Р. Справочник по непараметрической статистике. — М.: Финансы и статистика, 1982.
19. Сидоренко Е. В. Методы математической обработки в психологии. — С-Пб.: Речь, 2001.
20. Суходольский Г. В. Основы математической статистики для психологов. — Л.: ЛГУ, 1972.

21. Турчин В. М. Математична статистика. Навч. посіб. — К.: Видавничий центр „Академія”, 1999.
22. Фелингер А. Ф. Статистические алгоритмы в социологических исследованиях. — Новосибирск: Наука, 1985.
23. Феллер В. Введение в теорию вероятностей и ее приложения: В 2 т. — М.: Мир, 1964, 1967.
24. Хастинг Н., Пикон Дж. Справочник по статистическим распределениям: пер. с англ. — М.: Статистика, 1980.
25. Холлендер М., Вулф Д. А. Непараметрические методы статистики: пер. с англ. — М.: Финансы и статистика, 1983.
26. Эренберг А. Анализ и интерпретация статистических данных. — М.: Финансы и статистика, 1981.
27. Горбань С.Ф., Снижко Н.В. Теория вероятностей и математическая статистика.- К., МАУП.- 2008,168с.
28. Чорней В.Е. Практикум по теории вероятностей и математической статистике. –К.,МАУП.-2005, 265с.
29. Лещинский О.Л., Рязанцева В.В., Юнькова О.О. Економетрія – К.,МАУП.-2009, 208с.

МАУП



MAYUN