

НАЦІОНАЛЬНА МЕТАЛУРГІЙНА АКАДЕМІЯ УКРАЇНИ
КАФЕДРА ПРИКЛАДНОЇ МАТЕМАТИКИ ТА ОБЧИСЛЮВАЛЬНОЇ ТЕХНІКИ

Навчальний посібник

щодо вивчення дисципліни

***"Методи прикладного статистичного
аналізу "***

для студентів усіх спеціальностей

Затверджено на засіданні кафедри
Протокол № від . 2017 р.

Дніпропетровськ

2017

Навчальний посібник щодо вивчення дисципліни "Методи прикладного статистичного аналізу" / Укл.: Г.Г. Швачич, В.С. Коноваленков, О.В. Соболєнко, Т.М.Забоєова, В.І. Христ'ян, Є.Є. Єгорцева – Дніпропетровськ: НМетАУ, 2017. – 178 с.

Укладачі: Г.Г. Швачич, докт. техн. наук, проф.,
В.С. Коноваленков, канд. техн. наук, доц.,
О.В. Соболєнко, канд. техн. наук, доц.,
Т.М.Забоєова, ст.викл.,
В.І. Христ'ян, ст. викл.,
Є.Є. Єгорцева, асистент

Відповідальний за випуск: Г.Г. Швачич, докт. техн. наук, проф.,
зав. каф. ПМ та ОТ

Зміст

Вступ.....	4
Завдання № 1. Вибірковий метод	5
Завдання № 2. Первинна обробка статистичних даних дискретної випадкової ознаки	11
Завдання № 3. Первинна обробка статистичних даних безперервної випадкової ознаки	20
Завдання № 4. Статистичне оцінювання параметрів	33
Завдання № 5. Перевірка статистичних гіпотез. Критерій згоди χ^2 (Критерій Пірсона)	47
Завдання № 6. Перевірка статистичних гіпотез. Критерій згоди λ (Критерій Колмогорова)	79
Завдання № 7. Дисперсійний аналіз	89
Завдання № 8. Регресійний аналіз. Парна лінійна регресія. етап структуризації математичної моделі	135
Завдання № 9. Регресійний аналіз. Парна лінійна регресія. етап параметризації математичної моделі	146
Вказівки щодо виконання контрольної роботи	166
Вибір варіанту контрольної роботи	166
Приклад оформлення титульного аркуша контрольної роботи	167
Рекомендована література.....	168

Вступ

Відповідно до навчальних програм підготовки магістрів викладається дисципліна "Методи прикладного статистичного аналізу", що є базовою при вивченні методів статистичної обробки результатів експериментальних досліджень.

Навчальні програми цього курсу передбачають проведення лабораторних занять, у процесі яких студенти закріплюють лекційний матеріал і набувають необхідні навички застосування статистичних методів при обробці реальних експериментальних даних. Одночасно закріплюються набуті раніше навички роботи із засобами інформаційних технологій.

Даний практикум містить необхідні теоретичні положення, характерні приклади і варіанти завдань для кожного з занять по розділах: первинна статистична обробка результатів спостережень і перевірка статистичних гіпотез про закони розподілу, вибірках і числових характеристиках.

Практикум призначений для студентів усіх спеціальностей, які вивчають навчальний курс «Методи прикладного статистичного аналізу», і передбачає користування ним в процесі лабораторних занять. Практикум містить довідкові дані, необхідні при виконанні завдань. Даний практикум може виявитися корисним студентам при виконанні курсових і дипломних проектів або навчально-дослідних робіт, пов'язаних з проведенням експериментальних досліджень і статистичною обробкою результатів спостережень.

Завдання № 1
Вибірковий метод
Короткий теоретичний матеріал

Вибірковий метод полягає в тому, що з сукупності елементів відбирається яка - небудь частина. Вся сукупність однорідних елементів, що підлягає вивченню, називається генеральною сукупністю. Частина випадково відібраних елементів називається вибірковою сукупністю чи вибіркою.

Число об'єктів в генеральній сукупності або в вибірці називається їх об'ємом. Генеральна сукупність може бути кінцевою або нескінченною.

При кінцевій генеральній сукупності кожен її елемент представляється реально існуючим індивідумом або наявним значенням випадкової величини. Обсяг кінцевої генеральної сукупності дорівнює числу проведених спостережень (дослідів) над випадковою величиною. Часто, особливо в технічних додатках, у якості генеральної сукупності зручно розглядати безліч значень випадкової величини, які могли б з'явитися при проведенні спостережень над випадковою величиною. Число спостережень над випадковою величиною теоретично може бути як завгодно великим, тому генеральна сукупність значень спостережуваної випадкової величини нескінченна. Вибірка об'єму у такому розумінні буде називатися сукупністю значень випадкової величини, що спостерігаються в досліді. Ці значення, що є елементами вибірки, будемо називати спостережуваними або вибірковими значеннями. Проводячи відбір, слід прагнути до того, щоб вибірка була репрезентативною (представницькою), тобто щоб вона створювала уявлення про генеральну сукупність. В іншому випадку на основі вибірки неможливо зробити висновок щодо всієї сукупності. При нескінченній генеральній сукупності репрезентативність вибірки залежить від властивостей спостережуваної випадкової величини і забезпечується спеціальними заходами при проведенні спостережень, в число яких входять: стабільність умов проведення дослідів, мінімізація помилок вимірювальних засобів і ін. При кінцевій генеральній сукупності для забезпечення репрезентативності вибірки необхідно проводити відбір об'єктів випадковим чином. Принцип випадкового відбору можна здійснити за допомогою жереба, або таблиці випадкових чисел.

ЗАВДАННЯ:

З генеральної сукупності обсягу N сформувати вибірку сукупності обсягу n .

Для чого:

1. Сформувати таблицю випадкових чисел (табл.1).
2. Об'єднати кількість стовпців, відповідних до розрядності заданої вибіркової сукупності.
3. Вибрати послідовність необхідних чисел з таблиці, пропускаючи числа, що перевищують найбільший номер вибірки. Реалізувати процедуру вибору чисел з поверненням, тобто при читанні чисел в таблиці пропускаються ті числа, які вже зустрічалися раніше.
4. Упорядкувати отримані значення номерів вибіркової сукупності.

Приклад. Нехай з 8000 одиниць, що складають генеральну сукупність, вирішено відібрати 10 одиниць. Дослідник, пронумерувавши 8000 одиниць, може скористатися таблицею випадкових чисел (таблиця). Така таблиця складається з послідовності цілих чисел: 0,1,..., 9. Об'єднуючи два стовпця цієї таблиці, ми отримаємо послідовність чисел, утворених з сукупності, що містить 102 чисел: 00,..., 99. Аналогічно для трьох, чотирьох або будь-якого більшого числа стовпців.

Для наведеної вище генеральної сукупності необхідно об'єднати чотири стовпці і вибрати послідовність чотиризначних чисел з таблиці, пропускаючи числа, що перевищують 8000, поки не

наберемо 10 чисел. Якщо здійснюється вибір без повернення, то при читанні чисел в таблиці ми повинні пропускати ті числа, які вже зустрічалися раніше.

Отже, користуючись таблицею 1, відбираються об'єкти з номерами:

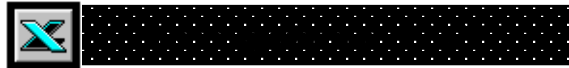
1009, 3754, 0842, 9901, 1280, 6606, 3106, 8526, 6357, 7379, 9852, 1180, 8345, 8868, 9959, 6548.

Підкреслені номери пропускаються.

Отримані номери вибіркового значень упорядковуються:

1009, 3754, 0842, 9901, 1280, 6606, 3106, 8526, 6357, 7379, 9852, 1180, 8345, 8868, 9959, 6548.
0842, 1009, 1180, 1280, 3106, 3754, 6357, 6548, 6606, 7379.

Розв'язання задачі в середовищі EXCEL:



1. Сформувати таблицю випадкових чисел.

Для формування таблиці випадкових чисел скористаємося надбудовою середовища EXCEL «Аналіз даних». Для чого скористаємося пунктом верхнього меню середовища EXCEL Сервіс - Надбудови ... (рис.1.1).

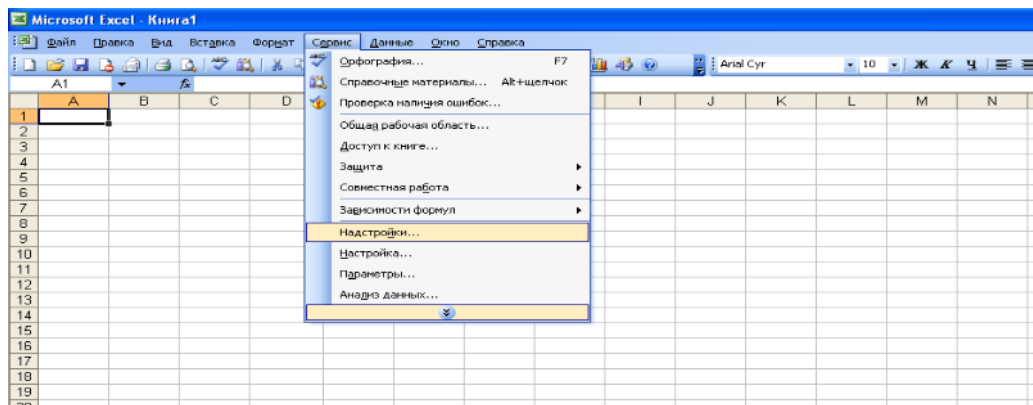


Рис.1.1. Вікно середовища EXCEL Сервіс - Надбудови ... У пункті меню

«Надстройки» вибирають в доступних надбудовах «Пакет аналізу» (рис.1.2).

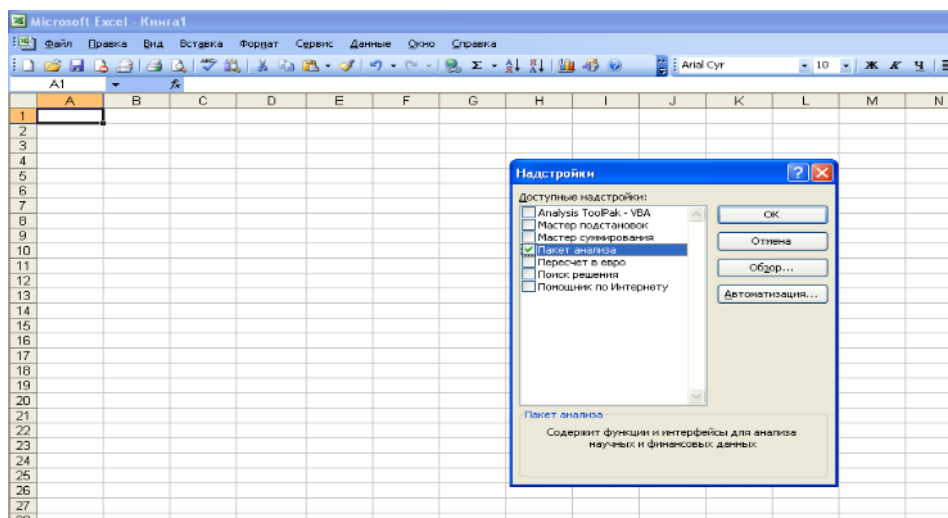


Рис. 1.2. Активізація пакета аналізу даних

У верхній лінійці пункту меню «Сервис» з'являється вкладення «Аналіз даних ...» (рис.1.3).

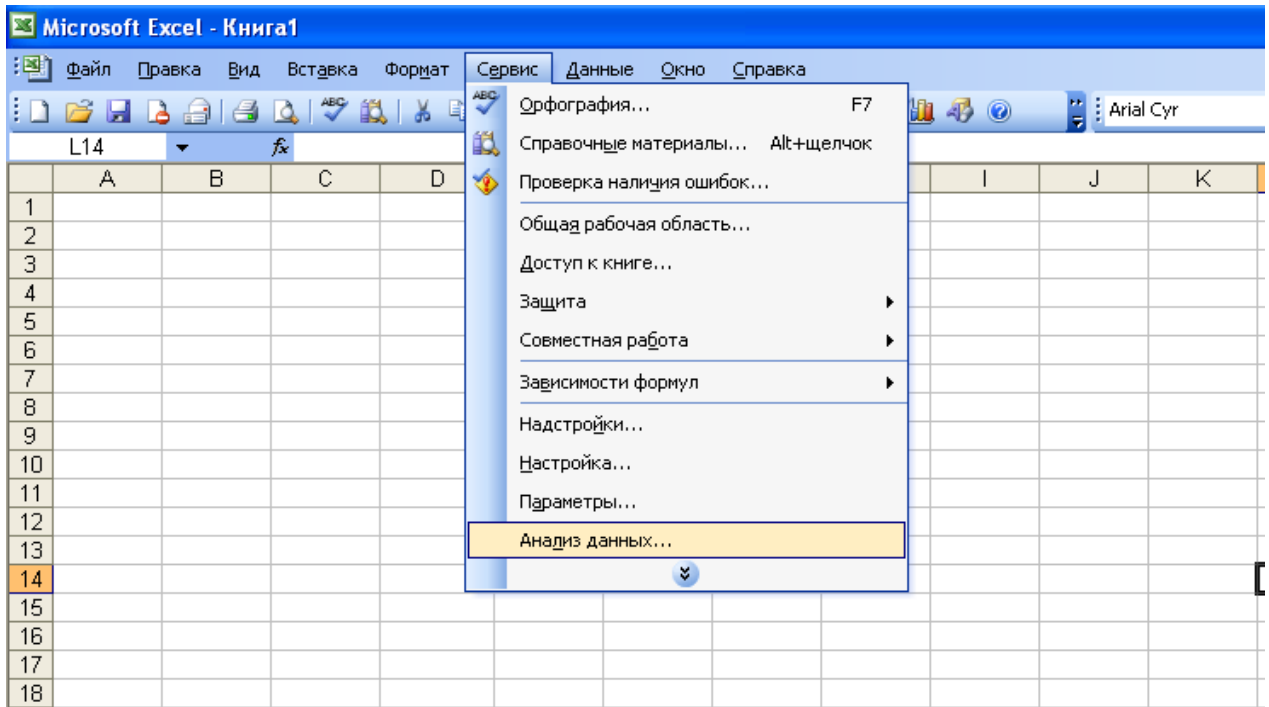


Рис. 1.3. Анализ данных пункту меню «Сервис».

Встановити активної клітинку A1 і звернутися до пункту меню «Сервис» і вкладення «Аналіз даних ...». При цьому у вікні «Аналіз даних» звернутися до пункту меню «Генерация случайных чисел»

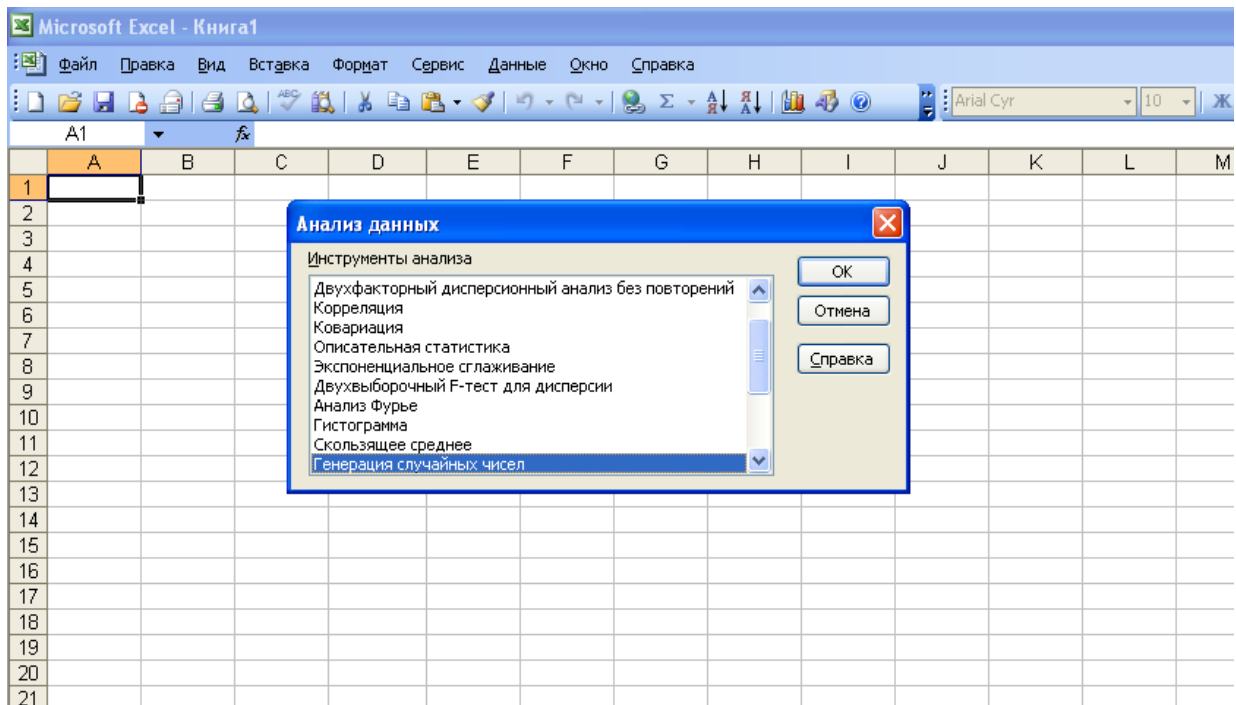


Рис. 1.4. Генерация случайных чисел в среде EXCEL

Далі у вікні «Генерация случайных чисел» вказати число випадкових чисел і закон розподілу (рис.1.5).

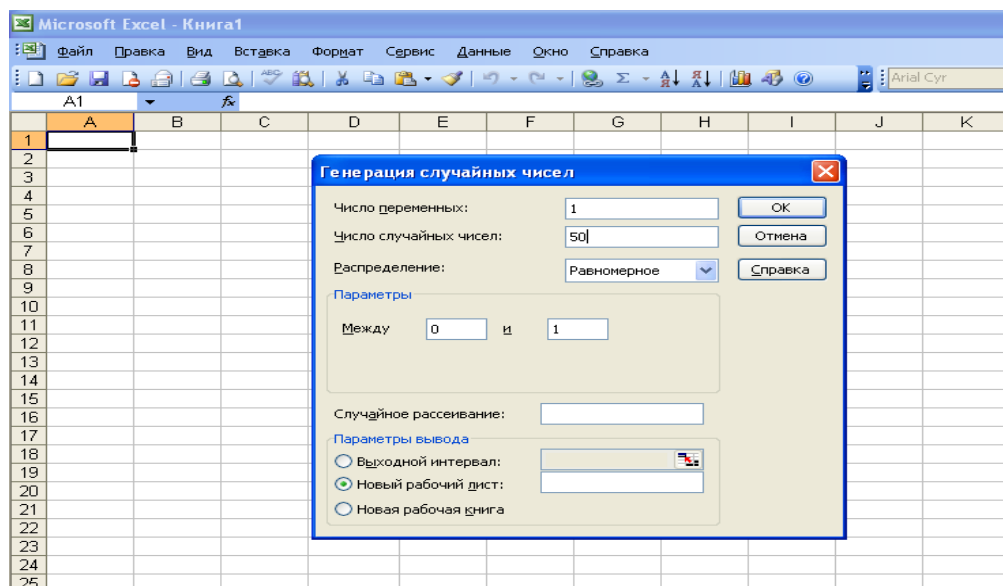


Рис. 1.5. Вибір параметрів генератора випадкових чисел
В діапазоні клітинок A1-A50 будуть розташовуватися випадкові числа (рис. 1.6).

	A	B	C	D	E	F	G	H	I	J
10	0,177313									
11	0,47322									
12	0,809809									
13	0,447066									
14	0,134892									
15	0,150853									
16	0,729331									
17	0,020692									
18	0,849147									
19	0,407941									
20	0,643544									
21	0,567248									
22	0,304605									
23	0,867275									
24	0,454146									
25	0,571673									
26	0,976775									
27	0,81225									
28	0,682485									
29	0,855159									
30	0,746269									
31	0,067843									
32	0,95172									
33	0,753838									
34	0,487838									
35	0,860225									
36	0,533158									

Рис. 1.6. Генерована таблиця випадкових чисел

1. Об'єднати кількість стовпців, що відповідають розрядності заданої вибіркової сукупності.
2. Вибрати послідовність необхідних чисел з таблиці, пропускаючи числа, що перевищують найбільший номер вибірки. Реалізувати процедуру вибору чисел з поверненням, тобто при читанні чисел в таблиці пропускаються ті числа, які вже зустрічалися раніше.



В клітинці C1 ввести множник для виділення відповідної розрядності заданої вибіркової сукупності (в даному випадку 10000).

В клітинку E1 ввести формулу =ОТБР(A1*\$C\$1) і розмножити її для діапазону C1:C50.

Серед отриманих значень виділити жовтим кольору номери вибірки, які задовольняють заданій умові (рис. 1.7).

	A	B	C	D	E	F	G	H	I	J	K
1	0,963134		10000		9631						
2	0,526017				5260						
3	0,482009				4820						
4	0,255531				2555						
5	0,783441				7834						
6	0,835414				8354						
7	0,729179				7291						
8	0,52916				5291						
9	0,294198				2941						
10	0,177313				1773						
11	0,47322				4732						
12	0,809809				8098						
13	0,447066				4470						
14	0,134892				1348						
15	0,150853				1508						
16	0,729331				7293						
17	0,020692				206						
18	0,849147				8491						
19	0,407941				4079						

Рис. 1.7. Номери вибірки, які задовольняють заданій умові

1. Впорядкувати отримані значення номерів вибіркової сукупності. Виконати сортування виділених даних і записати їх в стовпець G.

Варіанти індивідуальних завдань.

З генеральної сукупності об'єму N здійснити вибірку об'єму n (вказати номери відібраних елементів).

№ вар.	N	n
1	500	7
2	600	8
3	700	9
4	400	10
5	300	6
6	200	8
7	1000	11
8	850	12
9	460	9
10	9150	21
11	320	8
12	970	24
13	150	9
14	225	11

15	750	19
16	890	21
17	445	18
18	235	16
19	650	14
20	750	12
21	630	15
22	460	14
23	400	12
24	350	11
25	550	15
26	720	16
27	500	12
28	480	14
29	370	13
30	220	16
31	260	17
32	240	20
33	280	14
34	340	10
35	410	12
36	270	15
37	190	10
38	160	10
40	330	14
41	670	18
42	760	20
43	630	17
44	720	19
45	740	21
46	680	18
47	790	17
48	820	20
49	1240	22
50	476	14

Випадкові числа, рівномірно розподілені в інтервалі $(0;1)$ $[x 10^5]$

10097	32533	76520	13586	34673	54876	80959	09117
37542	04805	64894	74296	24805	24037	20636	10402
08422	68953	19645	09303	23209	02560	15953	34764
99019	02529	09376	70715	38311	31165	88676	74397
12807	99970	80157	36147	64032	36653	98951	16877
66065	74717	34072	76850	36697	36170	65813	39885
31060	10805	45571	82406	35303	42614	86799	07439
85269	77602	02051	65692	68665	74818	73053	85247
63573	32135	05325	47048	90553	57548	28468	28709
73796	45753	03529	64778	35808	34282	60935	20344
98520	17767	14905	68607	22109	40555	60970	94333
11805	05431	39808	27732	50725	68248	29405	42401
83452	99634	06288	98083	13746	70078	18475	40610
88685	40200	86507	58401	36766	67951	90364	76493
99594	67348	87517	64969	91826	08928	93785	61368
65481	17674	17468	50950	58047	76974	73039	57186
80124	35635	17727	08015	45318	22374	21115	78253
74350	99817	77402	77214	43236	00210	45521	64237
69915	26803	66252	29148	36936	87203	76621	13990
09893	20505	14225	68514	46427	56788	96297	78822
91499	14523	68479	27686	46162	83554	94750	89923
80336	94598	26940	36858	70297	34135	53140	33340
44104	81949	85157	47954	32979	26575	57600	40881
12550	73742	11100	02040	12860	74697	96644	89439
63606	49329	16505	34484	40219	52563	43651	77082
61196	90446	26457	47774	51924	33729	65391	59593
15474	45266	95270	79953	59367	83848	82396	10118
94557	28573	67897	54387	54622	44431	91190	42592
42481	16213	97344	08721	16868	48767	03071	12059
23523	78317	73208	73208	68935	91416	26252	29663
04493	52494	75246	75246	45862	51025	61962	79335
00549	97654	64051	64051	96119	63896	64692	82391
35963	15307	26898	26898	33351	35462	77974	50024
59808	09391	45427	45427	83609	49700	13021	24892
46058	85236	01390	01390	77281	44077	93910	83647
32179	00597	87379	87379	05567	07007	86743	17157
69234	61406	20117	20117	15956	60000	18743	92423
19565	41430	01758	01758	45204	21585	66674	36806
45155	14938	19476	19476	75379	94543	59047	90033
94864	31994	36168	36168	07246	81553	01540	35456

Завдання № 2

Первинна обробка статистичних даних дискретної випадкової ознаки

Краткий теоретичний матеріал

Значна частина методів прикладного статистичного аналізу пов'язана з необхідністю опису і вивчення великої сукупності об'єктів. Вибірковий метод полягає в тому, що з великої сукупності об'єктів вибирається якась частина. Вся сукупність однорідних елементів, що підлягає вивченню, називається **генеральною сукупністю**.

Наприклад, генеральною сукупністю можуть бути всі жителі Дніпра, місячна продукція заводу, студенти даного потоку і т. д.

Генеральна сукупність може бути кінцевою і нескінченною. При кінцевій генеральній сукупності кожен її елемент представляється реально існуючим індивідумом або наявним значенням випадкової величини. Обсяг кінцевої генеральної сукупності дорівнює числу проведених спостережень (дослідів) над випадковою величиною. Часто, особливо в технічних додатках, у якості генеральної сукупності зручно розглядати множину значень випадкової величини, які могли б з'явитися при проведенні спостережень над випадковою величиною. Число спостережень над випадковою величиною теоретично може бути як завгодно великим, тому генеральна сукупність значень спостережуваної випадкової величини нескінченна.

Сукупність обмеженого числа значень випадкової величини, отриманих в результаті експерименту, називається **вибірковою сукупністю**, або вибіркою з генеральної сукупності.

Закони, побудовані за вибірковими даними, називаються емпіричними. Емпіричний закон встановлюють зв'язок між вибірковими значеннями $x_1, x_2, \dots, x_n$ і відповідними відносними частотами $P_i^* = \frac{m_i}{n}$, де n – об'єм вибірки, тобто $n = \sum_{i=1}^k m_i$, а m_i – числа, що показують скільки разів дане значення x_i зустрічалось серед вибіркових даних. Відносна частота є статистичним аналогом ймовірності ($P_i^* \xrightarrow{\text{вер. } n \rightarrow \infty} P_i$).

Якщо вибіркові значення розташувати в порядку їх зростання, то такий ряд називається **не згрупованим варіаційним рядом**. Порядковий номер елемента в варіаційному ряду називається **рангом**.

При великій кількості спостережень вибіркові значення групуються.

Якщо спостерігається дискретна випадкова величина X , то, позначивши через x_i – можливі значення випадкової величини, а через m_i – число, що показує скільки раз спостерігалось значення x_i (це число іноді називають частотою), в результаті угруповання отримують дискретний варіаційний ряд.

Дискретний варіаційний ряд має вигляд

x_i	x_1	x_2	$\dots$	x_k
m_i	m_1	m_2	$\dots$	m_k

$$, \text{ де } \sum_{i=1}^k m_i = n$$

Якщо таблицю згрупованого варіаційного ряду доповнити рядком відносних частот, то отримаємо статистичний ряд.

Дискретний статистичний ряд має вигляд.

x_i	x_1	x_2	...	x_k
m_i	m_1	m_2	...	m_k
P_i^*	P_1^*	P_2^*	...	P_i^*

, де $\sum_{i=1}^k m_i = n$
 $P_i^* = \frac{m_i}{n}, \sum_{i=1}^k P_i^* = 1$

Графічно дискретний статистичний ряд зображується у вигляді *полігону*. **Полігон** - це ламана, що з'єднує точки з координатами (x_i, P_i^*) .

Дискретний статистичний ряд є аналогом ряду розподілу, а полігон - аналогом багатокутника розподілу.

Статистичним аналогом функції розподілу $F(x)$ є емпірична функція розподілу $F^*(x) = P^*(X < x)$, що дорівнює частоті вибірових значень, менших деякого фіксованого значення x .

Для дискретного статистичного ряду емпірична функція задається наступним чином:

$$F^*(x) = \begin{cases} 0, & \text{при } x \leq x_1 \\ P_1^*, & \text{при } x_1 < x \leq x_2 \\ P_1^* + P_2^*, & \text{при } x_2 < x \leq x_3 \\ \dots & \\ 1, & \text{при } x_k < x < \infty \end{cases}$$

ЗАВДАННЯ:

За заданою вибірковою сукупністю побудувати:

1. Дискретний варіаційний ряд.
2. Дискретний статистичний ряд.
3. Полігон відносних частот.
4. Емпіричну функцію розподілу і її графік.

Приклад 1.

Результати вибіркового контролю якості продукції дали наступне число бракованих виробів по днях місяця за 25 робочих днів:

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
m_i	1	0	1	0	7	0	1	1	0	0	0	6	1	1	0	2	0	0	1	1	1	0	2	1	0

Зазначимо можливі значення випадкової величини X - числа бракованих виробів в порядку їх зростання і підрахуємо число появ кожного значення (можливі значення величина X більше 7 вважаємо малоймовірними і їх не розглядаємо)

x_i	0	1	2	3	4	5	6	7
m_i	11	10	2	0	0	0	1	1

Побудована таблиця і є дискретним варіаційним рядом.

Побудуємо дискретний статистичний ряд для заданого прикладу. Тут абсолютні частоти p_i^* знаходяться як $p_i^* = \frac{m_i}{n}$,

x_i	0	1	2	3	4	5	6	7
m_i	11	10	2	0	0	0	1	1
p_i^*	0,44	0,4	0,08	0	0	0	0,04	0,04

Дискретний статистичний ряд графічно оформлюється у вигляді полігону. Полігон будується наступним чином: на осі абсцис відкладаються значення статистичного ряду, в кожному з них будується ордината, величина якої дорівнює або пропорційна частоті відповідного значення. Кінці ординат з'єднуються відрізками прямих.

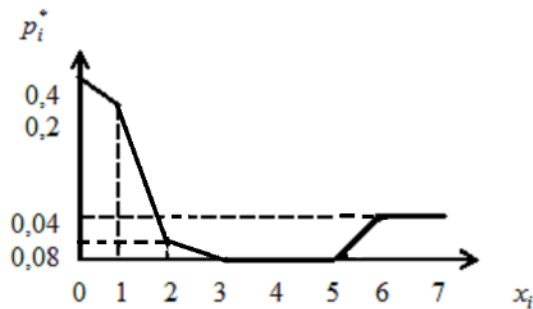


Рис.2.1. Полігон відносних частот

Побудуємо емпіричну функцію розподілу за даним розподілом вибірки, представленому дискретним варіаційним рядом

$$\begin{array}{c|ccc} x_i & 1 & 4 & 6 \\ \hline m_x & 10 & 15 & 25 \end{array}; \quad n = 10 + 15 + 25 = 50;$$

$$x < 1; F^*(x) = \frac{m^*}{n} = \frac{0}{n} = 0; \quad (\text{менше 1 значення не спостерігалися } m^* = 0);$$

$$1 < x \leq 4; F^*(x) = \frac{m^*}{n} = \frac{10}{50} = 0,2; \quad (\text{менше 4 спостерігалось одне значення - 1, але воно зустрічалось 10 раз; } m^* = 10);$$

$$4 < x \leq 6; F^*(x) = \frac{m^*}{n} = \frac{25}{50} = 0,5. \quad (\text{менше 6 два значення; } x_1 = 1 \text{ спостерігалось 10 раз; } x = 4 \text{ спостерігалось 15 раз; } m^* = 10 + 15 = 25).$$

$$6 < x < \infty; F^*(x) = \frac{m^*}{n} = 1.$$

Зобразимо графік цієї функції:

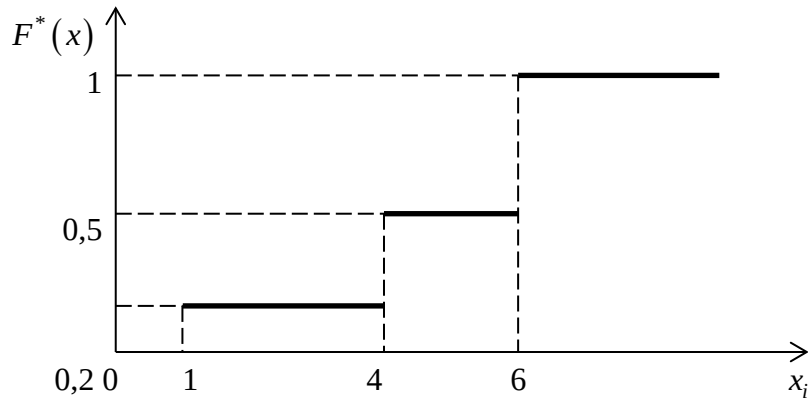


Рис.2.2. Емпірична функція розподілу

Приклад 2. X – кількість працюючих завалочних машин в цеху.

Вихідні дані за 10 днів: 2, 3, 1, 0, 2, 2, 3, 1, 1, 1.

Побудуємо статистичний ряд.

Для цього підрахуємо, скільки разів дане значення зустрілося в вибірці. Наприклад, значення $x_1 = 0$, зустрічалося один раз, значить $m_1 = 0$. Потім розділимо частоти на об'єм вибірки (кількість відібраних елементів). У нашому випадку $n=10$.

x_i	0	1	2	3
m_i	1	4	3	2
P_i^*	0,1	0,4	0,3	0,2

$$\sum_{i=1}^4 m_i = 1 + 4 + 3 + 2 = 10;$$

$$P_i^* = \frac{m_i}{n}, \quad \sum_{i=1}^4 P_i^* = 1$$

Полігон має вигляд

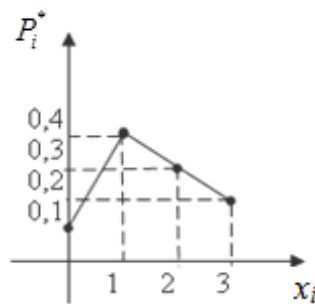


Рис.2.3. Полігон відносних частот

Емпірична функція розподілу

$$F^*(x) = \begin{cases} 0 & , \quad x \leq 0 ; \\ 0,1 & , \quad 0 < x \leq 1; \\ 0,1+0,4 & , \quad 1 < x \leq 2; \\ 0,1+0,4+0,3 & , \quad 2 < x \leq 3; \\ 1 & , \quad 3 < x < \infty; \end{cases}$$

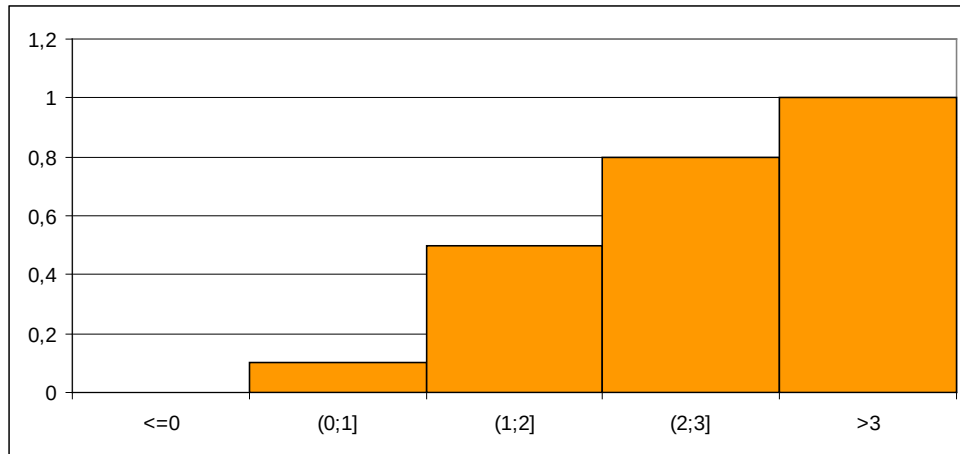


Рис. 2.4. Графік емпіричної функції розподілу

Розв'язання задачі в середовищі EXCEL:

За заданою вибірковою сукупністю побудувати:

1. Дискретний варіаційний ряд.



- У комірку A5 і нижче занести свій варіант вибірки.
- Вибірковому діапазону даних присвоїти ім'я. Для цього виділяється необхідний діапазон даних і в поле імен рядка введення формул середовища EXCEL вводиться ім'я даному діапазону (наприклад, вибірка).
- У комірку C5 вводять $N =$
- У комірку D5 вводять формулу $=\text{СЧЁТ}(\text{вибірка})$ і визначають об'єм вибірки (рис.2.5).

	A	B	C	D	E	F	G	H	I	J
1										
2										
3										
4										
5	2		N=	10						
6	3									
7	1									
8	0									
9	2									
10	2									
11	3									
12	1									
13	1									
14	1									
15										
16										
17										
18										

Рис.2.5. Формування вибірки та визначення її об'єму

За заданою вибірковою сукупністю побудувати:.

1. Дискретний статистичний ряд.



– Скопіювати в буфер обміну комірки A5: A14, скопіювати в комірку F5 і виконати сортування за зростанням (кнопка **A→Я**).

– В комірки I6 і нижче занести *варіанти*, в J6 і нижче - їх абсолютні частоти.

Абсолютні частоти визначити за допомогою функції **=СЧЁТЕСЛИ(вибірка; I6)** яка заноситься в комірку J6. Потім ця функція тиражується на необхідний діапазон

-В клітинці K6 і нижче підрахувати *відносні частоти*.

Microsoft Excel - Задача_2														
Файл Правка Вид Вставка Формат Сервис Данные Окно Справка														
G21 fx Arial Cyr 10 Ж К Ч														
	A	B	C	D	E	F	G	H	I	J	K	L	M	
1														
2														
3														
4														
5	2		N=	10		0								
6	3					1								
7	1					1								
8	0					1								
9	2					1								
10	2					2								
11	3					2								
12	1					2								
13	1					3								
14	1					3								
15														
16														
17														
18														

Рис.2.6. Формування дискретного статистичного ряду

За заданою вибірковою сукупністю побудувати:

1. Полігон відносних частот.



За даними з статистичного ряду будується графік. На осі абсцис відкладається x_i , на осі ординат – p_i (або w_i). Поєднуючи ламаною лінією ці точки, отримуємо фігуру, яка називається *полігон*.

Викликаючи «Мастер Диаграмм», «Точечная Диаграмма», будують *полігон частот*. Редагують графік таким чином, щоб він був максимально схожий на наведений на рис.2.7.

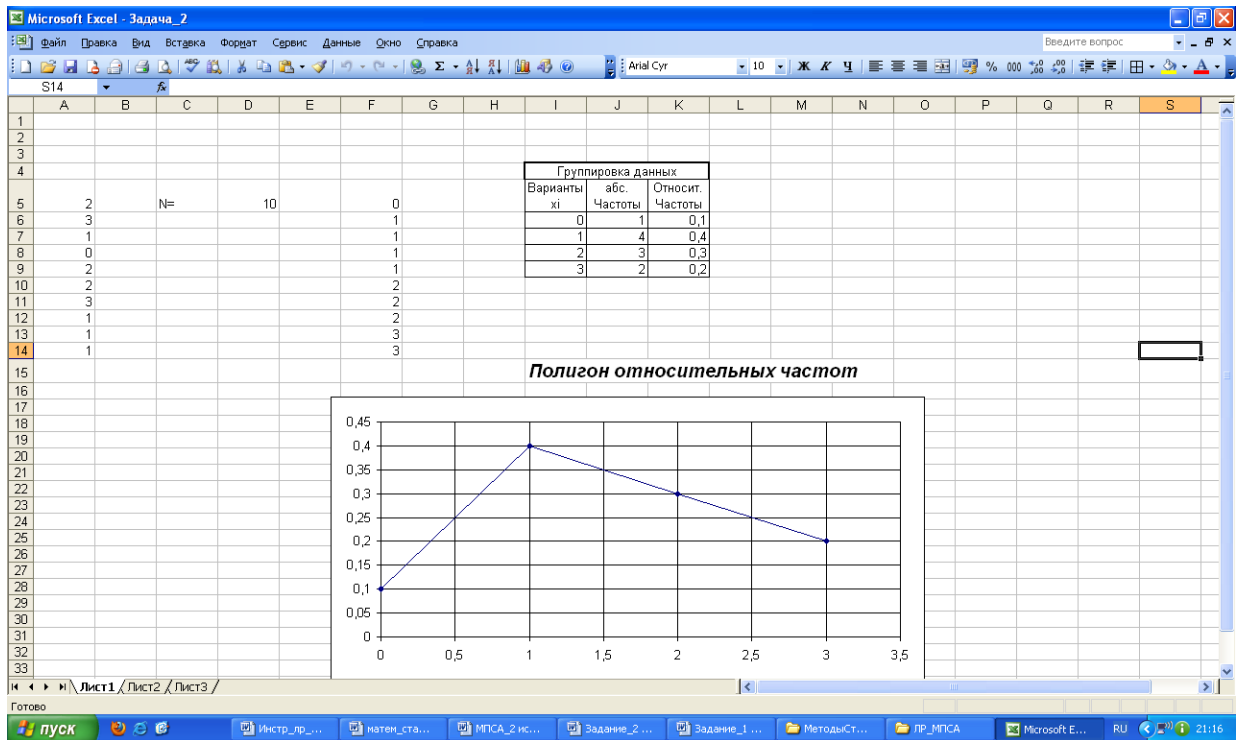


Рис.2.7. Полігон відносних частот

За заданою вибірковою сукупністю побудувати:

- Емпіричну функцію розподілу і її графік



В діапазон комірок D43: G47 ввести аналітичну формулу емпіричної функції розподілу (рис.2.8).

Функция распределения		
	0	при $x < 0$
	0,1	при $0 \leq x < 1$
$F(x) =$	0,5	при $1 \leq x < 2$
	0,8	при $2 \leq x < 3$
	1	при $x \geq 3$

Рис. 2.8. Аналітична формула емпіричної функції розподілу

- В діапазон комірок C49: F49 ввести варіанти (x_i) випадкової величини.
- В діапазон комірок B50:F50 ввести значення емпіричної функції розподілу.
- В діапазон комірок B51:F51 ввести значення інтервалів емпіричної функції розподілу.

Для цієї мети необхідно скористатися функціями:

= "<="&C49	для комірки B51
= "("&C49&" ";"&D49&""]"	для комірки C51
= "("&D49&" ";"&E49&""]"	для комірки D51
= "("&E49&" ";"&F49&""]"	для комірки E51
= ">"&F49	для комірки F51

Побудувати емпіричну функцію розподілу (рис. 2.9) за підготовленими даними:

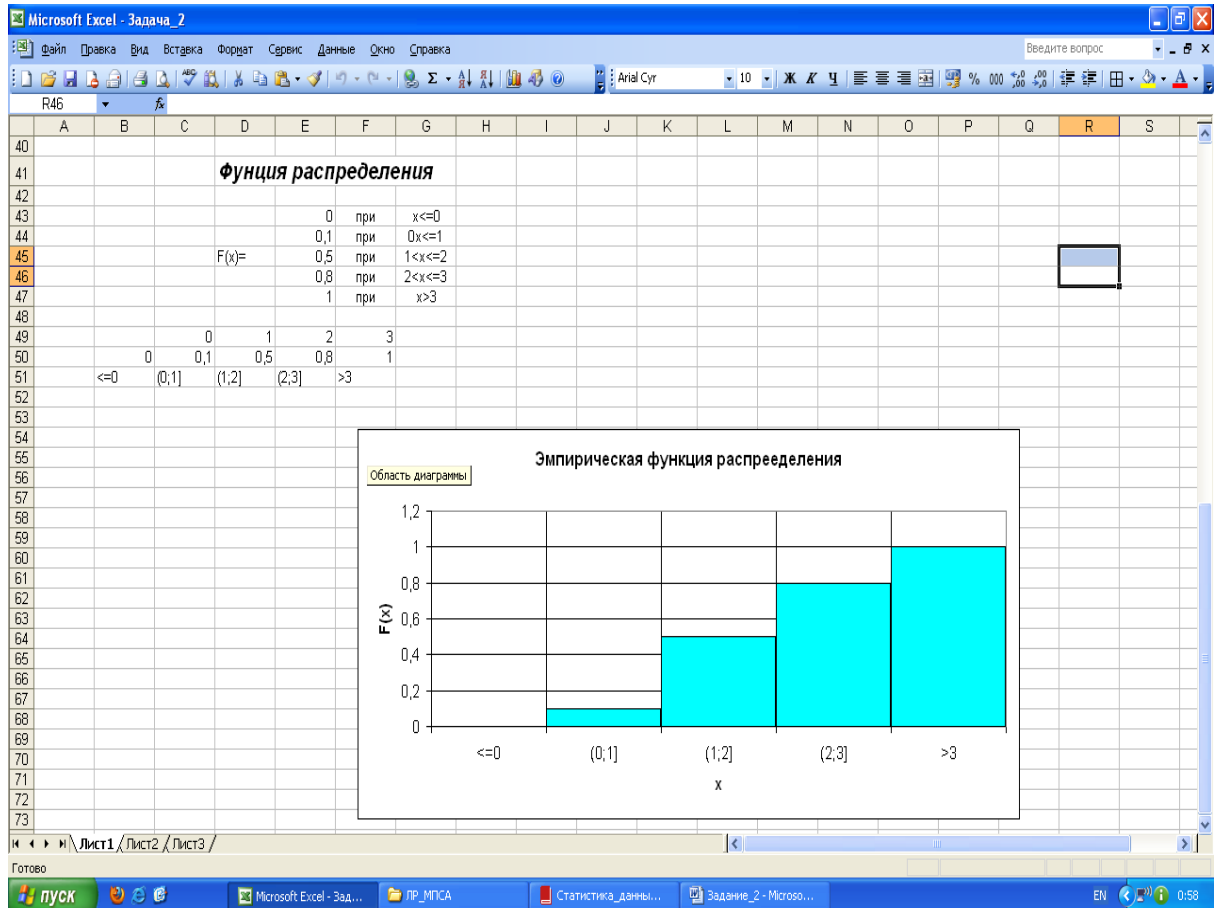


Рис.2.9. Графік емпіричної функції розподілу

Особенности формирования и форматирования графика эмпирической функции распределения.

1. Викликаючи «Мастер Диаграмм», «Гистограмма», будують емпіричну функцію розподілу.
2. Призначають стовпцями нульові перекриття і ширину зазорів. Виконують цю операцію на сторінці Параметри команды Формат рядов данных (рис.2.10).

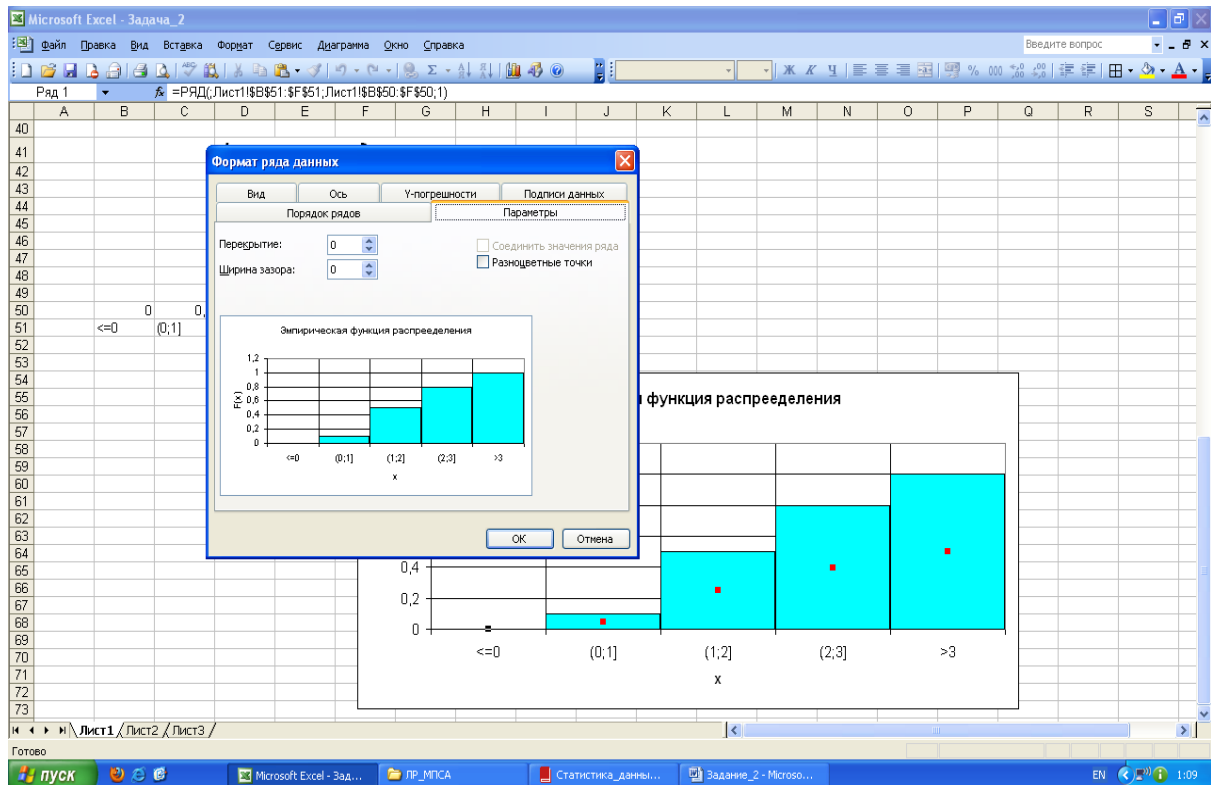


Рис. 2.10. Визначення ширини зазорів діаграми

3. Обирають для стовпців діаграми необхідну заливку. Виконують цю операцію на сторінці **Вид команды Формат ряда данных** (рис.2.11).

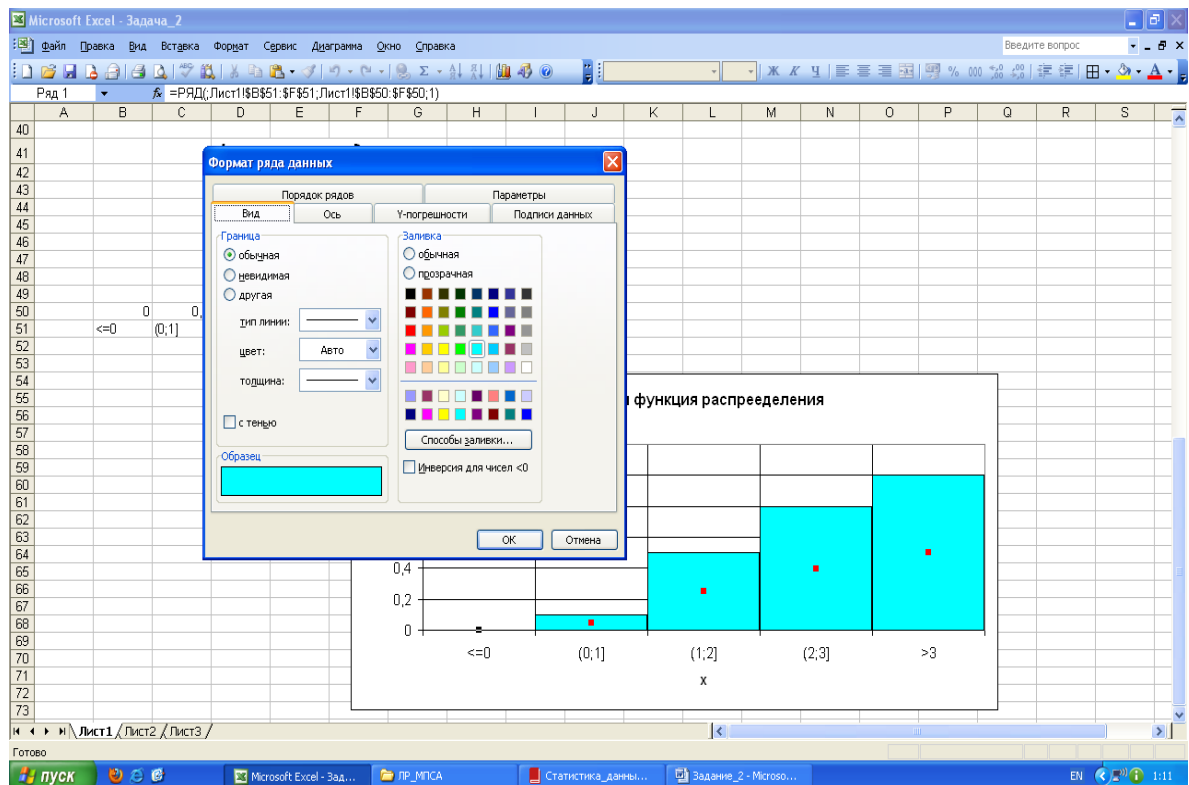


Рис. 2.11. Вибір кольору заливки стовпців діаграми

Варіанти індивідуальних завдань.

Побудувати дискретний статистичний ряд, полігон і $F^*(x)$ для випадкової ознаки X - кількості працюючих завалочних машин в кожному цеху, якщо є відомості по 10 цехах за 15 днів:

1. 2,1,4,1,1,2,1,3,2,1,2,1,3,1,2.
2. 1,2,2,3,2,3,3,2,4,4,1,1,2,3,2.
3. 2,1,2,1,2,2,1,4,2,3,4,1,2,1,3.
4. 1,3,3,2,3,2,2,4,3,4,1,2,4,3,3.
5. 1,2,1,1,2,1,4,2,3,2,3,3,4,2,1.
6. 3,4,2,3,1,3,2,4,4,3,2,1,2,3,4.
7. 2,1,2,1,1,3,4,1,2,1,4,3,4,2,3.
8. 3,2,3,4,3,1,3,2,3,4,1,2,1,3,4.
9. 2,1,2,1,3,1,2,3,2,3,3,1,4,1,2.
10. 2,2,4,1,1,2,1,1,2,2,1,2,3,2,3.
11. 1,1,4,3,1,2,2,3,2,1,2,1,3,4,2.
12. 1,1,2,3,2,2,3,2,4,1,1,1,2,3,2.
13. 2,1,2,1,3,3,1,3,2,3,4,1,2,4,3.
14. 1,3,3,2,4,2,3,4,3,4,2,2,4,1,3.
15. 2,2,1,3,2,3,4,2,3,2,1,3,4,2,1.
16. 2,4,2,2,1,3,2,4,4,3,2,1,2,4,4.
17. 1,1,3,1,1,3,4,2,2,1,3,3,4,2,3.
18. 2,2,1,4,3,1,2,2,3,4,1,2,3,3,4.
19. 1,1,1,4,2,1,1,3,2,3,3,1,2,3,3.
20. 2,2,3,1,3,2,4,1,2,4,1,1,3,2,3.
21. 2,1,3,1,2,2,1,3,2,2,2,1,3,1,4.
22. 1,2,2,2,2,3,3,4,4,4,1,1,2,3,2.
23. 1,1,2,1,3,3,1,4,2,4,4,1,2,1,3.
24. 1,2,3,2,4,2,2,4,3,4,2,2,4,3,4.
25. 1,3,1,1,3,1,4,2,2,2,3,3,4,2,1.
26. 3,2,2,3,1,3,2,4,4,4,2,1,2,1,4.
27. 2,1,1,1,1,3,4,1,3,1,4,3,4,2,2.
28. 3,3,3,4,3,1,4,2,3,4,1,2,3,3,4.
29. 1,1,2,1,3,1,1,4,2,3,3,4,2,1,2.
30. 1,3,4,1,3,2,2,4,2,2,1,3,3,2,3.
31. 1,1,4,1,2,2,1,2,2,1,2,1,3,1,3.
32. 1,3,2,3,2,3,3,4,4,4,1,1,2,3,2.
33. 3,1,2,1,3,2,1,4,2,3,4,3,2,1,3.
34. 1,2,3,2,3,4,2,4,3,4,1,2,4,3,4.
35. 1,3,1,1,2,1,3,2,3,4,3,3,4,2,4.
36. 1,2,2,4,1,3,2,4,4,3,4,1,4,2,4.
37. 1,1,2,1,1,3,4,1,3,1,4,3,4,3,3.
38. 3,1,3,4,3,4,3,2,3,4,1,2,4,3,4.
39. 2,1,2,3,3,1,1,4,4,3,3,4,2,1,2.
40. 1,2,4,1,3,3,2,4,3,2,1,3,4,2,3.
41. 1,2,4,1,2,2,3,2,2,1,2,3,3,1,4.
42. 1,2,2,3,2,3,3,4,4,4,1,1,2,4,4.
43. 1,1,2,3,3,4,1,4,2,4,4,3,2,1,3.
44. 1,1,3,2,4,4,2,4,3,4,1,4,4,3,4.
45. 1,3,1,2,2,1,2,2,3,4,3,3,4,4,4.
46. 1,3,2,4,1,3,3,4,4,3,4,1,4,2,4.
47. 1,3,2,1,4,3,4,1,3,1,4,3,4,3,4.
48. 3,2,3,4,3,4,3,2,3,4,1,2,4,4,4.
49. 2,1,2,1,2,1,1,3,2,3,3,1,2,1,2.
50. 1,2,4,1,3,2,2,1,2,2,1,1,3,2,3.

Завдання № 3
Первинна обробка статистичних даних неперервної випадкової ознаки
Короткий теоретичний матеріал

Якщо величина, що спостерігається, є неперервною, то угруповання вибірових значень проводять наступним чином. На першому етапі складають інтервальний статистичний ряд. Для побудови інтервального варіаційного ряду необхідно визначити довжину інтервалу (Δx_i) і саму ліву точку, від якої відкладається інтервал. Ці питання найчастіше вирішує експериментатор так, щоб число інтервалів було не надто малим, але й не занадто великим. Якщо всі інтервали мають однакову довжину (Δx), то можна скористатися для визначення формулою Стерджесса

$$\Delta x = \frac{x_{\max} - x_{\min}}{1 + 3,322 \lg n},$$

де: $x_{\min}$ – найменше значення у вибірці; $x_{\max}$ – найбільше значення у вибірці.

Величину Δx можна округляти.

Ліва межа першого інтервалу визначається наступним чином:

$$x^{(1)} = x_{\min} - \frac{\Delta x}{2}.$$

Потім отримують межі інтервалів:

$$x^{(2)} = x^{(1)} + \Delta x;$$

$$x^{(3)} = x^{(2)} + \Delta x;$$

...

Такий процес розбиття реалізують до тих пір, поки величина $x^{(k+1)}$ стане більшою $x_{\max}$.

Таким чином отримують **інтервальний варіаційний ряд**, який має наступний вигляд:

J_i	$[x^{(1)}; x^{(2)})$	$[x^{(2)}; x^{(3)})$	...	$[x^{(k)}; x^{(k+1)})$
m_i	m_1	m_2	...	m_k

Тут m_i показує, скільки вибірових значень потрапляє в i -й інтервал. Квадратна дужка означає, що тут включається ліва межа інтервалу і не включається права (хоча можна інтервали сформулювати і інакше, наприклад, включити праву і не включити ліву межу).

Маючи інтервальний варіаційний ряд, легко підрахувати відносні частоти. Якщо таблицю згрупованого варіаційного ряду доповнити рядком відносних частот, то отримують інтервальний статистичний ряд.

Інтервальний статистичний ряд представляється таблицею виду:

J_i	$[x^{(1)}; x^{(2)})$	$[x^{(2)}; x^{(3)})$	...	$[x^{(k)}; x^{(k+1)})$	
m_i	m_1	m_2	...	m_k	
P_i^*	P_1^*	P_2^*	...	P_i^*	

Середній рядок (рядок m_i) можна опускати.

Графічно **інтервальний статистичний ряд** зображують у вигляді полігону і гістограми.

Полігон - це ламана, що з'єднує точки з координатами $(\bar{x}_i, P_i^*)$, де $\bar{x}_i$ - середина i -того інтервалу.

Гістограма - це плоска фігура, що складається з прямокутників, підстави яких дорівнюють довжині інтервалу, а висота дорівнює або пропорційна відносній частоті. Щоб площа гістограми

дорівнювала одиниці, беруть $h_i = \frac{P_i^*}{\Delta x_i}$.

За допомогою інтервального статистичного ряду, а саме, скориставшись полігоном або гістограмою, можна судити про щільність розподілу $f(x)$.

Статистичним аналогом функції розподілу $F(x)$ є емпірична функція розподілу $F^*(x) = P^*(X \leq x)$, що дорівнює частоті вибірових значень, менших деякого фіксованого значення x .

Емпірична функція розподілу для **інтервального статистичного ряду** задається наступним чином:

$$F^*(x) = \begin{cases} 0, & \text{при } x = x^{(1)} \\ P_1^*, & \text{при } x = x^{(2)} \\ P_1^* + P_2^*, & \text{при } x = x^{(3)} \\ \dots & \\ 1, & \text{при } x = x^{(k+1)} \end{cases}$$

ЗАВДАННЯ:

За заданою вибіркової сукупності побудувати:

1. Інтервальний варіаційний ряд.
2. Інтервальний статистичний ряд.
3. Полігон і гістограма відносних частот.
4. Емпіричну функцію розподілу і її графік.

Приклад 1. Представлені вихідні дані по 40 плавов. X - продуктивність мартенівської печі.
 16,0; 16,4; 16,1; 16,3; 16,4; 16,1; 16,2; 17,0; 16,8; 16,7; 16,6; 16,9; 16,7; 17,4; 17,5; 18,0; 17,6; 17,9;
 15,0; 15,4; 15,3; 15,6; 15,9; 16,0; 15,7; 16,6; 16,8; 16,7; 16,5; 17,0; 15,2; 15,1; 16,4; 16,3; 16,2; 15,6;
 15,9; 15,8; 15,9; 16,4.

Виконати первинну обробку безперервної випадкової ознаки.

Побудуємо інтервальний варіаційний ряд. Для цього визначимо довжину інтервалу за формулою Стерджесса.

$$\Delta x = \frac{x_{\max} - x_{\min}}{1 + 3,322 \lg n} = \frac{18,0 - 15,0}{1 + 3,322 \lg 40} = \frac{3,0}{1 + 3,322 \cdot 1,6} = 0,42.$$

Округлимо крок: $\Delta x = 0,4$.

Далі формують межі інтервалів:

$$x^{(1)} = x_{\min} - \frac{\Delta x}{2} = 15,0 - \frac{0,4}{2} = 14,8;$$

$$x^{(2)} = x^{(1)} + \Delta x = 14,8 + 0,4 = 15,2;$$

$$x^{(3)} = x^{(2)} + \Delta x = 15,2 + 0,4 = 15,6;$$

...

Останнє $x^{(k+1)}$ має бути більше або дорівнює $x_{\max}$.

Формують інтервальний варіаційний ряд, підрахувавши абсолютні частоти (скільки значень потрапляє в кожен інтервал):

J_i	[14,8;15,2)	[15,2;15,6)	[15,6;16,0)	[16,0;16,4)	[16,4;16,8)	[16,8;17,2)	[17,2;17,6)	[17,6;18,0)
m_i	3	4	7	10	8	3	3	2

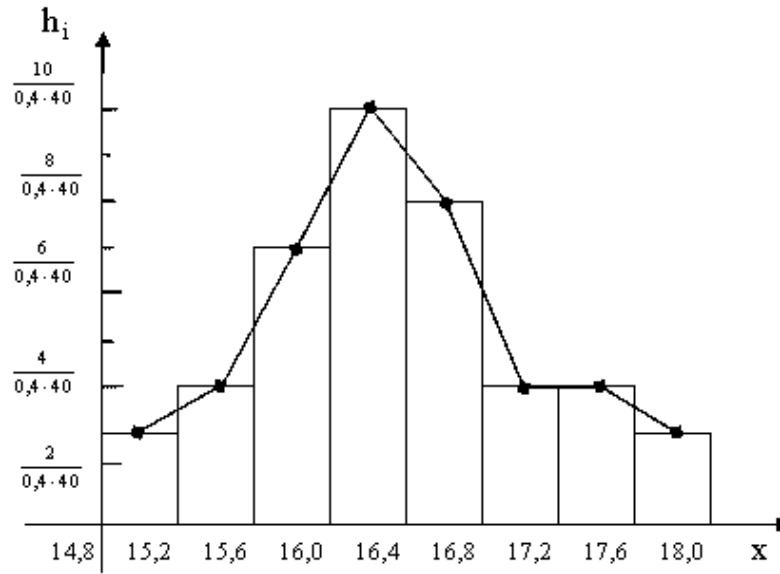
Для побудови інтервального статистичного ряду відносні частоти p_i^* знаходять із співвідношення: $p_i^* = \frac{m_i}{n}$. Тоді інтервальний статистичний ряд буде представлений у вигляді такої таблиці:

J_i	[14,8;15,2)	[15,2;15,6)	[15,6;16,0)	[16,0;16,4)	[16,4;16,8)	[16,8;17,2)	[17,2;17,6)	[17,6;18,0)
m_i	3	4	7	10	8	3	3	2
p_i^*	$\frac{3}{40} = 0,075$	$\frac{4}{40} = 0,1$	$\frac{7}{40} = 0,175$	$\frac{10}{40} = 0,25$	$\frac{8}{40} = 0,2$	$\frac{3}{40} = 0,075$	$\frac{3}{40} = 0,075$	$\frac{2}{40} = 0,05$

Графічно інтервальний статистичний ряд можна зобразити у вигляді гістограми та полігона. Висота прямокутників гістограми визначається з співвідношення:

$$h_i = \frac{p_i^*}{\Delta x_i} = \frac{p_i^*}{0,4}.$$

Полігон, як правило, будують на тому ж рисунку, з'єднавши середини верхніх сторін прямокутника, тому що вид полігону не зміниться, якщо замість P_i^* взяти $\frac{P_i^*}{\Delta x}$.



Емпірична функція розподілу визначається своїми значеннями на кінцях інтервалів:

$$F^*(14,8) = 0; F^*(15,2) = p_1^* = 3/40;$$

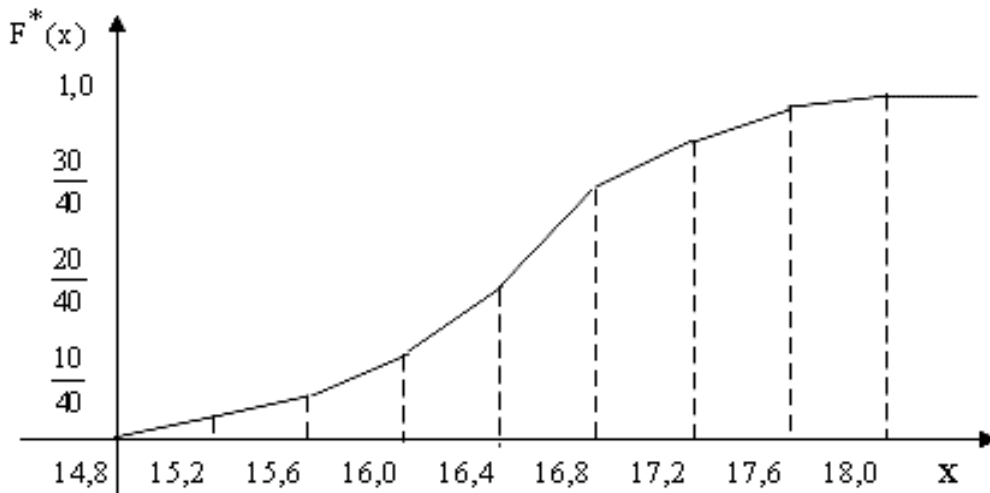
$$F^*(15,6) = \sum_1^2 p_i^* = 7/40; F^*(16,0) = \sum_1^3 p_i^* = 14/40;$$

$$F^*(16,4) = \sum_1^4 p_i^* = 24/40; F^*(16,8) = \sum_1^5 p_i^* = 32/40;$$

$$F^*(17,2) = \sum_1^6 p_i^* = 35/40; F^*(17,6) = \sum_1^7 p_i^* = 38/40;$$

$$F^*(18,0) = \sum_1^8 p_i^* = 1.$$

Поєднавши отримані точки плавною кривою або відрізками прямої, отримаємо графік емпіричної функції розподілу $F^*(x)$:



Розв'язання завдання в середовищі EXCEL:

За заданою вибіркової сукупності побудувати:

1. Інтервальний варіаційний ряд



- У комірку B5 і нижче занести свій варіант вибірки.
- Вибірковому діапазону даних присвоїти ім'я. Для цього виділяється необхідний діапазон даних і в полі імен рядка введення формул середовища EXCEL вводиться ім'я даному діапазону (наприклад, виборка_1).
- Виконати сортування даних (виборка_1) по зростанню (Кнопка А*Я).
- У комірку D5 вводять N =- У комірку E5 вводять формулу =СЧЁТ(виборка)
- і визначають об'єм вибірки (рис.3.1).

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L	M
4													
5		15		n=	40								
6		15,1											
7		15,2											
8		15,3											
9		15,4											
10		15,5											
11		15,6											
12		15,7											
13		15,8											
14		15,9											
15		15,9											
16		15,9											
17		15,9											
18		16											
19		16,1											
20		16,1											
21		16,2											
22		16,2											
23		16,3											
24		16,3											
25		16,4											
26		16,4											

Рис.3.1. Формування вибірки та визначення її обсягу

Для інтервального групування даних проводять такі попередні розрахунки:

1. У комірки G5 і H5 вводять мінімальне і максимальне значення вибіркової сукупності
2. В комірці I5 визначають розмах варіювання $R = X_{max} - X_{min}$.
3. В комірці J5 визначають число інтервалів за формулою:

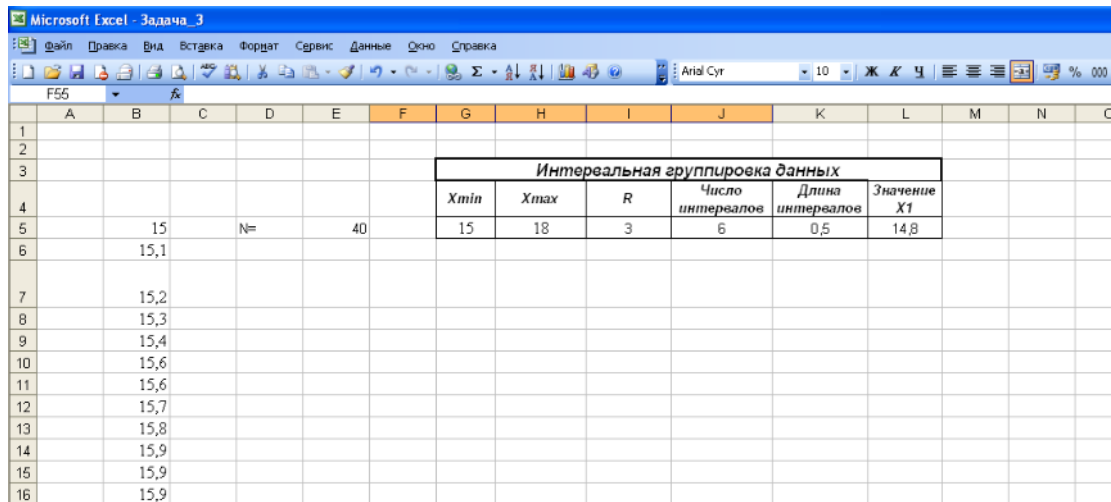
$$k \approx 1 + 3,322 \lg(n).$$

4. В комірці K5 визначають довжину інтервалів за формулою:

$$\Delta x = R / k.$$

При цьому величину Δx округлюють до одного десяткового знака.

5. Значення $x^{(1)}$ визначають на основі співвідношення $x^{(1)} = x_{min} - \frac{\Delta x}{2}$ і заносять у комірку L5, округлюючи до одного десяткового знака (рис.3.2).



Интервальный группировка данных					
X_{min}	X_{max}	R	Число интервалов	Длина интервалов	Значение X_1
15	18	3	6	0,5	14,8

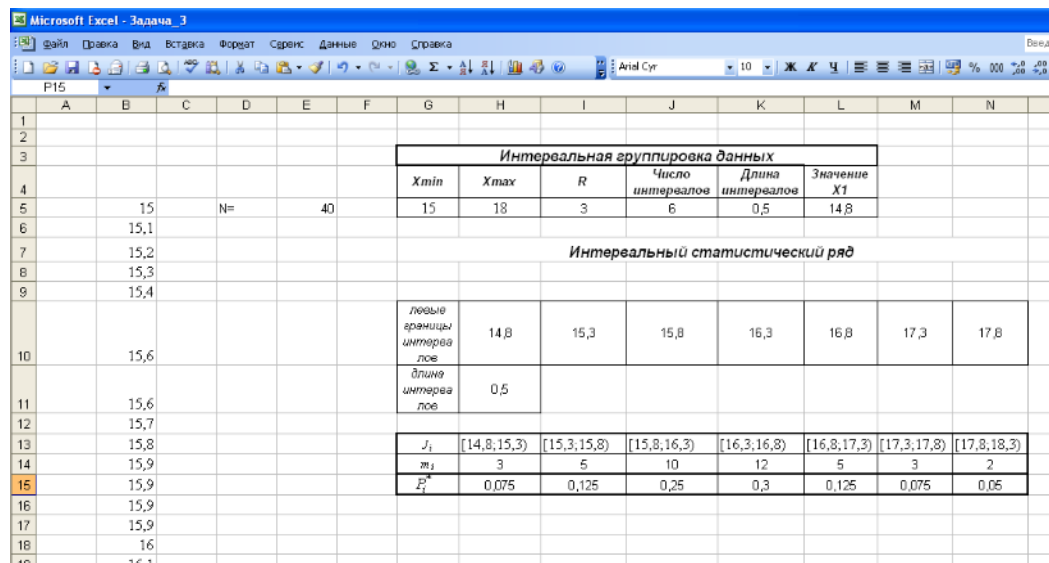
Рис.3.2. Формування даних для інтервального угруповання

За заданою вибірковою сукупністю побудувати:

1. *Інтервальний статистичний ряд*



- У комірки H10 і вправо ввести ліві межі інтервалів.
- У комірки H11 ввести довжину інтервалів.
- У комірки G13, G14, G15 ввести заголовки ряду, відповідно J_i , m_i , P_i^* .
- Починаючи з комірки H13 і по рядку вправо вводять значення інтервалів. При цьому в комірку H13 вводять формулу $=["\&H10\&";"\&H10+\$H\$11\&"]$ і потім тиражують її по рядку вправо.
- Починаючи з комірки H14 і по рядку вправо вводять значення абсолютних частот.
- Починаючи з комірки H15 і по рядку вправо вводять значення відносних частот. При цьому в комірку H13 вводять формулу $=H15/\$K\5 і потім тиражують її по рядку вправо (рис.3.3).



Интервальный группировка данных					
X_{min}	X_{max}	R	Число интервалов	Длина интервалов	Значение X_1
15	18	3	6	0,5	14,8

Интервальный статистический ряд							
левые границы интервалов	14,8	15,3	15,8	16,3	16,8	17,3	17,8
длина интервалов	0,5						
J_i	[14,8;15,3)	[15,3;15,8)	[15,8;16,3)	[16,3;16,8)	[16,8;17,3)	[17,3;17,8)	[17,8;18,3)
m_i	3	5	10	12	5	3	2
P_i^*	0,075	0,125	0,25	0,3	0,125	0,075	0,05

Рис.3.3 Інтервальний статистичний ряд

За заданою вибірковою сукупністю побудувати:

1. Полігон і гістограму відносних частот.



Для побудови гістограм інтервальний статистичний ряд доповнюють рядком висот прямокутників $h_i = \frac{P_i^*}{\Delta x_i}$. Для цієї мети в комірку Н16 вводять формулу =Н15/\$К\$5 і потім тиражують її по рядку вправо (рис.3.4).

[illegible]

Рис. 3.4. Формування висот прямокутників гистограми

Викликаючи «Мастер Диаграмм», «Гистограмма», будують гістограму відносних частот (рис. 3.5).

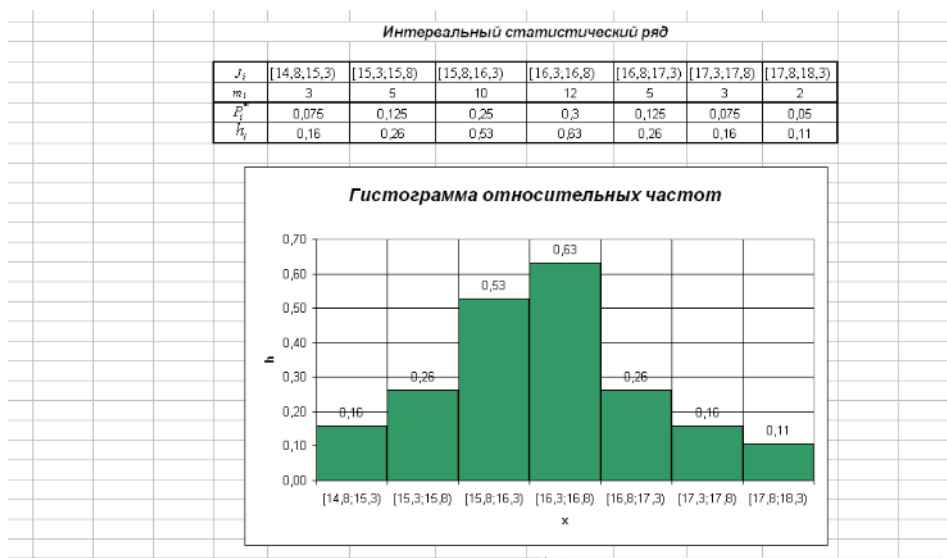


Рис. 5. Гістограма відносних частот

Для побудови полігону відносних частот будують допоміжний ряд, де J_{icp} визначає середину відповідного інтервалу (рис.3.6).

Вспомогательный ряд для построения полигона относительных частот							
J_{icp}	15,05	15,55	16,05	16,25	17,05	17,25	18,05
m_i	3	5	10	12	5	3	2
P_i	0,075	0,125	0,25	0,3	0,125	0,075	0,05

Рис.3.6. Допоміжний ряд для побудови полігону відносних частот

Викликаючи «Мастер Диаграмм», «Точечная Диаграмма», будують полігон відносних частот. Редагують графік таким чином, щоб він був максимально схожий на наведений на рисунку 3.7.

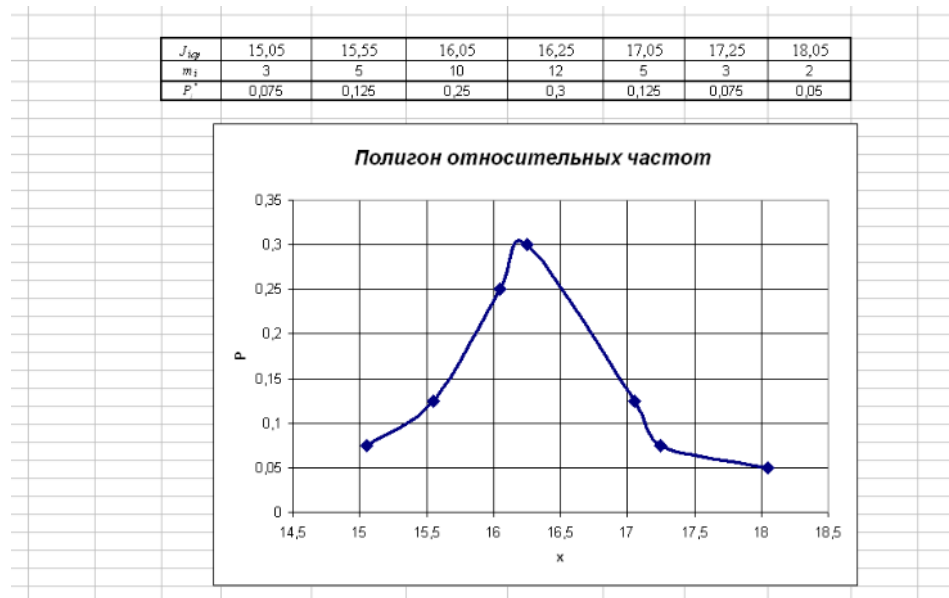


Рис. 3.7. Полігон відносних частот

За заданою вибірковою сукупністю побудувати:

1. Емпіричну функцію розподілу і її графік.



У діапазон комірок H82: M89 ввести аналітичну формулу емпіричної функції розподілу (рис.3.8).

Эмпирическая функция распределения			
F*(x)=	0	при	x= 14,8
	0,075	при	x=15,3
	0,2	при	x=15,8
	0,45	при	x=16,3
	0,75	при	x=16,8
	0,875	при	x=17,3
	0,95	при	x=17,8
	1	при	x=18,3

Рис.3.8. Аналітична формула емпіричної функції розподілу

Для побудови емпіричної функції розподілу будують допоміжну таблицю, транспонірую значения емпіричної функції розподілу і її аргументи (рис. 3.9).

Эмпирическая функция распределения													
F*(x)=	0	при	x= 14,8										
	0,075	при	x=15,3										
	0,2	при	x=15,8										
	0,45	при	x=16,3										
	0,75	при	x=16,8										
	0,875	при	x=17,3										
	0,95	при	x=17,8										
	1	при	x=18,3										
Вспомогательная таблица для построения эмпирической функции распределения													
0	0,075	0,2	0,45	0,75	0,875	0,95	1						
x= 14,8	x=15,3	x=15,8	x=16,3	x=16,8	x=17,3	x=17,8	x=18,3						

Рис. 3.9. Допоміжна таблиця для побудови емпіричної функції розподілу
Викликаючи «Мастер Диаграмм», «График», Будують емпіричну функцію розподілу.
Редагують графік таким чином, щоб він був максимально схожий на наведений на рис.3.10.



Рис.3.10. Графік емпіричної функції розподілу

Варіанти індивідуальних завдань.

Побудувати інтервальний статистичний ряд, гістограму, полігон і $F^*(x)$ для випадкової ознаки X - продуктивності мартенівської печі (в т / год), якщо є дані по 40 плавок для кожної з 10 мартенівської печей:

- 1) 16,1; 16,2; 16,4; 15,3; 15,4; 15,0; 15,6; 15,7; 15,6; 15,8
16,6; 16,7; 16,6; 16,9; 16,9; 16,8; 16,7; 17,1; 17,2; 17,6;
17,8; 16,9; 16,7; 16,9; 16,6; 17,8; 16,1; 16,1; 16,2; 15,9;
15,8; 15,9; 16,9; 16,8; 16,6; 16,9; 16,4; 16,4; 16,3; 16,1;
- 2) 18,0; 15,9; 15,8; 15,2; 15,9; 16,1; 16,4; 15,3; 16,3; 16,2
16,9; 16,7; 16,6; 16,8; 17,2; 17,1; 17,4; 15,0; 15,1; 15,4;
17,6; 16,6; 16,5; 15,7; 15,8; 15,6; 15,1; 15,4; 15,5; 15,0;
15,0; 17,4; 17,0; 15,0; 15,4; 15,3; 15,2; 15,3; 15,2; 15,1
- 3) 15,4; 15,0; 15,1; 15,2; 15,5; 15,6; 15,7; 15,7; 16,0; 16,2
16,3; 16,4; 16,5; 16,4; 16,6; 16,9; 16,9; 17,0; 16,7; 17,1
17,3; 17,2; 17,4; 17,5; 17,6; 17,7; 17,8; 17,9; 16,4; 15,0;
15,2; 15,2; 15,4; 16,0; 16,8; 16,7; 17,3; 18,0; 17,6; 17,8.
- 4) 17,9; 17,8; 15,0; 15,1; 15,2; 15,1; 16,3; 16,2; 17,3; 17,1;
16,6; 16,7; 16,6; 15,4; 15,1; 15,2; 15,6; 15,9; 15,9; 17,7;
17,8; 16,3; 16,1; 16,2; 15,9; 15,8; 15,9; 16,9; 16,5; 16,8;
15,1; 15,1; 15,5; 15,5; 16,0; 16,5; 16,5; 17,0; 17,5; 17,5.
- 5) 17,4; 15,0; 15,1; 15,4; 15,3; 15,7; 15,9; 16,0; 16,2; 16,4;
15,2; 16,5; 15,0; 15,2; 15,3; 15,5; 15,3; 15,4; 18,0; 17,6;
17,1; 15,2; 17,5; 16,6; 16,7; 16,9; 15,9; 15,7; 15,2; 15,3;
16,0; 16,4; 16,2; 15,9; 16,0; 15,0; 16,3; 16,2; 15,5; 15,0.
- 6) 15,4; 15,0; 15,2; 15,1; 15,5; 15,6; 15,9; 16,0; 15,7; 15,8;
17,0; 17,7; 17,9; 17,1; 17,2; 16,1; 16,3; 16,2; 16,4; 16,2;
16,8; 16,9; 16,6; 17,0; 17,5; 17,4; 15,4; 15,5; 15,2; 15,9;
15,9; 15,7; 16,4; 16,9; 16,8; 15,0; 15,4; 15,2; 15,9; 17,0
- 7) 15,1; 15,0; 15,7; 15,8; 16,0; 16,4; 16,0; 16,1; 16,5; 16,9;
17,0; 17,4; 17,2; 17,6; 17,9; 17,7; 18,0; 15,0; 16,3; 15,4;
16,1; 16,2; 16,3; 16,3; 17,4; 17,3; 16,7; 16,9; 16,9; 17,9;
17,8; 15,4; 15,3; 17,0; 18,0; 17,8; 17,3; 17,4; 15,5; 18,0.
- 8) 17,1; 17,3; 17,4; 15,0; 15,3; 15,2; 15,4; 15,9; 16,8; 16,7;
16,1; 16,2; 16,3; 16,4; 17,9; 17,6; 16,9; 16,7; 16,7; 17,0;
17,6; 17,7; 17,6; 16,6; 17,0; 16,7; 16,9; 16,4; 16,3; 17,4;
15,8; 15,6; 15,7; 15,9; 16,6; 16,9; 18,8; 17,9; 17,8; 17,4.
- 9) 16,1; 16,0; 15,6; 15,7; 15,9; 15,0; 15,2; 15,1; 18,0; 17,9;
16,9; 16,8; 17,0; 17,1; 15,6; 15,7; 15,8; 15,3; 15,4; 15,3;
16,2; 16,4; 16,3; 16,7; 16,6; 16,9; 17,4; 17,9; 17,7; 16,4;
16,3; 16,2; 17,0; 16,9; 15,4; 15,3; 15,2; 15,7; 15,9; 15,9;
- 10) 16,4; 16,2; 16,3; 16,1; 16,2; 16,9; 16,8; 16,7; 16,5; 16,6;
17,1 17,0; 17,2; 17,9; 15,0; 15,9; 15,7; 15,6; 15,8; 16,1;
16,2; 16,2; 16,3; 16,4; 17,4; 17,3; 16,9; 16,8; 16,6; 16,7;
15,7; 15,6; 16,0; 16,9; 16,7; 16,3; 16,2; 17,0; 16,5; 16,9;
- 11) 16,1; 16,2; 16,4; 15,3; 15,4; 15,0; 15,6; 15,7; 15,6; 15,8
16,6; 16,4; 16,6; 16,8; 16,9; 16,8; 16,7; 17,1; 17,2; 17,6;

17,8; 16,9; 16,7; 16,9; 16,2; 17,8; 16,1; 16,1; 16,2; 15,9;
15,9; 15,9; 16,9; 16,8; 16,6; 16,9; 16,5; 16,4; 16,3; 16,2;

12) 18,1; 15,9; 15,8; 15,3; 15,9; 16,2; 16,4; 15,3; 16,3; 16,2
16,9; 16,7; 16,7; 16,8; 17,2; 17,4; 17,4; 15,0; 15,1; 15,4;
17,6; 16,6; 16,5; 15,7; 15,8; 15,6; 15,1; 15,4; 15,5; 15,0;
15,4; 17,4; 17,2; 15,0; 15,4; 15,3; 15,2; 15,3; 15,2; 15,1

13) 15,6; 15,0; 15,1; 15,5; 15,5; 15,6; 15,7; 15,7; 16,0; 16,2
16,3; 16,4; 16,5; 16,4; 16,6; 16,9; 16,9; 17,0; 16,9; 17,1
17,3; 17,2; 17,4; 17,5; 17,6; 17,7; 17,8; 17,9; 16,4; 15,0;
15,2; 15,2; 15,4; 16,1; 16,8; 16,7; 17,3; 18,2; 17,6; 17,9.

14) 17,4; 17,4; 15,0; 15,1; 15,2; 15,1; 16,3; 16,2; 17,3; 17,1;
16,6; 16,7; 16,6; 15,4; 15,1; 15,2; 15,6; 15,9; 15,9; 17,7;
17,8; 16,3; 16,1; 16,2; 16,9; 15,8; 15,9; 16,9; 16,5; 16,8;
15,1; 15,1; 15,5; 15,5; 16,0; 16,6; 16,5; 17,4; 17,5; 17,5.

15) 17,4; 15,1; 15,1; 15,4; 15,3; 15,7; 15,9; 16,0; 16,2; 16,4;
15,2; 16,5; 15,0; 15,2; 15,3; 15,5; 15,3; 15,4; 18,0; 17,6;
17,1; 15,2; 17,5; 16,6; 16,7; 16,8; 15,9; 15,7; 15,2; 15,3;
16,0; 16,8; 16,2; 15,9; 16,0; 15,0; 16,3; 16,8; 15,5; 15,0.

16) 15,5; 15,0; 15,2; 15,1; 15,5; 15,6; 15,9; 16,0; 15,7; 15,8;
17,0; 17,7; 17,9; 17,1; 17,2; 16,1; 16,3; 16,2; 16,4; 16,1;
16,8; 16,9; 16,1; 17,0; 17,5; 17,4; 15,4; 15,5; 15,2; 15,9;
15,5; 15,7; 16,4; 16,9; 16,8; 15,0; 15,4; 15,2; 15,9; 17,0

17) 15,2; 15,2; 15,7; 15,8; 16,0; 16,4; 16,0; 16,1; 16,5; 16,9;
17,0; 17,4; 17,2; 17,6; 17,9; 17,7; 18,0; 15,0; 16,3; 15,4;
16,1; 16,2; 16,4; 16,4; 17,4; 17,3; 16,4; 16,9; 16,9; 17,9;
17,8; 15,4; 15,3; 17,0; 18,0; 17,8; 17,3; 17,4; 15,5; 18,0.

18) 17,2; 17,3; 17,4; 15,0; 15,3; 15,2; 15,4; 15,9; 16,8; 16,7;
16,1; 16,2; 16,3; 16,4; 17,9; 17,6; 16,9; 16,7; 16,7; 17,0;
17,6; 17,2; 17,6; 16,6; 17,2; 16,9; 16,9; 16,4; 16,3; 17,4;
15,8; 15,7; 15,7; 15,9; 16,6; 16,9; 18,8; 17,9; 17,8; 17,4.

19) 16,1; 16,2; 15,7; 15,7; 15,9; 15,0; 15,2; 15,1; 18,0; 17,9;
16,9; 16,8; 17,0; 17,0; 15,6; 15,7; 15,8; 15,3; 15,4; 15,3;
16,2; 16,4; 16,3; 16,7; 16,6; 16,9; 17,4; 17,9; 17,7; 16,4;
16,3; 16,2; 17,0; 16,8; 15,5; 15,3; 15,5; 15,7; 15,5; 15,9;

20) 16,4; 16,1; 16,3; 16,1; 16,2; 16,9; 16,8; 16,7; 16,5; 16,6;
17,1 17,1; 17,2; 17,9; 15,0; 15,9; 15,7; 15,6; 15,8; 16,1;
16,2; 16,2; 16,1; 16,4; 17,4; 17,3; 16,9; 16,8; 16,6; 16,7;
15,7; 15,6; 16,1; 16,9; 16,7; 16,3; 16,1

21) 16,2; 16,2; 16,4; 15,3; 15,3; 15,0; 15,6; 15,7; 15,6; 15,3
16,6; 16,7; 16,6; 16,9; 16,9; 16,8; 16,7; 17,1; 17,2; 17,6;
17,8; 16,9; 16,7; 16,9; 16,7; 17,8; 16,2; 16,1; 16,2; 15,8;
15,8; 15,9; 16,9; 16,8; 16,7; 16,9; 16,4; 16,4; 16,3; 16,2;

22) 18,3; 15,8; 15,8; 15,2; 15,9; 16,2; 16,4; 15,3; 16,3; 16,2

16,2; 16,7; 16,6; 16,8; 17,2; 17,1; 17,4; 15,0; 15,1; 15,4;
 17,6; 16,6; 16,2; 15,7; 15,8; 15,6; 15,1; 15,4; 15,5; 15,1;
 15,0; 17,4; 17,0; 15,0; 15,4; 15,3; 15,3; 15,3; 15,2; 15,1

23) 15,5; 15,1; 15,1; 15,2; 15,5; 15,6; 15,8; 15,7; 16,0; 16,2
 16,3; 16,4; 16,5; 16,5; 16,6; 16,9; 16,9; 17,0; 16,7; 17,1
 17,3; 17,2; 17,4; 17,5; 17,6; 17,7; 17,8; 17,9; 16,4; 15,0;
 15,2; 15,2; 15,5; 16,0; 16,8; 16,7; 17,2; 18,0; 17,6; 17,2.

24) 17,9; 17,9; 15,0; 15,2; 15,2; 15,1; 16,3; 16,2; 17,3; 17,1;
 16,6; 16,7; 16,6; 15,4; 15,2; 15,2; 15,6; 15,9; 15,9; 17,7;
 17,8; 16,3; 16,1; 16,3; 15,9; 15,8; 15,9; 16,3; 16,5; 16,8;
 15,1; 15,2; 15,5; 15,5; 16,0; 16,5; 16,3; 17,0; 17,5; 17,5.

25) 17,3; 15,1; 15,1; 15,4; 15,1; 15,7; 15,9; 16,0; 16,2; 16,4;
 15,2; 16,5; 15,0; 15,2; 15,3; 15,5; 15,1; 15,4; 18,0; 17,6;
 17,1; 15,2; 17,5; 16,6; 16,7; 16,9; 15,1; 15,7; 15,2; 15,3;
 16,0; 16,4; 16,2; 15,9; 16,0; 15,0; 16,3; 16,2; 15,1; 15,0.

26) 15,4; 15,2; 15,2; 15,1; 15,5; 15,6; 15,9; 16,0; 15,7; 15,8;
 17,0; 17,7; 17,9; 17,2; 17,2; 16,1; 16,3; 16,2; 16,4; 16,2;
 16,8; 16,9; 16,6; 17,0; 17,5; 17,4; 15,4; 15,5; 15,2; 15,9;
 15,9; 15,7; 16,4; 16,9; 16,9; 15,0; 15,4; 15,2; 15,9; 17,0

27) 15,0; 15,0; 15,7; 15,8; 16,0; 16,4; 16,1; 16,1; 16,1; 16,9;
 17,0; 17,4; 17,2; 17,6; 17,9; 17,7; 18,1; 15,0; 16,3; 15,4;
 16,1; 16,2; 16,3; 16,3; 17,4; 17,3; 16,7; 16,9; 16,9; 17,9;
 17,8; 15,4; 15,3; 17,1; 18,1; 17,8; 17,1; 17,4; 15,5; 18,2.

28) 17,1; 17,3; 17,2; 15,0; 15,3; 15,2; 15,4; 15,9; 16,8; 16,7;
 16,1; 16,2; 16,4; 16,4; 17,2; 17,6; 16,9; 16,7; 16,7; 17,0;
 17,6; 17,7; 17,6; 16,7; 17,0; 16,7; 16,9; 16,3; 16,3; 17,4;
 15,8; 15,8; 15,7; 15,7; 16,6; 16,9; 18,8; 17,9; 17,8; 17,4.

29) 16,2; 15,5; 15,6; 15,6; 15,9; 15,1; 15,2; 15,1; 18,0; 17,9;
 16,9; 16,8; 17,0; 17,1; 15,6; 15,7; 15,8; 15,1; 15,4; 15,3;
 16,2; 16,3; 16,3; 16,7; 16,6; 16,9; 17,4; 17,9; 17,7; 16,4;
 16,3; 16,2; 17,0; 16,9; 15,4; 15,1; 15,2; 15,7; 15,1; 15,9;

30) 16,2; 16,2; 16,3; 16,2; 16,2; 16,9; 16,8; 16,7; 16,5; 16,6;
 17,1 17,0; 17,2; 17,9; 15,0; 15,9; 15,7; 15,6; 15,8; 16,1;
 16,2; 16,2; 16,3; 16,4; 17,4; 17,3; 16,3; 16,6; 16,6; 16,7;
 15,7; 15,6; 16,0; 16,9; 16,7; 16,3; 16,6

31) 16,2; 16,2; 16,4; 15,4; 15,4; 15,0; 15,6; 15,7; 15,6; 15,8
 16,6; 16,7; 16,6; 16,9; 16,9; 16,8; 16,7; 17,1; 17,2; 17,6;
 17,8; 16,9; 16,7; 16,7; 16,6; 17,8; 16,1; 16,7; 16,2; 15,9;
 15,8; 15,9; 16,9; 16,8; 16,6; 16,9; 16,4; 16,4; 16,7; 16,1;

32) 18,1; 15,8; 15,8; 15,2; 15,9; 16,1; 16,4; 15,8; 16,3; 16,2
 16,9; 16,7; 16,6; 16,8; 17,2; 17,1; 17,4; 15,0; 15,1; 15,4;
 17,6; 16,6; 16,5; 15,7; 15,8; 15,6; 15,1; 15,4; 15,5; 15,0;
 15,0; 17,4; 17,0; 15,0; 15,4; 15,3; 15,2; 15,3; 15,8; 15,2

33) 15,5; 15,5; 15,1; 15,2; 15,5; 15,6; 15,7; 15,7; 16,3; 16,2
 16,3; 16,4; 16,5; 16,4; 16,6; 16,9; 16,9; 17,0; 16,7; 17,1

17,3; 17,2; 17,4; 17,5; 17,6; 17,7; 17,8; 17,9; 16,4; 15,0;
15,2; 15,2; 15,4; 16,3; 16,8; 16,3; 17,3; 18,0; 17,6; 17,8.

34) 17,8; 17,8; 15,1; 15,1; 15,2; 15,1; 16,3; 16,2; 17,3; 17,1;
16,6; 16,7; 16,6; 15,4; 15,1; 15,2; 15,6; 15,9; 15,9; 17,7;
17,8; 16,3; 16,1; 16,2; 15,1; 15,8; 15,1; 16,9; 16,5; 16,8;
15,1; 15,1; 15,5; 15,5; 16,0; 16,5; 16,5; 17,7; 17,5; 17,5.

35) 17,4; 15,1; 15,1; 15,4; 15,3; 15,7; 15,9; 16,0; 16,2; 16,4;
15,2; 16,5; 15,3; 15,2; 15,3; 15,5; 15,3; 15,4; 18,0; 17,6;
17,1; 15,2; 17,5; 16,7; 16,7; 16,9; 15,9; 15,7; 15,2; 15,3;
16,7; 16,4; 16,2; 15,9; 16,0; 15,1; 16,3; 16,2; 15,5; 15,0.

36) 15,4; 15,4; 15,2; 15,1; 15,5; 15,4; 15,9; 16,0; 15,7; 15,8;
17,0; 17,7; 17,9; 17,1; 17,2; 16,1; 16,3; 16,2; 16,4; 16,2;
16,8; 16,9; 16,6; 17,0; 17,5; 17,4; 15,4; 15,5; 15,2; 15,9;
15,9; 15,7; 16,4; 16,9; 16,8; 15,0; 15,4; 15,4; 15,9; 17,1

37) 15,0; 15,0; 15,7; 15,7; 16,0; 16,4; 16,0; 16,1; 16,5; 16,9;
17,0; 17,4; 17,2; 17,6; 17,9; 17,7; 18,0; 15,7; 16,3; 15,4;
16,1; 16,2; 16,3; 16,3; 17,4; 17,3; 16,7; 16,9; 16,9; 17,9;
17,8; 15,4; 15,7; 17,0; 18,0; 17,8; 17,3; 17,5; 15,5; 18,0.

38) 17,2; 17,3; 17,4; 15,0; 15,3; 15,2; 15,4; 15,9; 16,8; 16,7;
16,1; 16,2; 16,4; 16,4; 17,9; 17,6; 16,9; 16,7; 16,7; 17,0;
17,6; 17,7; 17,6; 16,4; 17,2; 16,7; 16,9; 16,4; 16,3; 17,4;
15,8; 15,6; 15,7; 15,9; 16,6; 16,9; 18,8; 17,9; 17,8; 17,4.

39) 16,1; 15,2; 15,6; 15,7; 15,9; 15,0; 15,2; 15,1; 18,0; 17,9;
16,9; 16,8; 17,0; 17,1; 15,2; 15,7; 15,2; 15,3; 15,4; 15,3;
16,2; 16,4; 16,3; 16,7; 16,6; 16,9; 17,7; 17,9; 17,7; 16,4;
16,3; 16,2; 17,7; 16,9; 15,4; 15,3; 15,2; 15,7; 15,9; 15,9;

40) 16,4; 16,4; 16,3; 16,1; 16,2; 16,9; 16,8; 16,7; 16,5; 16,6;
17,1 17,0; 17,2; 17,2; 15,0; 15,9; 15,7; 15,6; 15,8; 16,1;
16,2; 16,2; 16,3; 16,4; 17,4; 17,3; 16,9; 16,8; 16,6; 16,7;
15,7; 15,6; 16,7; 16,9; 16,7; 16,3; 16,7

41) 16,3; 16,2; 16,4; 15,3; 15,4; 15,0; 15,6; 15,7; 15,6; 15,8
16,6; 16,7; 16,6; 16,9; 16,3; 16,8; 16,7; 17,1; 17,2; 17,6;
17,8; 16,9; 16,7; 16,9; 16,6; 17,8; 16,1; 16,1; 16,2; 15,9;
15,8; 15,9; 16,3; 16,8; 16,6; 16,3; 16,4; 16,4; 16,3; 16,1;

42) 18,0; 15,9; 15,6; 18,0; 15,9; 16,1; 16,4; 15,3; 16,3; 16,2
16,9; 16,7; 16,6; 16,8; 17,2; 17,1; 17,4; 15,0; 15,1; 15,4;
17,6; 16,6; 16,5; 15,7; 15,8; 15,6; 15,6; 15,4; 15,5; 15,0;
15,0; 17,0; 17,0; 15,0; 15,4; 15,3; 15,2; 15,6; 15,2; 15,1

43) 15,4; 15,2; 15,1; 15,2; 15,5; 15,6; 15,7; 15,7; 16,0; 16,2
16,3; 16,4; 16,5; 16,4; 16,6; 16,9; 16,9; 17,0; 16,7; 17,1
17,3; 17,2; 17,4; 17,5; 17,6; 17,8; 17,8; 17,9; 16,4; 15,0;
15,2; 15,2; 15,2; 16,0; 16,8; 16,7; 17,3; 18,0; 17,8; 17,8.

44) 17,9; 17,9; 15,1; 15,1; 15,2; 15,1; 16,3; 16,2; 17,3; 17,1;

16,6; 16,7; 16,6; 15,3; 15,1; 15,3; 15,6; 15,9; 15,9; 17,7;
 17,8; 16,3; 16,1; 16,2; 15,9; 15,8; 15,9; 16,9; 16,5; 16,8;
 15,3; 15,1; 15,5; 15,3; 16,0; 16,5; 16,5; 17,0; 17,5; 17,5.

45) 17,2; 15,0; 15,1; 15,4; 15,3; 15,7; 15,9; 16,0; 16,2; 16,4;
 15,2; 16,5; 15,0; 15,2; 15,3; 15,5; 15,3; 15,4; 18,0; 17,6;
 17,2; 15,2; 17,2; 16,6; 16,7; 16,9; 15,9; 15,7; 15,2; 15,3;
 16,2; 16,4; 16,2; 15,9; 16,2; 15,0; 16,3; 16,2; 15,5; 15,0.

46) 15,4; 15,4; 15,2; 15,1; 15,5; 15,6; 15,9; 16,0; 15,7; 15,8;
 17,7; 17,7; 17,9; 17,1; 17,2; 16,1; 16,3; 16,2; 16,4; 16,2;
 16,8; 16,9; 16,6; 17,7; 17,5; 17,4; 15,4; 15,2; 15,2; 15,9;
 15,9; 15,4; 16,4; 16,9; 16,8; 15,7; 15,4; 15,3; 15,7; 17,7

47) 15,1; 15,3; 15,7; 15,8; 16,0; 16,4; 16,0; 16,1; 16,5; 16,9;
 17,0; 17,4; 17,2; 17,6; 17,9; 17,7; 18,0; 15,0; 16,3; 15,4;
 16,1; 16,2; 16,3; 16,3; 17,4; 17,3; 16,7; 16,9; 16,9; 17,9;
 17,8; 15,3; 15,3; 17,0; 18,1; 17,8; 17,3; 17,4; 15,3; 18,1.

48) 17,1; 17,4; 17,4; 15,0; 15,3; 15,2; 15,4; 15,9; 16,8; 16,7;
 16,1; 16,2; 16,3; 16,4; 17,9; 17,6; 16,9; 16,7; 16,7; 17,0;
 17,6; 17,7; 17,4; 16,6; 17,4; 16,7; 16,9; 16,4; 16,3; 17,4;
 15,8; 15,6; 15,7; 15,8; 16,6; 16,9; 18,4; 17,9; 17,8; 17,4.

49) 16,1; 15,3; 15,6; 15,7; 15,3; 15,0; 15,2; 15,1; 18,0; 17,9;
 16,9; 16,8; 17,0; 17,1; 15,6; 15,7; 15,8; 15,3; 15,4; 15,3;
 16,2; 16,4; 16,3; 16,7; 16,6; 16,9; 17,4; 17,9; 17,7; 16,4;
 16,3; 16,2; 17,0; 16,9; 15,4; 15,3; 15,3; 15,7; 15,9; 15,3;

50) 16,2; 16,2; 16,3; 16,1; 16,2; 16,9; 16,8; 16,7; 16,5; 16,6;
 17,1 17,0; 17,2; 17,9; 15,0; 15,9; 15,7; 15,6; 15,8; 16,1;
 16,2; 16,2; 16,3; 16,4; 17,4; 17,3; 16,2; 16,8; 16,2; 16,7;
 15,7; 15,6; 16,0; 16,9; 16,7; 16,3; 16,2

Статистичне оцінювання параметрів

Короткий теоретичний матеріал

Нехай X - випадкова ознака, підлеглий закону розподілу $F(x, \theta)$, де θ - параметр розподілу, числове значення якого невідомо. Істинне значення цього параметра можна визначити, дослідивши всі елементи генеральної сукупності, що не представляється можливим. За вибірковими даними можна визначити тільки наближене значення параметра, яке називають значенням оцінки параметра. Завдання статистичного оцінювання параметрів полягає у виборі такої оцінюючої функції, значення якої найближче до оцінюваного параметру. Таким чином, оцінка параметра - це однозначно визначена функція вибіркових даних (оціночна функція), а значення оцінки параметра - це значення цієї функції для конкретної вибірки. Оцінка параметра θ позначається $\tilde{\theta}$.

Така оцінка, значення якої визначається одним числом, називається точковою. Оцінка, значення якої визначається двома числами (кінцями інтервалу), називається *інтервальною*.

Точкова оцінка вважається найкращою, якщо вона:

- незміщена, тобто її математичне очікування дорівнює оцінюваному параметру, а саме

$$M(\tilde{\theta}) = \theta,$$

- спроможна, якщо вона задовольняє закону великих чисел, тобто сходиться по ймовірності до оцінюваного параметру:

$$\lim_{n \rightarrow \infty} P(\tilde{\theta}_n - \theta) < \varepsilon \rightarrow 1.$$

Якщо оцінка спроможна, то практично достовірно, що при великому обсязі вибірки (n) $\tilde{\theta} \approx \theta$.

- ефективна, тобто якщо вона має найменшу серед усіх незміщених оцінок дисперсію:

$$D(\tilde{\theta}) \rightarrow \min.$$

Це означає, що ефективна оцінка має мінімальне розсіювання щодо істинного значення параметра.

Точкове оцінювання основних числових характеристик випадкових величин

Найбільш загальні числові характеристики випадкової величини - це статистичні моменти, які визначаються наступним чином:

- початковий момент r - го порядку

$$\alpha_r = \bar{x}^r = \frac{\sum_{i=1}^n x_i^r}{n} \quad (\text{або} \quad \frac{\sum_{i=1}^K m_i \cdot x_i^r}{n})$$

- центральний момент r - го порядку

$$\mu_r = \frac{\sum_{i=1}^n (x_i - \bar{x})^r}{n} \quad (\text{або} \quad \frac{\sum_{i=1}^K m_i (x_i - \bar{x})^r}{n})$$

де x_i - значення випадкової величини в i - м спостереженні; n - число спостережень.

Звідси отримують оцінки для математичного очікування і дисперсії випадкової величини

$$\tilde{m}_x = \bar{x} = \begin{cases} \frac{1}{n} \sum_{i=1}^n x_i, & \text{якщо } x_1, x_2, \dots, x_n, \\ \frac{1}{n} \sum_{i=1}^k m_i x_i, & \text{якщо } \left| \begin{array}{c} x_i \ x_1 \ x_2 \ \dots \ x_k \\ m_i \ m_1 \ m_2 \ \dots \ m_k \end{array} \right|, \sum_{i=1}^k m_i = n \\ \frac{1}{n} \sum_{i=1}^k m_i \bar{x}_i, & \text{якщо } \left| \begin{array}{c} J_i [x^{(1)}, x^{(2)}] \dots [x^{(k)}, x^{(k+1)}] \\ m_i \quad m_1 \quad \dots \quad m_k \end{array} \right| \end{cases}$$

Вибіркова дисперсія визначається так:

$$D_B = \begin{cases} \frac{1}{n} \sum_{i=1}^n (x_i - \tilde{m}_x)^2, & \text{якщо дані } x_1, x_2, \dots, x_n, \\ \frac{1}{n} \sum_{i=1}^k m_i (x_i - \tilde{m}_x)^2, & \text{якщо дані } \left| \begin{array}{c} x_i \ x_1 \ x_2 \ \dots \ x_k \\ m_i \ m_1 \ m_2 \ \dots \ m_k \end{array} \right|, \\ \frac{1}{n} \sum_{i=1}^k m_i (\bar{x}_i - \tilde{m}_x)^2, & \text{якщо дані } \left| \begin{array}{c} J_i [x^{(1)}, x^{(2)}] \dots [x^{(k)}, x^{(k+1)}] \\ m_i \quad m_1 \quad \dots \quad m_k \end{array} \right| \end{cases},$$

де $\bar{x}_i$ – середина i – го інтервалу.

Вибіркова дисперсія не є незміщеною оцінкою дисперсії генеральної сукупності, тому в якості оцінки $\tilde{D}_x$ береться виправлена вибірка дисперсія, тобто

$$\tilde{D}_x = \frac{n}{n-1} D_B.$$

При $n \geq 50$ можна брати $\tilde{D}_x = D_B$.

Оцінки параметрів розподілів:

- нормального $\tilde{m} = \tilde{m}_x; \tilde{\delta} = \sqrt{\tilde{D}_x}$;
- показового $\tilde{\lambda} = 1/\tilde{m}_x$;
- рівномірного $\tilde{a} = \tilde{m}_x - \sqrt{3} \cdot \sqrt{\tilde{D}_x}; \tilde{b} = \tilde{m}_x + \sqrt{3} \cdot \sqrt{\tilde{D}_x}$;

Інтервальне оцінювання основних числових характеристик випадкових величин

Знайти точне значення оцінок математичного очікування, дисперсії і стандартного відхилення за дослідними даними в принципі неможливо, так як в вибірковій сукупності отримують тільки частину інформації про випадкову величину.

Коли замість математичного очікування з досвіду визначають вибірку середню, то допускають похибку. Оцінити її можна за допомогою довірчого інтервалу $(\tilde{m}_x - \varepsilon; \tilde{m}_x + \varepsilon)$ і довірчої ймовірності β – ймовірність того, що істинне значення математичного очікування лежить в цьому інтервалі.

Інтервальною називають оцінку, яка визначається двома числами – кінцями інтервалу, що містить оцінюваний параметр.

Довірчим називають інтервал, який з заданою довірчою ймовірністю β містить істинне значення оцінюваного параметра. Для інтервальної оцінки математичного очікування нормально розподіленої генеральної сукупності служить довірчий інтервал, який визначається наступним чином:

а) при відомому середньому квадратичному відхиленні σ генеральної сукупності

$$\bar{X} - t_{\beta} \frac{\sigma}{\sqrt{n}} < m_x < \bar{X} + t_{\beta} \frac{\sigma}{\sqrt{n}},$$

де t_{β} – таке значення аргументу функції Лапласа $\Phi(X)$, при якому

$$\Phi\left(t_{\beta}\right) = \frac{\beta}{2}, \text{ тобто } t_{\beta} = \arg \Phi\left(\frac{\beta}{2}\right); \text{ величина } t_{\beta} \cdot \frac{\sigma}{\sqrt{n}} = \varepsilon \text{ називається точністю оцінки;}$$

б) при невідомим σ :

$$\bar{X} - t_{\beta, n} \frac{\bar{\sigma}}{\sqrt{n}} < \tilde{m}_x < \bar{X} + t_{\beta, n} \frac{\bar{\sigma}}{\sqrt{n}},$$

де $\bar{\sigma}$ – незміщена оцінка параметра σ , знайдена по виборці; $t_{\beta, n}$ – знаходиться по таблиці розподілу Стьюдента (див. додаток) для двосторонньої критичної області при $\alpha = 1 - \beta$ и $f = n - 1$.

Довірчий інтервал для дисперсії визначається нерівністю:

$$\frac{n\tilde{D}_x}{\chi_2^2} < \tilde{D}_x < \frac{n\tilde{D}_x}{\chi_1^2},$$

де χ_1^2 – значення випадкової величини, що має розподіл χ^2 (див. додаток) з числом ступенів свободи $f_1 = n - 1$ при рівні значимості $\alpha_1 = 1 - \frac{\alpha}{2}$; де $\alpha = 1 - \beta$;

χ_2^2 – значення випадкової величини, що має розподіл χ^2 з числом ступенів свободи $f_2 = n - 1$, при рівні значимості $\alpha_2 = \frac{\alpha}{2}$.

Значення χ_1^2 и χ_2^2 по заданим β і n знаходять по таблиці розподілу χ^2 . Значення χ_1^2 и χ_2^2 іноді називають критичними точками або критичними значеннями розподілу χ^2 для α_1 і α_2 та відомому числі ступенів свободи розподілу.

ЗАВДАННЯ:

За результатами обробки дискретної (завдання 2) і безперервної (завдання 3) випадкової ознаки знайти:

1. Числові характеристики вибіркової сукупності дискретної випадкової ознаки по відомим аналітичним співвідношенням.
2. Числові характеристики вибіркової сукупності дискретної випадкової ознаки за стандартними формулами середовища EXCEL.
3. Числові характеристики вибіркової сукупності безперервної випадкової ознаки по відомим аналітичним співвідношенням.
4. Числові характеристики вибіркової сукупності безперервної випадкової ознаки за стандартними формулами середовища EXCEL.
5. Довірчі інтервали для оцінки математичного очікування і дисперсії безперервної випадкової ознаки.

Приклад 1.

X – число працюючих завалочних машин в цеху.

Дискретний варіаційний ряд (із завдання №2):

x_i	0	1	2	3
m_i	1	4	3	2

За результатами обробки дискретної (завдання 2) випадкової ознаки знайти:

1. Числові характеристики вибірковій сукупності дискретної випадкової ознаки по відомим аналітичним співвідношенням

Визначимо об'єм виробки: $n = \sum_{i=1}^k m_i = 10$

Середнє число працюючих завалочних машин в цеху

$$\bar{m}_x = \frac{\sum_{i=1}^k m_i x_i}{n} = \frac{1 \cdot 0 + 4 \cdot 1 + 3 \cdot 2 + 2 \cdot 3}{10} = 1.6$$

Відхилення від середнього значення

$$\begin{aligned} \tilde{D}_x &= \frac{n}{n-1} D_g = \frac{n}{n-1} \cdot \frac{\sum_{i=1}^k m_i (x_i - \bar{m}_i)^2}{n} = \frac{\sum_{i=1}^k m_i (x_i - \tilde{m}_x)^2}{n-1} = \\ &= \frac{1 \cdot (0 - 1.6)^2 + 4 \cdot (1 - 1.6)^2 + 3 \cdot (2 - 1.6)^2 + 2 \cdot (3 - 1.6)^2}{10 - 1} \approx 0.93; \end{aligned}$$

$$\bar{\sigma}_x = \sqrt{\tilde{D}_x} = \sqrt{0.93} \approx 0.96$$

Приклад 2.

X – продуктивність мартенівської печі.

Інтервальний варіаційний ряд (із завдання №2):

$\tilde{x}_i$	15	15.4	15.8	16.2	16.6	17	17.4	17.8
I_i	[14,8;15,2)	[15,2;15,6)	[15,6;16,0)	[16,0;16,4)	[16,4;16,8)	[16,8;17,2)	[17,2;17,6)	[17,6;18,0)
m_i	3	4	7	10	8	3	3	2

За результатами обробки безперервної (завдання 3) випадкової ознаки знайти:

1. Числові характеристики вибірковій сукупності безперервної випадкової ознаки по відомим аналітичним співвідношенням

2. Довірчі інтервали для оцінки математичного очікування і дисперсії безперервної випадкової ознаки.

Визначимо середину кожного інтервалу і запишемо над кожним інтервалом, наприклад,

$$\tilde{x}_1 = \frac{x^{(1)} + x^{(2)}}{2} = \frac{14.8 + 15.2}{2} = 15, \text{ або}$$

$$\tilde{x}_1 = x^{(1)} + \frac{\Delta x}{2} = 14.8 + \frac{0.4}{2} = 14.8 + 0.2 = 15$$

Визначимо об'єм виробки: $T = \sum_{i=1}^k m_i = 40$

Середня продуктивність мартенівської печі

$$\tilde{m}_x = \frac{\sum_{i=1}^k m_i \tilde{x}_i}{n} = \frac{3 \cdot 15 + 4 \cdot 15.4 + 7 \cdot 15.8 + 10 \cdot 16.2 + 8 \cdot 16.6 + 3 \cdot 17.4 + 2 \cdot 17}{40} = 16.27$$

Відхилення від середнього значення

$$\begin{aligned} \tilde{D}_x &= \frac{\sum_{i=1}^k m_i (\tilde{x}_i - \tilde{m}_x)^2}{n-1} = \frac{3 \cdot (15 - 16.27)^2 + 4 \cdot (15.4 - 16.27)^2 + 7 \cdot (15.8 - 16.27)^2 +}{39} \\ &+ \frac{10 \cdot (16.2 - 16.27)^2 + 8 \cdot (16.6 - 16.27)^2 + 3 \cdot (17.4 - 16.27)^2 + 2 \cdot (17.8 - 16.27)^2}{39} = \\ &= 0.52; \end{aligned}$$

$$\bar{\sigma}_x = \sqrt{\tilde{D}_x} = \sqrt{0.52} = 0.72$$

Щоб побудувати довірчий інтервал для m_x , необхідно знати $\tilde{m}_x$ і ε :

$$\tilde{m}_x = 16.27 \text{ (из п.2)}$$

$$t_{\beta, n} = t_{\text{табл. двуст}}(\alpha = 1 - 0.95 = 0.05; f = 40 - 1 = 39) = 2.02$$

$$\varepsilon = 2.02 \cdot 0.72 / \sqrt{40} = 0.23$$

Довірчий інтервал для m_x :

$$(\tilde{m}_x - \varepsilon; \tilde{m}_x + \varepsilon);$$

$$(16.27 - 0.23; 16.27 + 0.23);$$

$$I_{m_x} = (16.04; 16.5)$$

Визначимо довірчий інтервал для $\tilde{D}_x$

$$\frac{(n-1)\tilde{D}_x}{\chi_2^2} < \tilde{D}_x < \frac{(n-1)\tilde{D}_x}{\chi_1^2}$$

$$\chi_1^2 = \chi_{\text{табл}}^2(\alpha_1 = 1 - \frac{1-0.95}{2} = 0.975; f_1 = 40 - 1 = 39) = 23.6;$$

$$\chi_2^2 = \chi_{\text{табл}}^2(\alpha_2 = \frac{1-0.95}{2} = 0.025; f_2 = 40 - 1 = 39) = 58.1;$$

$$\frac{(40-1) \cdot 0,52}{58,1} < \tilde{D}_x < \frac{(40-1) \cdot 0,52}{23,6};$$

$$0,33 < \tilde{D}_x < 0,86; I_{D_x} = (0,33; 0,86).$$

Розв'язання задачі в середовищі EXCEL:

ЗАВДАННЯ:

За результатами обробки дискретної (завдання 2) випадкової ознаки знайти:

1. Числові характеристики вибіркової сукупності дискретної випадкової ознаки по відомим аналітичним співвідношенням



- до робочої області Лист1 копіюють інформацію за завданням 2 і перейменовують його на DSP
- до робочої області Лист2 копіюють інформацію за завданням 3 і перейменовують його на NSP.
- лист3 перейменовують в СНС і подальші операції виконують в робочій області даного листа.
- у комірці B6, B8, B10, B12 вводять найменування точкових характеристик вибіркової сукупності.
- у комірку C6 вводять формулу для визначення оцінки математичного очікування:

$$=СУММ(DSP!выборка)/DSP!D5$$

Вибіркову дисперсію визначають в комірці C8 за формулою:

$$=КВАДРОТКЛ(DSP!выборка)/DSP!D5$$

- Виправлену оцінку дисперсії визначають в комірці C10 за формулою:

$$=DSP!D5/(DSP!D5-1)*CHC!C8$$

- Оцінку стандартного відхилення визначають за формулою:

$$=КОРЕНЬ(C10)$$

За результатами обробки дискретної (завдання 2) випадкової ознаки знайти:

1. Числові характеристики вибіркової сукупності дискретної випадкової ознаки за стандартними формулами середовища EXCEL.

- у комірки G6, G10, G12 вводять найменування точкових характеристики вибіркової сукупності.

- у комірку Н6 вводять формулу для визначення оцінки математичного очікування:

$$=CP3H4Ч(DSP!выборка)$$

- вибірккову дисперсію визначають в комірці Н8 за формулою:

$$=ДИСП(DSP!выборка)$$

- оцінку стандартного відхилення визначають за формулою:

$$=СТАНДОТКЛОН(DSP!выборка)$$

При цьому робоча область середовища EXCEL має вигляд, представлений на рис.4.1.

Точечные оценки дискретного случайного признака	
Вычисленные по:	
аналитическим соотношениям	стандартным формулам среды
$\tilde{m}_x = 1,6$	$\tilde{m}_x = 1,6$
$\tilde{D}_x = 0,84$	
$\tilde{\sigma}_x = 0,93$	$\tilde{\sigma}_x = 0,93$
$\tilde{\sigma} = 0,97$	$\tilde{\sigma} = 0,97$

Рис.4.1. Визначення точкових оцінок вибіркової сукупності дискретної випадкової ознаки

За результатами обробки неперервної (завдання 3) випадкової ознаки знайти:

1. Числові характеристики вибіркової сукупності неперервної випадкової ознаки за відомим аналітичним співвідношенням.



- у комірки B23, B25, B27, B29 вводять найменування точкових характеристик вибіркової сукупності.

- у комірку C23 вводять формулу для визначення оцінки математичного очікування:

$$=(СУММПРОИЗВ(NSP!H45:N45;NSP!H46:N46))/NSP!E5$$

або

$$=СУММПРОИЗВ(NSP!H45:N45;NSP!H47:N47)$$

- оцінку вибіркової дисперсії визначають в комірці C25 за формулою:

$$=(\text{КВАДРОТКЛ}(\text{выборка_1}))/\text{NSP!E5}$$

- виправлену оцінку дисперсії визначають в комірці C27 за формулою:

$$=\text{NSP!E5}/(\text{NSP!E5}-1)*\text{CHC!C25}$$

- оцінку стандартного відхилення визначають за формулою

$$=\text{КОРЕНЬ}(\text{C27})$$

За результатами обробки неперервної (завдання 3) випадкової ознаки знайти:

1. Числові характеристики вибіркової сукупності неперервної випадкової ознаки за відомим аналітичним співвідношенням.



- у комірки G23, G27, G29 вводять найменування точкових характеристик вибіркової сукупності.

- у комірку H23 вводять формулу для визначення оцінки математичного очікування:

$$=\text{СРЗНАЧ}(\text{NSP!B5:B44})$$

- вибірккову дисперсію визначають в комірці H27 за формулою:

$$=\text{ДИСП}(\text{выборка_1})$$

- оцінку стандартного відхилення визначають в комірці H29 за формулою:

$$=\text{СТАНДОТКЛОН}(\text{выборка_1})$$

При цьому робоча область середовища EXCEL має вигляд, представлений на рис.4.2.

Корень	А	В	С	Д	Е	Ф	Г	Н	И	Ж	К	Л	М	Н
16														
17		Точечные оценки непрерывного случайного признака												
18														
19		Вычисленные по:												
20		аналитическим соотношениям						стандартным формулам среды						
21														
22														
23		M ₀ =		16,29				M ₀ =		16,35				
24		D ₀ =		0,52				D ₀ =		0,53				
25		σ ₀ =		0,53				σ ₀ =		0,53				
26		σ ₀ =		0,53				σ ₀ =		0,53				
27		σ ₀ =		0,53				σ ₀ =		0,53				
28		σ ₀ =		0,53				σ ₀ =		0,53				
29		σ ₀ =		0,73				σ ₀ =		орка_1)				
30														
31														
32														
33														
34														
35														
36														
37														

Рис.4.2 Визначення точкових оцінок вибіркової сукупності неперервної випадкової ознаки

ЗАВДАННЯ:

За результатами обробки неперервної (завдання 3) випадкової ознаки знайти:

1. Довірчі інтервали для оцінки математичного очікування і дисперсії неперервної випадкової ознаки.



- у поточному листі сформувати таблиці для визначення числа ступенів свободи, довірчих ймовірностей, t – статистики Стюдента і значення ε (рис.4.3).
- у комірці C40 визначити число ступенів свободи за формулою:

$$=NSP!E5-1$$

- у комірку F40 ввести значення довірчої ймовірності для довірчого інтервалу,
- у комірці F41 обчислити значення довірчої ймовірності для t – статистики Стюдента.
- по таблиці критичних точок розподілу Стюдента з числом ступенів свободи (0,05; 39) визначити значення t – статистики Стюдента і внести його до комірки L40.
- у комірці I40 обчислити значення параметра ε за формулою:

$$=L40*H29/КОРЕНЬ(NSP!E5)$$

- у комірки F47 і H47 ввести формули для визначення довірчого інтервалу для оцінки математичного очікування:

$$=C23-I40 \quad =C23+I40$$

Microsoft Excel - Задача_4

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

Л55

A	B	C	D	E	F	G	H	I	J	K	L	M
31												
32												
33												
34												
35												
36												
37												
38												
39												
40												
41												
42												
43												
44												
45												
46												
47												
48												
49												
50												
51												
52												

Доверительные интервалы

для оценки математического ожидания

число степеней свободы	Доверительная вероятность	Значение ϵ	t статистика Стьюдента
$\nu =$ 39	$\beta =$ 0,95	$\epsilon =$ 0,23	$t_{\beta, \nu} =$ 2,02
	$\alpha =$ 0,05		

Доверительный интервал

16,05	$\leq \tilde{m}_k \leq$	16,52
-------	-------------------------	-------

Рис.4.3. Довірчі інтервали для оцінки математичного очікування

- у поточному листі сформувати таблиці для визначення числа ступенів свободи, рівнів значимості, критерію Пірсона (рис.4.4).
- у комірках C57, C58 визначити число ступенів свободи за формулою:

$$=NSP!E5-1$$

- у комірки F57 і F58 ввести значення довірчих ймовірностей для довірчого інтервалу і t -статистики Стюдента.

- у комірки I57 і I58 ввести значення довірчих ймовірностей для критерію Пірсона α_1, α_2 за формулами:

$$=1-F58/2 \quad =F58/2$$

- по таблиці критичних точок розподілу Пірсона з числом ступенів свободи (0,975; 39) і (0,025; 39) визначають значення χ^2 розподілу і вносять до комірок I57 і I58.

- у комірки F64 і H64 ввести формули для визначення довірчого інтервалу для оцінки дисперсії:

$$=C57*C27/L58 \quad =C57*C27/L57$$

Доверительные интервалы для оценки дисперсии									
число степеней свободы		Уровень значимости		Уровень значимости		Критерий Пирсона			
f1=	39	β =	0,95	α_1 =	0,975	χ^2_1 =	23,65		
f2=	39	α =	0,05	α_2 =	0,025	χ^2_2 =	58,12		
Доверительный интервал									
		0,36		$\leq \hat{D}_x \leq$		0,88			

Рис.4.4. Довірчі інтервали для оцінки дисперсії

Значення t -критерію Стюдента при рівні значущості

(0,01; 0,05; 0,1; 0,15; 0,20; 0,25; 0,30)

Таблиця 4.1

n	P - 0,01	P - 0,05	P - 0,1	P - 0,15	P - 0,2	P - 0,25	P - 0,3
1	63,66	12,71	6,31	4,17	3,08	2,41	1,96
2	9,92	4,30	2,92	2,28	1,89	1,60	1,39
3	5,84	3,18	2,35	1,92	1,64	1,42	1,25
4	4,60	2,78	2,13	1,78	1,53	1,34	1,19
5	4,03	2,57	2,02	1,70	1,48	1,30	1,16
6	3,71	2,45	1,94	1,65	1,44	1,27	1,13
7	3,50	2,36	1,89	1,62	1,41	1,25	1,12
8	3,36	2,31	1,86	1,59	1,40	1,24	1,11
9	3,25	2,26	1,83	1,57	1,38	1,23	1,10
10	3,17	2,23	1,81	1,56	1,37	1,22	1,09
11	3,11	2,20	1,80	1,55	1,36	1,21	1,09

12	3,05	2,18	1,78	1,54	1,36	1,21	1,08
13	3,01	2,16	1,77	1,53	1,35	1,20	1,08
14	2,98	2,14	1,76	1,52	1,35	1,20	1,08
15	2,95	2,13	1,75	1,52	1,34	1,20	1,07
16	2,92	2,12	1,75	1,51	1,34	1,19	1,07
17	2,90	2,11	1,74	1,51	1,33	1,19	1,07
18	2,88	2,10	1,73	1,50	1,33	1,19	1,07
19	2,86	2,09	1,73	1,50	1,33	1,19	1,07
20	2,85	2,09	1,72	1,50	1,33	1,18	1,06
21	2,83	2,08	1,72	1,49	1,32	1,18	1,06
22	2,82	2,07	1,72	1,49	1,32	1,18	1,06
23	2,81	2,07	1,71	1,49	1,32	1,18	1,06
24	2,80	2,06	1,71	1,49	1,32	1,18	1,06
25	2,79	2,06	1,71	1,49	1,32	1,18	1,06
26	2,78	2,06	1,71	1,48	1,31	1,18	1,06
27	2,77	2,05	1,70	1,48	1,31	1,18	1,06
28	2,76	2,05	1,70	1,48	1,31	1,17	1,06
29	2,76	2,05	1,70	1,48	1,31	1,17	1,06
30	2,75	2,04	1,70	1,48	1,31	1,17	1,05
31	2,74	2,04	1,70	1,48	1,31	1,17	1,05
32	2,74	2,04	1,69	1,47	1,31	1,17	1,05
33	2,73	2,03	1,69	1,47	1,31	1,17	1,05
34	2,73	2,03	1,69	1,47	1,31	1,17	1,05
35	2,72	2,03	1,69	1,47	1,31	1,17	1,05
36	2,72	2,03	1,69	1,47	1,31	1,17	1,05
37	2,72	2,03	1,69	1,47	1,30	1,17	1,05
38	2,71	2,02	1,69	1,47	1,30	1,17	1,05
39	2,71	2,02	1,68	1,47	1,30	1,17	1,05
40	2,70	2,02	1,68	1,47	1,30	1,17	1,05
41	2,70	2,02	1,68	1,47	1,30	1,17	1,05
42	2,70	2,02	1,68	1,47	1,30	1,17	1,05
43	2,70	2,02	1,68	1,47	1,30	1,17	1,05
44	2,69	2,02	1,68	1,47	1,30	1,17	1,05
45	2,69	2,01	1,68	1,46	1,30	1,17	1,05
46	2,69	2,01	1,68	1,46	1,30	1,17	1,05
47	2,68	2,01	1,68	1,46	1,30	1,16	1,05
48	2,68	2,01	1,68	1,46	1,30	1,16	1,05
49	2,68	2,01	1,68	1,46	1,30	1,16	1,05
50	2,68	2,01	1,68	1,46	1,30	1,16	1,05

Значення t-критерію Стьюдента при рівні значущості

(0,01; 0,05; 0,1; 0,15; 0,20; 0,25; 0,30)

Таблиця 4.2

n	P - 0,01	P - 0,05	P - 0,1	P - 0,15	P - 0,2	P - 0,25	P - 0,3
51	2,68	2,01	1,68	1,46	1,30	1,16	1,05
52	2,67	2,01	1,67	1,46	1,30	1,16	1,05
53	2,67	2,01	1,67	1,46	1,30	1,16	1,05
54	2,67	2,00	1,67	1,46	1,30	1,16	1,05
55	2,67	2,00	1,67	1,46	1,30	1,16	1,05
56	2,67	2,00	1,67	1,46	1,30	1,16	1,05
57	2,66	2,00	1,67	1,46	1,30	1,16	1,05
58	2,66	2,00	1,67	1,46	1,30	1,16	1,05
59	2,66	2,00	1,67	1,46	1,30	1,16	1,05
60	2,66	2,00	1,67	1,46	1,30	1,16	1,05
61	2,66	2,00	1,67	1,46	1,30	1,16	1,05
62	2,66	2,00	1,67	1,46	1,30	1,16	1,05
63	2,66	2,00	1,67	1,46	1,30	1,16	1,05
64	2,65	2,00	1,67	1,46	1,29	1,16	1,04
65	2,65	2,00	1,67	1,46	1,29	1,16	1,04
66	2,65	2,00	1,67	1,46	1,29	1,16	1,04

67	2,65	2,00	1,67	1,46	1,29	1,16	1,04
68	2,65	2,00	1,67	1,46	1,29	1,16	1,04
69	2,65	1,99	1,67	1,46	1,29	1,16	1,04
70	2,65	1,99	1,67	1,46	1,29	1,16	1,04
71	2,65	1,99	1,67	1,46	1,29	1,16	1,04
72	2,65	1,99	1,67	1,46	1,29	1,16	1,04
73	2,64	1,99	1,67	1,45	1,29	1,16	1,04
74	2,64	1,99	1,67	1,45	1,29	1,16	1,04
75	2,64	1,99	1,67	1,45	1,29	1,16	1,04
76	2,64	1,99	1,67	1,45	1,29	1,16	1,04
77	2,64	1,99	1,66	1,45	1,29	1,16	1,04
78	2,64	1,99	1,66	1,45	1,29	1,16	1,04
79	2,64	1,99	1,66	1,45	1,29	1,16	1,04
80	2,64	1,99	1,66	1,45	1,29	1,16	1,04
81	2,64	1,99	1,66	1,45	1,29	1,16	1,04
82	2,64	1,99	1,66	1,45	1,29	1,16	1,04
83	2,64	1,99	1,66	1,45	1,29	1,16	1,04
84	2,64	1,99	1,66	1,45	1,29	1,16	1,04
85	2,63	1,99	1,66	1,45	1,29	1,16	1,04
86	2,63	1,99	1,66	1,45	1,29	1,16	1,04
87	2,63	1,99	1,66	1,45	1,29	1,16	1,04
88	2,63	1,99	1,66	1,45	1,29	1,16	1,04
89	2,63	1,99	1,66	1,45	1,29	1,16	1,04
90	2,63	1,99	1,66	1,45	1,29	1,16	1,04
91	2,63	1,99	1,66	1,45	1,29	1,16	1,04
92	2,63	1,99	1,66	1,45	1,29	1,16	1,04
93	2,63	1,99	1,66	1,45	1,29	1,16	1,04
94	2,63	1,99	1,66	1,45	1,29	1,16	1,04
95	2,63	1,99	1,66	1,45	1,29	1,16	1,04
96	2,63	1,98	1,66	1,45	1,29	1,16	1,04
97	2,63	1,98	1,66	1,45	1,29	1,16	1,04
98	2,63	1,98	1,66	1,45	1,29	1,16	1,04
99	2,63	1,98	1,66	1,45	1,29	1,16	1,04
100	2,63	1,98	1,66	1,45	1,29	1,16	1,04

Таблиця критичних точок розподілу Пірсона «хі-квадрат»)

Таблиця 4.3

k / α	0,01	0,025	0,05	0,95	0,975	0,99
1	6,63	5,02	3,84	0,00	0,00	0,00
2	9,21	7,38	5,99	0,10	0,05	0,02
3	11,34	9,35	7,81	0,35	0,22	0,11
4	13,28	11,14	9,49	0,71	0,48	0,30
5	15,09	12,83	11,07	1,15	0,83	0,55
6	16,81	14,45	12,59	1,64	1,24	0,87
7	18,48	16,01	14,07	2,17	1,69	1,24
8	20,09	17,53	15,51	2,73	2,18	1,65
9	21,67	19,02	16,92	3,33	2,70	2,09
10	23,21	20,48	18,31	3,94	3,25	2,56
11	24,72	21,92	19,68	4,57	3,82	3,05
12	26,22	23,34	21,03	5,23	4,40	3,57
13	27,69	24,74	22,36	5,89	5,01	4,11

14	29,14	26,12	23,68	6,57	5,63	4,66
15	30,58	27,49	25,00	7,26	6,26	5,23
16	32,00	28,85	26,30	7,96	6,91	5,81
17	33,41	30,19	27,59	8,67	7,56	6,41
18	34,81	31,53	28,87	9,39	8,23	7,01
19	36,19	32,85	30,14	10,12	8,91	7,63
20	37,57	34,17	31,41	10,85	9,59	8,26
21	38,93	35,48	32,67	11,59	10,28	8,90
22	40,29	36,78	33,92	12,34	10,98	9,54
23	41,64	38,08	35,17	13,09	11,69	10,20
24	42,98	39,36	36,42	13,85	12,40	10,86
25	44,31	40,65	37,65	14,61	13,12	11,52
26	45,64	41,92	38,89	15,38	13,84	12,20
27	46,96	43,19	40,11	16,15	14,57	12,88
28	48,28	44,46	41,34	16,93	15,31	13,56
29	49,59	45,72	42,56	17,71	16,05	14,26
30	50,89	46,98	43,77	18,49	16,79	14,95
31	52,19	48,23	44,99	19,28	17,54	15,66
32	53,49	49,48	46,19	20,07	18,29	16,36
33	54,78	50,73	47,40	20,87	19,05	17,07
34	56,06	51,97	48,60	21,66	19,81	17,79
35	57,34	53,20	49,80	22,47	20,57	18,51
36	58,62	54,44	51,00	23,27	21,34	19,23
37	59,89	55,67	52,19	24,07	22,11	19,96
38	61,16	56,90	53,38	24,88	22,88	20,69
39	62,43	58,12	54,57	25,70	23,65	21,43
40	63,69	59,34	55,76	26,51	24,43	22,16
41	64,95	60,56	56,94	27,33	25,21	22,91
42	66,21	61,78	58,12	28,14	26,00	23,65
43	67,46	62,99	59,30	28,96	26,79	24,40
44	68,71	64,20	60,48	29,79	27,57	25,15
45	69,96	65,41	61,66	30,61	28,37	25,90
46	71,20	66,62	62,83	31,44	29,16	26,66
47	72,44	67,82	64,00	32,27	29,96	27,42
48	73,68	69,02	65,17	33,10	30,75	28,18
49	74,92	70,22	66,34	33,93	31,55	28,94
50	76,15	71,42	67,50	34,76	32,36	29,71

Задача № 5
Перевірка статистичних гіпотез. Критерій згоди χ^2 (Критерій Пірсона)
Короткий теоретичний матеріал

Деяка пропозиція щодо властивостей генеральної сукупності, з якої зроблена вибірка, називається **статистичною гіпотезою**. Процедура обґрунтованого зіставлення висловленої гіпотези з наявними вибірковими даними здійснюється за допомогою того чи іншого статистичного критерію і називається перевіркою статистичних гіпотез.

Зазвичай за вибірковими даними обчислюється деяка характеристика (міра розподілу), яка порівнюється з еталоном (критичним значенням), заданих таблицею. Тому перевірка статистичних гіпотез зводиться до процедури зіставлення обчислювальної характеристики з деякими табличними значеннями. Зазвичай вона полягає в тому, що одному або декільком параметрам генеральної сукупності приписується деяке значення або **робиться пропозиція щодо виду закону розподілу генеральної сукупності або різних її властивостей**. Одна з гіпотез вибирається в якості вихідної і позначається H_0 . Наприклад, $H_0: m_x=0,5$. Тут двокрапка аналогічна до союзу «що». Гіпотеза, яка суперечить нульовий, називається конкуруючою (альтернативною) гіпотезою і позначається H_1 .

Шляхом статистичної перевірки необхідно встановити, наскільки дані, отримані з вибірки, узгоджуються з висловленою гіпотезою. Процедура перевірки гіпотези - це правило, за яким гіпотеза приймається або відкидається. Абсолютно надійне рішення щодо перевіряємої гіпотези отримати не можна. Необхідно заздалегідь допустити можливість помилкового рішення, причому можливі помилки двох родів: відкинути вірну гіпотезу H_0 при її правильності (помилка 1-го роду) і прийняти гіпотезу H_0 при правильності альтернативної гіпотези H_1 (помилка 2-го роду). Для визнання статистичного критерію перевірки висунутої гіпотези H_0 "хорошим", необхідно вимагати, щоб при цьому критерії ймовірність відкинути вірну гіпотезу α (що називається рівнем значущості) була мала, а ймовірність відкинути помилкову гіпотезу велика. З двох критеріїв, що відповідають одній і тій же ймовірності α , слід віддати перевагу тому, який дає більшу ймовірність відкинути гіпотезу H_0 , коли вона є хибною. Ймовірність помилки першого роду позначають α і називають **рівнем значущості**. Це досить мала ймовірність, при якій в даній задачі подію можна вважати практично неможливим. Величина α вибирається експериментатором в залежності від практичних ситуацій. При вирішенні більшості технічних завдань вибирається $\alpha=0,05$.

Процедура перевірки гіпотези H_0 полягає в наступному: вибирається деяка відповідна функція вибірових даних $K(x_1, x_2, \dots, x_n; H_0)$, яка визначається вибіркою та висунутою гіпотезою H_0 і називається статистичним критерієм. Функція $K(x_1, x_2, \dots, x_n; H_0)$ вибирається такою, щоб закон її розподілу при справедливості H_0 був відомий. Потім встановлюється область, ймовірність попадання значення статистичного критерію до якої у разі справедливості гіпотези H_0 дорівнює α , тобто ці значення є неможливими у однократному випадку. Ця область називається критичною. Якщо обчислене за вибіркою значення статистичного критерію потрапляє у критичну область, гіпотеза H_0 відкидається, у іншому випадку - немає підстав для відхилення гіпотези.

Точки, що відокремлюють критичну область від області прийняття гіпотези, називаються критичними. Розрізняють односторонні і двосторонні критичні області. Перевірка різних статистичних гіпотез проводиться у відповідності з наступною загальною схемою:

1. Формується нульова гіпотеза H_0 .
2. Вибирається статистичний критерій K .
3. Обчислюється спостережуване (розрахункове) значення критерію $K_{набл}$.
4. Залежно від рівня значущості α за таблицями знаходиться критичне значення критерію $K_{кр}$.
5. Порівнюється спостережуване і критичне значення критерію.
6. Робиться висновок щодо справедливості висунутої гіпотези:

якщо $K_{набл} \leq K_{кр}$, то H_0 вірна;

якщо $K_{набл} > K_{кр}$, то H_0 невірна.

Критерій для перевірки гіпотези про вид закону розподілу генеральної сукупності називається критерієм згоди. На практиці найчастіше застосовується критерій згоди χ^2 - Пірсона. У критерії згоди χ^2 - Пірсона за міру розбіжності між гіпотетичним (пропонованим) і емпіричним законами розподілу приймається сума квадратів відхилення відносної частоти від відповідної ймовірності, взятих з певною вагою $C_i = n/p_i$, т.е.

$$\sum_{i=1}^K (P_i^* - P_i)^2 \cdot C_i = \sum_{i=1}^K \left(\frac{m_i}{n} - P_i \right)^2 \cdot \frac{n}{P_i} = \sum_{i=1}^K \frac{(m_i - np_i)^2}{np_i}, \quad \text{яка при великих } n \text{ має } \chi^2$$

2 розподіл з числом ступенів свободи $f = k - e - 1$, де k - число інтервалів, якщо спостережувана величина неперервна і результати спостережень представлені інтервальним варіаційним рядом,

або число можливих значень спостережуваної величини, якщо вона дискретна і результати спостережень представлені дискретним варіаційним рядом; e - число параметрів гіпотетичного розподілу, які оцінюються за вибіркою. Причому:

$$n = \sum_{i=1}^K m_i \text{ – об'єм вибірки.}$$

Тут m_i – кількість значень ознаки, що потрапили у i -й інтервал, P_i - ймовірність попадання значень ознаки у i -й інтервал.

Ця випадкова величина, якщо $m_i \geq 5$ підпорядковується закону розподілу χ^2 з числом ступенів свободи $f = k-1$ (постійний зв'язок $n = \sum_{i=1}^K m_i$).

Для конкретної вибірки ми отримаємо певне (розрахункове, спостережене) значення критерію

$$\chi^2_{\text{набл}} = \sum_{i=1}^k \frac{(m_i - np_i)^2}{np_i}.$$

Якщо параметри пропонованого розподілу невідомі, то вони визначаються за вибірковими даними, а значить число ступенів свободи зменшиться на кількість параметрів, тобто

$f = k-e-1$, де e - кількість параметрів, що визначаються за вибірковими даними. Критичне значення визначається за таблицею χ^2 розподілу:

$$\chi^2_{\text{крит}} = \chi^2_{\text{табл}}(\alpha; f=k-e-1).$$

Якщо $\chi^2_{\text{набл.}} > \chi^2_{\text{крит}}$, то гіпотеза відкидається; якщо $\chi^2_{\text{набл.}} \leq \chi^2_{\text{крит}}$, то робимо висновок, що дані суперечать гіпотезі.

Для обчислення $\chi^2_{\text{набл}}$ зазвичай використовують наступну таблицю:

J_i		m_i	p_i	np_i	$(m_i - np_i)^2$	$\frac{(m_i - np_i)^2}{np_i}$
$X^{(i)}$	$X^{(i+1)}$					

$$\sum_{i=1}^K m_i = n \quad \sum_{i=1}^K p_i = 1$$

$$\chi^2_{\text{набл}} = \sum$$

Тут $p_i = p(X^{(i)} < X < X^{(i+1)})$ обчислюється за певними формулами у залежності від виду запропонованого розподілу. При цьому вид передбачуваного розподілу визначається за характером отриманої гістограми. Для цього порівнюють вид гістограми з графіками щільності відомих розподілів:



рівномірний

показове

нормальне

Якщо побудована гістограма виявиться несхожою ні на одне з наведених розподілів, то це означає, що потрібно брати інші закони розподілу. Їх досить багато, і відповідні графіки можна знайти у підручниках або довідниках по теорії ймовірностей.

У кожному розподілі у формулі щільності розподілу ймовірностей $f(x)$ присутні параметри розподілу:

Розподіл:			
	<i>рівномірний</i>	<i>показове</i>	<i>нормальне</i>
Формула щільності	$f(x) = \begin{cases} C & \text{при } x \in (a, b) \\ 0 & \text{при } x \notin (a, b) \end{cases}$	$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & \text{при } x \geq 0 \\ 0 & \text{при } x < 0 \end{cases}$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$
Параметри	<u>два</u> a, b	<u>один</u> λ	<u>два</u> a, σ

Зазвичай оцінюють по вибірці *математичне сподівання, дисперсію і стандартне відхилення*, тобто m_x, D_x, σ_x , при цьому використовують відповідно *вибіркове середнє* $\bar{x}$, *виправлену вибіркву дисперсію* s^2 і *виправлене вибіркве стандартне відхилення*: S :

$$\boxed{\tilde{m}_x = \bar{x}}, \quad \boxed{\tilde{D}_x = s^2 = \frac{n}{n-1} \cdot D_v}, \quad \boxed{\tilde{\sigma}_x = S = \sqrt{s^2}}.$$

Для будь-якого з розподілів *числові характеристики* виражаються через *параметри розподілу* відомими співвідношеннями:

	Розподіл:		
	<i>рівномірний</i>	<i>показове</i>	<i>нормальне</i>
<i>математичне сподівання</i>	$m_x = \frac{b+a}{2}$	$m_x = \frac{1}{\lambda}$	$m_x = a$
<i>дисперсія</i>	$D_x = \frac{(b-a)^2}{12}$	$D_x = \frac{1}{\lambda^2}$	$D_x = \sigma$
<i>стандартне відхилення</i>	$\sigma_x = \frac{b-a}{2\sqrt{3}}$	$\sigma_x = \frac{1}{\lambda}$	$\sigma_x = \sigma$

З цих співвідношень *параметри розподілу* можна виразити через знайдені за вибіркою *числовим характеристикам*:

	Розподіл:		
	<i>рівномірний</i>	<i>показове</i>	<i>нормальне</i>
<i>оцінки для параметрів</i>	$\begin{cases} \tilde{a} = \bar{x} - \sqrt{3} \cdot s \\ \tilde{b} = \bar{x} + \sqrt{3} \cdot s \end{cases}$	$\tilde{\lambda} = \frac{1}{\bar{x}}$	$\begin{cases} \tilde{a} = \bar{x} \\ \tilde{\sigma} = s \end{cases}$

За формулами передбачуваного теоретичного закону розподілу підраховують теоретичні частоти. Для цього потрібно підрахувати ймовірності влучень у кожен з інтервалів з використанням формул передбачуваного розподілу:

	Розподіл:		
	<i>рівномірний</i>	<i>показове</i>	<i>нормальне</i>
$P(\alpha < X < \beta) =$	$= F(\beta) - F(\alpha)$	$= e^{-\lambda\alpha} - e^{-\lambda\beta}$	$= \Phi\left(\frac{\beta-a}{\sigma}\right) - \Phi\left(\frac{\alpha-a}{\sigma}\right)$
<i>Що використовується</i>	Функція розподілу $F(x)$	Показова функція $e^{-\lambda x}$	Функція Лапласа $\Phi\left(\frac{x-a}{\sigma}\right)$

Для нормального розподілу формула придатна при для будь-яких α, β .

Для показового – для додатних. α, β .

Для рівномірного розподілу вид формули залежить від того, де розташовані α, β щодо границь інтервалу (a, b) , тому для нього зручніше скористатися *функцією розподілу* $F(x)$:

$$F(x) = \begin{cases} 0 & \text{при } x < a \\ \frac{x-a}{b-a} & \text{при } a < x < b \\ 1 & \text{при } x > b \end{cases}.$$

У будь-якому випадку ймовірність попадання у інтервал підраховується як різниця значень деякої функції на кінцях цього інтервалу.

Перевірка гіпотези про закон розподілу за критерієм χ^2 проводиться відповідно до зазначеної вище загальної схеми, проте, якщо вихідний статистичний матеріал представлений вибіркою у первинній формі, то для обґрунтованого висунення гіпотези H_0 будується емпіричний розподіл, по вигляду якого вибирається той чи інший гіпотетичний розподіл $F(x)$. У зв'язку з цим перевірку гіпотези за допомогою критерію χ^2 - Пірсона роблять у такий спосіб:

1. Будується статистичний ряд розподілу.
2. Статистичний ряд зображується графічно у вигляді гістограми або полігону.
3. По виду емпіричного розподілу висувається гіпотеза про вид закону розподілу генеральної сукупності.
4. Визначаються оцінки параметрів пропонованого закону розподілу.
5. Обчислюється спостережуване значення критерію χ^2 за формулою

$$\chi_{\text{набл}}^2 = \sum_{i=1}^K \frac{(m_i - np_i)^2}{np_i} \quad (\text{інтервали з } m_i < 5 \text{ об'єднуються}).$$

6. Задається рівень значущості α і обчислюємо число ступенів свободи $f = k - e - 1$.
7. По таблиці χ^2 – розподілу (табл.5.1) знаходиться $\chi_{\text{крит}}^2 = \chi_{\text{табл}}^2(\alpha; f = k - e - 1)$
8. Порівнюються $\chi_{\text{набл}}^2$ і $\chi_{\text{кр}}^2$, якщо $\chi_{\text{набл}}^2 > \chi_{\text{кр}}^2$, гіпотеза H_0 відкидається, якщо $\chi_{\text{набл}}^2 \leq \chi_{\text{кр}}^2$ гіпотеза H_0 приймається.

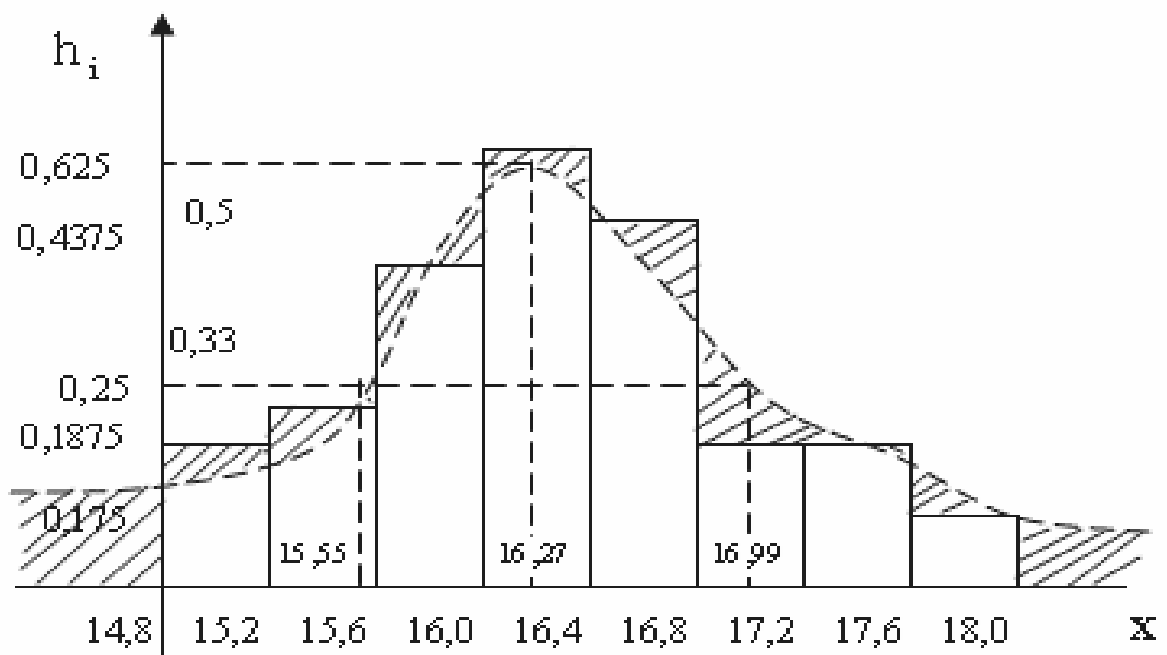
Розглянемо застосування критерію χ^2 - Пірсона на прикладі рішення деяких задач.

Приклад 1. При рівні значущості $\alpha = 0,05$ висловити гіпотезу про вид закону розподілу продуктивності мартенівської печі і перевіряти її за допомогою критерію χ^2 -Пірсона (використовувати інтервальный статистичний ряд, побудований раніше).

1) запишемо інтервальный статистичний ряд, доповнивши рядком $h_i = \frac{P_i^*}{\Delta x}$, який знадобиться при побудові гістограми:

J_i	[14,8;15,2)	[15,2;15,6)	[15,6;16,0)	[16,0;16,4)	[16,4;16,8)	[16,8;17,2)	[17,2;17,6)	[17,6;18,0)
m_i	3	4	7	10	8	3	3	2
P_i^*	0,075	0,1	0,175	0,25	0,2	0,075	0,075	0,05
$\frac{P_i^*}{\Delta x}$	0,1875	0,25	0,4375	0,625	0,5	0,1875	0,1875	0,175

2) будуюмо гістограму



3) з вигляду гістограми можна висловити гіпотезу H_0 : генеральна сукупність має нормальний розподіл;

4) нормальний розподіл має два параметри σ и m .

Їх оцінки:

$$\tilde{m} = \tilde{m}_x = \frac{1}{n} \sum_{i=1}^k m_i \tilde{x}_i = 16,27$$

$$\tilde{\sigma} = \tilde{\sigma}_x = \sqrt{\tilde{D}_x} = \sqrt{\frac{1}{n-1} \sum_{i=1}^k m_i (\tilde{x}_i - \tilde{m}_x)^2} = \sqrt{0,52} = 0,72;$$

Знаючи оцінки параметрів знаходимо точку екстремуму купола $(16,27; 0,56)$ $(m_x; 1/(\sigma \cdot \sqrt{2\pi}))$ і точки перегину $(15,55; 0,33), (16,99; 0,33)$ $(m_x \pm \sigma; 1/(\sigma \cdot e \cdot \sqrt{2\pi}))$.

5) $\chi^2_{\text{набл.}}$ визначається заштрихованою площею (чим вона більша, тим більше розбіжність). для обчислення $\chi^2_{\text{набл.}}$ складемо таблицю, об'єднавши інтервали с $m_i < 5$.

n=40

J _i		m _i	p _i	np _i	(m _i -np _i) ²	$\frac{(m_i - np_i)^2}{np_i}$
X ⁽ⁱ⁾	X ⁽ⁱ⁺¹⁾					
14,8	15,6	7	0,1162	4,648	5,5319	1,1902
15,6	16,0	7	0,1983	7,932	0,8686	0,1095
16,0	16,4	10	0,1969	7,876	4,5114	0,5128
16,4	17,2	11	0,3301	13,204	4,8576	0,3679
17,2	18,0	5	0,0985	3,94	1,1236	0,2852

$$\sum p_i = 0,94$$

$$\chi^2_{\text{набл}} = 2,4656$$

Імовірність p_i обчислюємо за формулою:

$$P_i = \Phi\left(\frac{x^{(i+1)} - \tilde{m}}{\tilde{\sigma}}\right) - \Phi\left(\frac{x^{(i)} - \tilde{m}}{\tilde{\sigma}}\right).$$

Причому, так як для нормального розподілу $-\infty < x < \infty$, то вважаємо, що

$$x^{(1)} = -\infty, \text{ а } x^{(6)} = +\infty,$$

$$\begin{aligned} P_1 &= \Phi\left(\frac{15,6 - 16,27}{0,72}\right) - \Phi\left(\frac{-\infty - 16,27}{0,72}\right) = \Phi(-0,93) - \Phi(-\infty) = -\Phi(0,93) + \Phi(\infty) = \\ &= -0,3238 + 0,5 = 0,1162 \end{aligned}$$

Значення $\Phi(x)$ визначаємо за таблицею 5.2.

$$\begin{aligned} P_2 &= \Phi\left(\frac{16,0 - 16,27}{0,72}\right) - \Phi\left(\frac{15,6 - 16,27}{0,72}\right) = \Phi(-0,32) - \Phi(-0,93) = -\Phi(0,32) + \Phi(0,93) = \\ &= -0,1255 + 0,3238 = 0,1983 \end{aligned}$$

$$\begin{aligned} P_3 &= \Phi\left(\frac{16,4 - 16,27}{0,72}\right) - \Phi\left(\frac{16,0 - 16,27}{0,72}\right) = \Phi(0,18) - \Phi(-0,32) = \Phi(0,18) + \Phi(0,32) = \\ &= 0,0714 + 0,1255 = 0,1969 \end{aligned}$$

$$P_4 = \Phi\left(\frac{17,2 - 16,27}{0,72}\right) - \Phi\left(\frac{16,4 - 16,27}{0,72}\right) = \Phi(1,29) - \Phi(0,18) = 0,4015 - 0,0714 = 0,3301$$

$$P_5 = \Phi\left(\frac{\infty - 16,27}{0,72}\right) - \Phi\left(\frac{17,2 - 16,27}{0,72}\right) = \Phi(\infty) - \Phi(1,29) = 0,5 - 0,4015 = 0,0985$$

6) Вважаємо $\alpha = 0,05$. Число ступенів свободи $f = k - e - 1 = 5 - 2 - 1 = 2$, так як залишилося після об'єднання 5 інтервалів, а число параметрів $e = 2$ (m і σ).

$$7) \chi^2_{\text{крит}} = \chi^2_{\text{табл}}(\alpha=0,05; f=2) = 6,0 \quad \chi^2_{\text{набл}} = 2,4656 < (\chi^2_{\text{крит}} = 6,0).$$

Висновок: вибіркові дані не суперечать гіпотезі про нормальний розподіл продуктивності даної мартенівської печі.

Розв'язок завдання у середовищі EXCEL:

1) будуюмо гістограму



Гістограма переноситься із завдання 3.

викликаючи «Мастер Диаграмм», «Гистограмма», будуюмо гістограму відносних частот (рис.5.1).

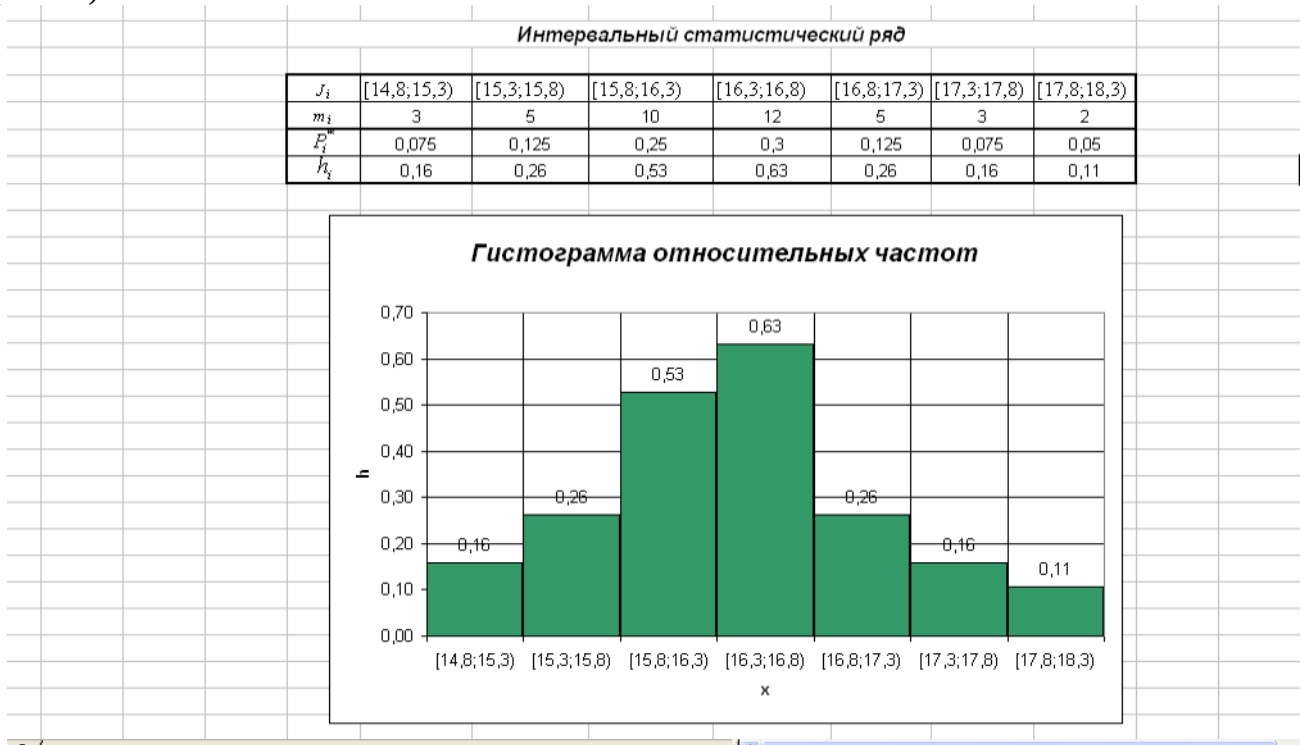


Рис. 5.1. Гістограма відносних частот

3) з вигляду гістограми можна висловити гіпотезу H_0 : генеральна сукупність має нормальний розподіл;

4) нормальний розподіл має два параметри σ і m .

$$\text{Їх оцінки визначають з четвертого завдання: } \tilde{m} = \tilde{m}_x = 16,35 \quad \tilde{\sigma} = \tilde{\sigma}_x = \sqrt{\tilde{D}_x} = 0,73.$$

5) Для обчислення $\chi^2_{\text{набл}}$ складемо таблицю, об'єднавши інтервали с $m_i < 5$.

При цьому слід завдання 3 перейменувати в завдання 5 і подальші операції виконувати, об'єднавши відповідні інтервали (рис.5.2).

Примітка: значення функції Лапласа можна взяти з таблиць або ж скористатися функцією НОРМРАСП категорія «Статистические». Для цього слід врахувати, що для стандартного нормального розподілу $\Phi(z) = F(z) - 0,5$.

Задача_5

Вид Вставка Формат Сервис Данные Справка

Англ Слуг 10 Ж К Ч

Введите вопрос

Расчетная таблица $\chi^2_{набл}$

$\chi^{(i)}$	$\chi^{(i+1)}$	m_i	$\Phi^{(i+1)}$	$\Phi^{(i)}$	P_i	$n P_i$	$(m_i - n P_i)^2$	$\frac{(m_i - n P_i)^2}{n P_i}$
14,8	15,8	8	-0,27	-0,50	0,23	9,02	1,05	0,12
15,8	16,3	10	-0,03	-0,27	0,25	9,88	0,01	0,00
16,3	16,8	12	0,23	-0,03	0,26	10,34	2,76	0,27
16,8	17,3	5	0,40	0,23	0,17	6,89	3,57	0,52
17,3	18,3	5	0,50	0,40	0,10	3,86	1,29	0,33
					Σ		$\chi^2_{набл} =$	1,24

Рис.5.2. Розрахункова таблиця визначення $\chi^2_{набл}$

Розрахункове значення $\chi^2_{набл}$ заносять в комірку I69.

6) Вважаємо $\alpha=0,05$. Число ступенів свободи $f=k-e-1=5-2-1=2$, оскільки залишилося після об'єднання 5 інтервалів, а число параметрів $e=2$ (m и σ). Тоді:

$$\chi^2_{крит} = \chi^2_{табл}(\alpha=0,05; f=2) = 6,0$$

У середовищі EXCEL критичне значення $\chi^2_{крит}$ визначають по функції

$$=ХИ2ОБР(0,05;2)$$

і критичне значення критерію заносять в комірку I72.

7) В комірку H75 вводять формулу

=ЕСЛИ(I69<=I72;"При заданном уровне значимости гипотеза H_0 не противоречит выборочным значениям"; "гипотеза H_0 отклоняется")

Результат розв'язання задачі наведено на рис.5.3.

Висновок: вибіркові дані не суперечать гіпотезі про нормальний розподіл продуктивності мартенівської печі.

На АТС проводилися спостереження за числом викликів в хвилину. Спостереження протягом години дали наступні результати:

Сформулювати гіпотезу про теоретичний розподіл числа викликів в хвилину і перевірити ступінь згоди теоретичного та емпіричного розподілів за критерієм χ^2 з рівнем значущості $\alpha = 0,05$.

- | | | | | | | | | |
|---------|-------|-------|-------|-------|-----|-------|---|-------|
| x_i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| m_i | 8 | 17 | 16 | 10 | 6 | 2 | 0 | 1 |
| p_i^* | 0,133 | 0,284 | 0,268 | 0,166 | 0,1 | 0,033 | 0 | 0,013 |

X_i	p_i^*
0	0.17
1	0.30
2	0.28
3	0.20
4	0.14
5	0.09
6	0.05
7	0.09

2. Більш ймовірними є малі значення. Можна припустити, що має місце закон розподілу Пуассона. Характерною ознакою закону Пуассона є рівність $m_x = D_x$.
3. Обчислимо:

$$m_x = \frac{0 \cdot 8 + 1 \cdot 17 + 2 \cdot 16 + 3 \cdot 10 + 4 \cdot 6 + 5 \cdot 2 + 6 \cdot 0 + 7 \cdot 1}{60} = 2 = \bar{X};$$

$$\tilde{D}_x = \frac{8(0-2)^2 + 17(1-2)^2 + 16(2-2)^2 + 10(3-2)^2 + 6(4-2)^2 + 2(5-2)^2 + 0(6-2)^2 + 1(7-2)^2}{60} \approx 2,1.$$

Можна вважати, що $m_x \approx D_x$. Висуваємо гіпотезу H_0 : число викликів має розподіл Пуассона.

4. Розподіл Пуассона залежить від одного параметра. Його оцінка $\tilde{a} = \bar{X} = 2$.

5. Обчислимо критерій $\chi^2_{набл}$, складемо розрахункову таблицю.

Теоретичні ймовірності обчислюємо за формулою:

$$p_i = p(x_i) = \frac{2^{x_i}}{x_i!} e^{-2}; \quad (e^{-2} = 0,1353).$$

X_i	m_i	P_i	np_i	$m_i - np_i$	$(m_i - p_i)^2$	$\frac{(m_i - np_i)^2}{np_i}$
0	8	0,1353	8,13	-0,1	0,01	0,0012
1	17	0,2707	16,24	0,76	0,5776	0,0356
2	16	0,2707	16,24	-0,24	0,576	0,0035
3	10	0,1804	10,82	-0,82	0,6724	0,0621
4	9	0,1429	8,57	0,43	0,1849	0,0215
и бо- лее						
	$\sum m_i = 60$	$\sum P_i = 1$	$\sum n_i = 60$			$\chi^2_{набл} = 0,1239$

6. Число ступенів свободи $f = k - e - 1 = 5 - 1 - 1 = 3$; $k = 5$, оскільки значення 5, 6, 7 мають мале число спостережень (< 5) і їх об'єднуємо зі значенням 4.

7. В таблиці 5.1 значенням $\alpha = 0,05$ і $f = 3$ відповідає $\chi^2_{кр} = 7,8$.

8. Висновок: так як $\chi^2_{набл} < \chi^2_{крит} (0,12 < 7,8)$, то гіпотеза H_0 не відкидається.

Застосування критерію χ^2 до інших розподілів відрізняється від наведених тільки способами обчислень величин p_i .

Для рівномірного розподілу:

$$P_1 = \frac{(X_1 - a)}{b - a}; \quad P_i = \frac{X_i - X_{i-1}}{b - a}; \quad (i = 2, \dots, k-1); \quad P_k = \frac{b - X_{k-1}}{b - a};$$

для показового розподілу:

$$P_i = e^{-\tilde{\lambda}^* x_i} - e^{-\tilde{\lambda}^* x_{i+1}}; (i = 1, 2, \dots, \kappa).$$

Розв'язання задачі в середовищі EXCEL:



1. У листі 2 завдання 5 формують дискретний статистичний ряд (рис.5.4).

Microsoft Excel - Задача_5

Г15

Дискретный статистический ряд

x_i	0	1	2	3	4	5	6	7
m_i	8	17	16	10	6	2	0	1
p_i^*	0.13	0.28	0.27	0.17	0.10	0.03	0.00	0.02

Рис.5.4. Дискретний статистичний ряд

2. Викликаючи «Мастер Диаграмм», «Точечная Диаграмма», будують *полигон относительных частот*. Редагують графік таким чином, щоб він був максимально схожий на наведений на рисунку 5.5.

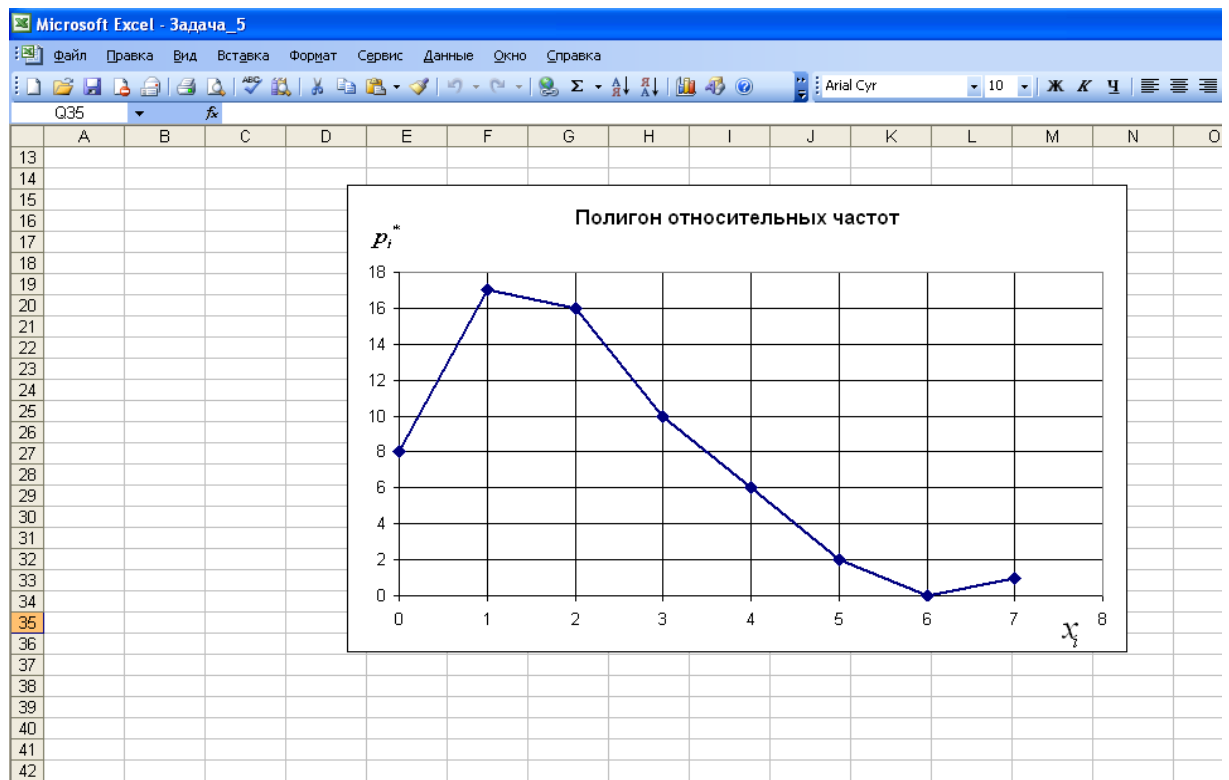


Рис. 5.5. Полігон відносних частот

3. Можна припустити, що має місце закон розподілу Пуассона в силу малих значень ймовірностей. Характерною ознакою закону Пуассона є рівність $m_x = D_x$.

В комірках E13 і I13 обчислюють оцінки математичного очікування і дисперсії (рис.5.6). Для оцінки математичного очікування застосовують функцію

$$=\text{СУММПРОИЗВ}(E7:L7;E8:L8)/60$$

x_i	0	1	2	3	4	5	6	7
m_i	8	17	16	10	6	2	0	1
p_i	0,13	0,28	0,27	0,17	0,10	0,03	0,00	0,02
Сум. кв.	32	17	0	10	24	18	0	25
Оц. Дисп.	2,1			Оц. М. о.	=B13			

Рис. 5.6. Обчислення точкових оцінок вибіркової сукупності

4. Розподіл Пуассона залежить від одного параметра. Його оцінка $\tilde{a} = \bar{X} = 2$.

5. Обчислюють критерій $\chi^2_{набл}$, складають при цьому розрахункову таблицю (рис.5.7). Теоретичні ймовірності обчислюємо за формулою:

$$p_i = p(x_i) = \frac{2^{x_i}}{x_i!} e^{-2}; \quad (e^{-2} = 0,1353).$$

x_i	m_i	p_i	$n p_i$	$(m_i - n p_i)$	$(m_i - n p_i)^2$	$\frac{(m_i - n p_i)^2}{n p_i}$
0	8	0,1353	8,12	-0,12	0,01	0,002
1	17	0,2706	16,24	0,76	0,58	0,036
2	16	0,2706	16,24	-0,24	0,06	0,003
3	10	0,1804	10,82	-0,82	0,68	0,063
4 и >	9	0,141	8,46	0,54	0,29	0,034
	60	0,9979	59,87		$\chi^2_{набл}$	0,138

Рис. 5.7. Розрахункова таблиця обчислення $\chi^2_{набл}$

Спостережуване значення критерію заносять до комірки I59.

6. Число ступенів свободи $f=k-e-1=5-1-1=3$; $k=5$, оскільки значення 5, 6, 7 мають мале число спостережень (<5) і їх об'єднуємо зі значенням 4.

У середовищі EXCEL критичне значення $\chi^2_{\text{крит}}$ визначають по функції

$$=\text{ХИ2ОБР}(0,05;3)$$

і критичне значення критерію заносять до комірки І62.

7. У комірки Н75 вводять формулу

$$=\text{ЕСЛИ}(\text{I59} \leq \text{I62}; \text{"При заданном уровне значимости ги-"} \\ \text{потеза } H_0 \text{ не противоречит выборочным значениям"; "гипотеза } H_0 \text{ отклоняется"})$$

Результат розв'язання задачі ілюструється на рис.5.8.

X_i	m_i	p_i	$n p_i$	$(m_i - n p_i)$	$(m_i - n p_i)^2$	$\frac{(m_i - n p_i)^2}{n p_i}$
0	8	0,1353	8,12	-0,12	0,01	0,002
1	17	0,2706	16,24	0,76	0,58	0,036
2	16	0,2706	16,24	-0,24	0,06	0,003
3	10	0,1804	10,82	-0,82	0,68	0,063
4 и >	9	0,141	8,46	0,54	0,29	0,034
	60	0,9979	59,87		$\chi^2_{\text{набл}}$	0,138
					$\chi^2_{\text{табл}}$	7,81
					При заданном уровне значимости гипотеза H_0 не противоречит выборочным значениям	

Рис.5.8. Результат розв'язання задачі у середовищі EXCEL

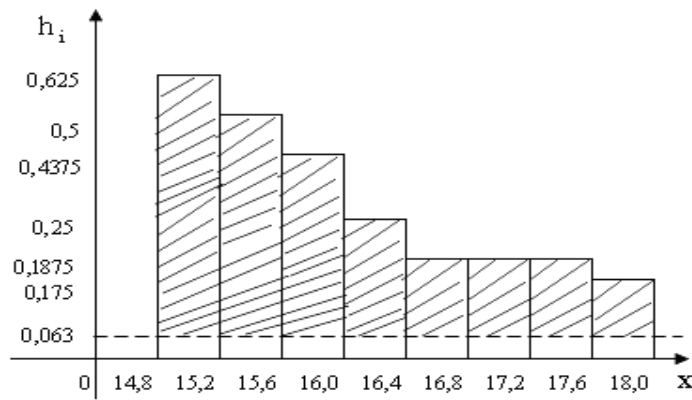
Приклад 3.

Для заданого ряду сформулювати гіпотезу про теоретичний розподіл і перевірити ступінь згоди теоретичного та емпіричного розподілів за критерієм χ^2 з рівнем значущості $\alpha = 0,05$.

1) вихідні дані мають вигляд:

J_i	[14,8;15,2)	[15,2;15,6)	[15,6;16,0)	[16,0;16,4)	[16,4;16,8)	[16,8;17,2)	[17,2;17,6)	[17,6;18,0)
m_i	10	8	7	4	3	3	3	2
P_i^*	0.25	0.2	0.175	0.1	0.075	0.075	0.075	0.05
$\frac{P_i^*}{\Delta x}$	0.625	0.5	0.4375	0.25	0.1875	0.1875	0.1875	0.175

2) будують гістограму



3) за виглядом гістограми висловлюється гіпотеза H_0 : генеральна сукупність має показовий закон розподілу;

4) показовий розподіл має один параметр λ .

Його оцінка визначається наступним чином:

$$\tilde{\lambda} = \frac{1}{\tilde{m}_x} = \frac{1}{\frac{1}{n} \sum_{i=1}^k m_i x_i} = \frac{n}{\sum_{i=1}^k m_i x_i} = \frac{40}{10 \cdot 15 + 8 \cdot 15,4 + 7 \cdot 15,8 + 4 \cdot 16,2 + 3 \cdot 16,6 + 3 \cdot 17 + 3 \cdot 17,4 + 2 \cdot 17,8} = \frac{40}{631,2} = 0,063$$

5) Знаючи λ будують теоретичну криву розподілу, потім заштриховують розбіжності між емпіричним законом розподілу і передбачуваним теоретичним. для обчислення $\chi^2_{\text{набл}}$ складемо таблицю, об'єднавши інтервали с $m_i < 5$.

J _i		m _i	p _i	np _i	(m _i -np _i) ²	$\frac{(m_i - np_i)^2}{np_i}$
X ⁽ⁱ⁾	X ⁽ⁱ⁺¹⁾					
14,8	15,2	10	0,6162	24,6480	214,5639	8,1051
15,2	15,6	8	0,0095	0,3800	58,0644	152,8010
15,6	16,4	11	0,0184	0,7360	109,3496	143,1381
16,4	17,2	6	0,0115	0,4600	30,6916	66,7209
17,2	18,0	5	0,3384	13,5360	72,8633	5,3829

$$\sum p_i = 0,994$$

$$\chi^2_{\text{набл}} = 376,148$$

Ймовірності $p_i = P(X^{(i)} < X < X^{(i+1)})$ обчислюються за формулою

$$P_i = e^{-\lambda x^{(i)}} - e^{-\lambda x^{(i+1)}}$$

причому, оскільки $0 < x < \infty$, то приймаємо $X^{(1)} = 0$; $X^{(6)} = \infty$.

Отже:

$$P_1 = e^{-0,063 \cdot 0} - e^{-0,063 \cdot 15,2} = 1 - 0,3838 = 0,6162;$$

$$P_2 = e^{-0,063 \cdot 15,2} - e^{-0,063 \cdot 15,6} = 0,3838 - 0,3743 = 0,0095$$

$$P_3 = e^{-0,63 \cdot 15,6} - e^{-0,063 \cdot 16,4} = 0,3743 - 0,3559 = 0,0184;$$

$$P_4 = e^{-0,063 \cdot 16,4} - e^{-0,063 \cdot 17,2} = 0,3559 - 0,3384 = 0,0115$$

$$P_5 = e^{-0,063 \cdot 17,2} - e^{-0,063 \cdot \infty} = 0,3384$$

6) вважаємо $\alpha=0,05$. Число степенів свободи $f=k-e-1=5-1-1=3$, оскільки залишилося після об'єднання $k=5$ інтервалів, а число параметрів $e=1$ (λ).

$$7) \chi^2_{\text{крит}} = \chi^2_{\text{табл}}(\alpha=0,05; f=3)=7,8 \quad (\chi^2_{\text{набл}}=376,148) > (\chi^2_{\text{крит}}=7,8)$$

Висновок: гіпотеза щодо показового розподілу продуктивності даної мартенівської печі відкидається. Очевидно тут більш відповідним був б β – розподіл.

Розв'язання задачі у середовищі EXCEL:



1. У листі 3 завдання 5 будують варіаційний статистичний ряд (рис.5.9).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1																
2																
3																
4																
5																
6																
7																
8																
9																
10																
11																
12																

Рис. 5.9. Варіаційний статистичний ряд

2. Викликаючи «Мастер Диаграмм», «Гистограмма», будують гістограму відносних частот (рис. 5.10).

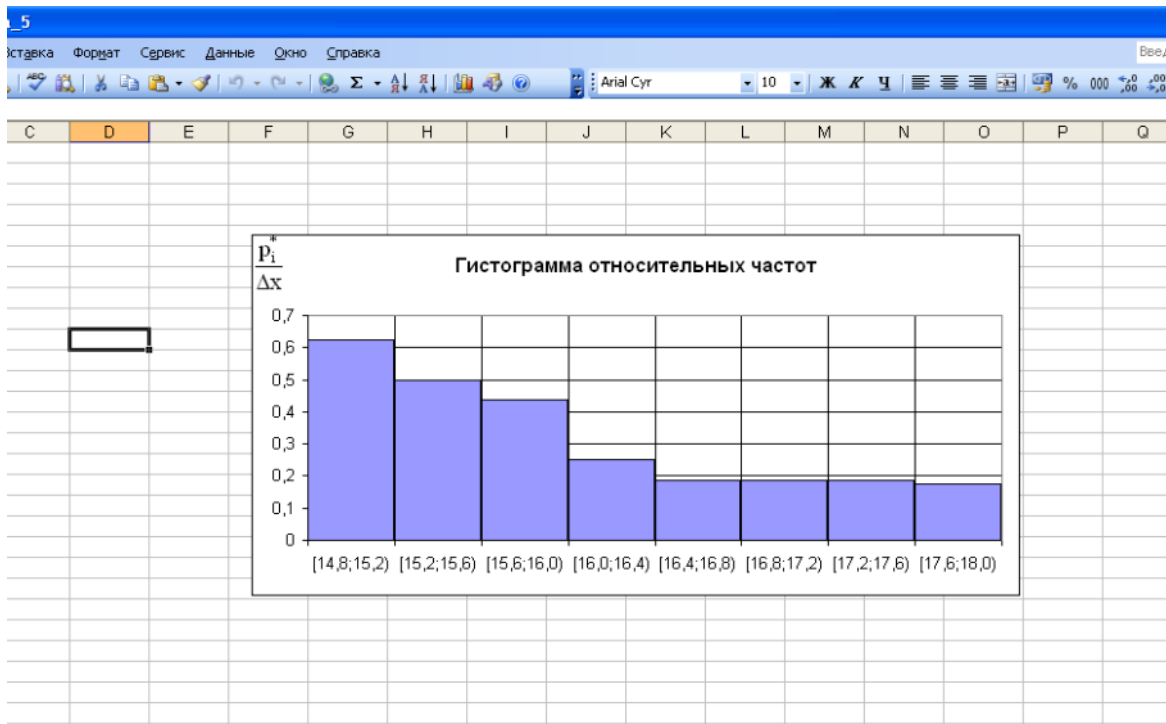


Рис.5.10. Гістограма відносних частот

3. За виглядом гістограми висловлюється гіпотеза H_0 : генеральна сукупність має показовий закон розподілу.

4. Показовий розподіл має один параметр λ .

Його оцінка визначається наступним чином:

$$\tilde{\lambda} = \frac{1}{\tilde{m}_x} = \frac{1}{\frac{1}{n} \sum_{i=1}^k m_i x_i} = \frac{n}{\sum_{i=1}^k m_i x_i}.$$

При цьому складають допоміжний ряд і на підставі функції

$$=40/\text{СУММПРОИЗВ}(H43:O43;H44:O44)$$

обчислюють оцінку параметра λ (рис.5.11).

$J_{\text{кр}}$	15	15,4	15,8	16,2	16,6	17	17,4	17,8
m_i	10	8	7	4	3	3	3	2

$\hat{\lambda} = 0,063$

Рис.5.11. Обчислення оцінки параметра λ

5. Знаючи значення параметра λ будують теоретичну криву розподілу, потім заштриховують розбіжності між емпіричним законом розподілу і передбачуваним теоретичним. Для обчислення $\chi^2_{\text{набл}}$ складають таблицю, об'єднавши інтервали з $m_i < 5$ (рис.5.12).

$X^{(i)}$	$X^{(i+1)}$	m_i	p_i	np_i	$(m_i - np_i)^2$	$\frac{(m_i - np_i)^2}{np_i}$
14,8	15,2	10	0,6162	24,648	214,5639	8,1051
15,2	15,6	8	0,0095	0,38	58,0644	152,801
15,6	16,4	11	0,0184	0,736	109,3496	143,1381
16,4	17,2	6	0,0115	0,46	30,6916	66,7209
17,2	18	5	0,3384	13,536	72,8633	5,3829

$\sum p_i = 0,994$ $\chi^2_{\text{набл}} = 376,148$

Рис. 5.12. Розрахункова таблиця $\chi^2_{\text{набл}}$

Ймовірності $p_i = P(X^{(i)} < X < X^{(i+1)})$ обчислюються за формулою:

$$P_i = e^{-\lambda x^{(i)}} - e^{-\lambda x^{(i+1)}}$$

, причому, оскільки $0 < x < \infty$, вважаємо $X^{(1)} = 0$; $X^{(6)} = \infty$.

Спостережуване значення критерію заносять в комірку K70.

6. Число ступенів свободи $f = k - e - l = 5 - 1 - 1 = 3$; $k = 5$, так як значення 5, 6, 7 мають мале число спостережень (< 5) і їх об'єднуємо зі значенням 4.

У середовищі EXCEL критичне значення визначають по функції
 $=\text{ХИ2ОБР}(0,05;3)$

і критичне значення критерію заносять в комірку K73.

7. В комірку K75 вводять формулу

=ЕСЛИ(K70<=K73;"При заданном уровне значимости гипотеза H_0 не противоречит выборочным значениям"; "гипотеза H_0 отклоняется")

Результат розв'язання задачі ілюструється на рис. 5.13.

J_i		m_i	p_i	np_i	$(m_i - np_i)^2$	$\frac{(m_i - np_i)^2}{np_i}$
$\chi^2(0)$	$\chi^2(0.05)$					
14,8	15,2	10	0,6162	24,648	214,5639	8,1051
15,2	15,6	8	0,0095	0,38	58,0644	152,801
15,6	16,4	11	0,0184	0,736	109,3496	143,1381
16,4	17,2	6	0,0115	0,46	30,6916	66,7209
17,2	18	5	0,3384	13,536	72,8633	5,3829
		$\sum p_i = 0,994$	$\chi^2_{набл} = 376,148$			
			$\chi^2_{крит} = 7,81$			
		гипотеза H_0 отклоняется				

Рис.5.13. Результат розв'язання задачі в середовищі EXCEL

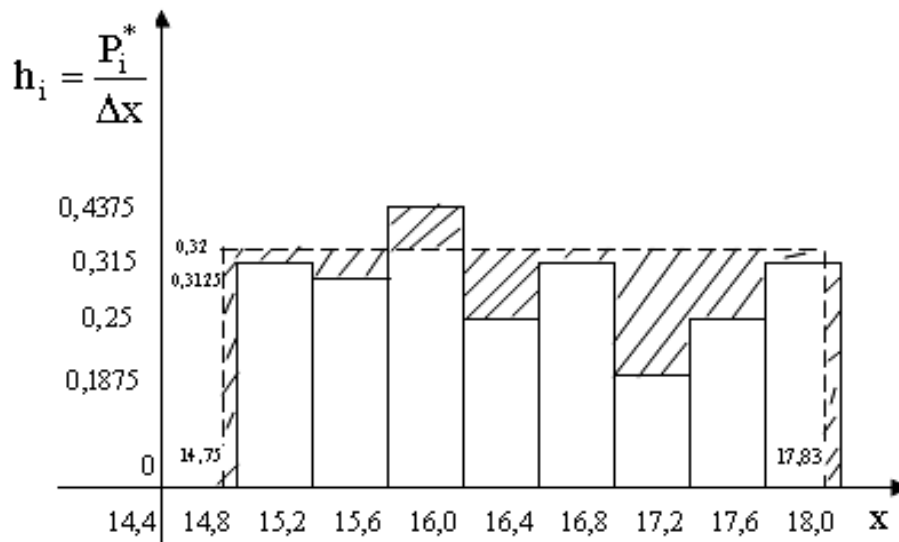
Висновок: гіпотеза про показовому розподілі продуктивності даної мартенівської печі відкидається. Очевидно, тут більш відповідним був б β – розподіл.

Приклад 4. Для заданого ряду сформулювати гіпотезу щодо теоретичного розподілу і перевірити ступінь згоди теоретичного та емпіричного розподілів за критерієм χ^2 з рівнем значущості $\alpha = 0,05$.

1) вихідні дані мають вигляд

J_i	[14,8;15,2)	[15,2;15,6)	[15,6;16,0)	[16,0;16,4)	[16,4;16,8)	[16,8;17,2)	[17,2;17,6)	[17,6;18,0)
m_i	5	5	7	4	6	3	5	5
p_i^*	0,125	0,125	0,175	0,1	0,15	0,075	0,125	0,125
$\frac{p_i^*}{\Delta x}$	0,3125	0,3125	0,4375	0,25	0,315	0,1875	0,3125	0,3125

2) побудуємо гістограму



3) з вигляду гістограми можна висловити гіпотезу H_0 : генеральна сукупність має рівномірний закон розподілу.

4) рівномірний закон розподілу має два параметри a і b . Їх оцінки визначаються на підставі наступних співвідношень:

$$\bar{a} = \bar{m}_x - \sqrt{3} \times \sqrt{\bar{D}_x}$$

$$\bar{b} = \bar{m}_x + \sqrt{3} \times \sqrt{\bar{D}_x}$$

Якщо $\tilde{m}_x$ и $\tilde{D}_x$ невідомі, то їх обчислюють за формулами:

$$\tilde{m}_x = \frac{1}{n} \sum_{i=1}^k m_i \tilde{x}_i = \frac{6 \cdot 15 + 5 \cdot 15,4 + 7 \cdot 15,8 + 4 \cdot 16,2 + 6 \cdot 16,6 + 3 \cdot 17 + 4 \cdot 17,4 + 5 \cdot 17,8}{40} = \frac{651}{40} = 16,29;$$

$$\tilde{D}_x = \frac{1}{n-1} \sum_{i=1}^k m_i (\tilde{x}_i - \tilde{m}_i)^2 = \frac{31,016}{39} = 0,7953;$$

$$\tilde{\sigma}_x = \sqrt{\tilde{D}_x} = \sqrt{0,7953} = 0,8918;$$

$$\tilde{a} = 16,29 - \sqrt{3} \cdot 0,8918 = 14,75;$$

$$\tilde{b} = 16,29 + \sqrt{3} \cdot 0,8918 = 17,83;$$

5) побудуємо криву рівномірного розподілу і заштрихуємо розбіжності між емпіричним законом розподілу і теоретичним. Крива рівномірного розподілу являє собою пряму проведену на висоті $\frac{1}{\tilde{b} - \tilde{a}} = \frac{1}{17,83 - 14,75} = 0,32$ від точки $\tilde{a} = 14,75$ до точки $\tilde{b} = 17,83$

Для обчислення $\chi^2_{набл}$ складемо таблицю, об'єднавши інтервали з $m_i < 5$

J_i	m_i	p_i	np_i	$(m_i - np_i)^2$	
-------	-------	-------	--------	------------------	--

$X^{(i)}$	$X^{(i+1)}$					$\frac{(m_i - np_i)^2}{np_i}$
14,8	15,2	6	0,1461	5,844	0,0243	
15,2	15,6	5	0,1298	5,192	0,0388	0,0071
15,6	16,0	7	0,1298	5,192	3,2689	0,6296
16,0	16,8	10	0,2596	10,384	0,1475	0,0142
16,8	17,6	7	0,2596	10,384	11,4515	1,1028
17,6	18,0	5	0,0747	2,987	4,0522	1,3566

$$\sum p_i = 0,9937$$

$$\chi^2_{\text{набл}} = 3,1103$$

Ймовірності $p_i = P(X^{(i)} < X < X^{(i+1)})$ обчислюємо за формулою:

$$p_i = \frac{X^{(i+1)} - X^{(i)}}{\tilde{b} - \tilde{a}},$$

причому, оскільки $\tilde{a} \leq X \leq \tilde{b}$, то $X^{(1)} = \tilde{a}$; $X^{(6)} = \tilde{b}$

Отже

$$P_1 = \frac{x^{(2)} - \tilde{a}}{\tilde{b} - \tilde{a}} = \frac{15,2 - 14,75}{17,83 - 14,75} = 0,1461$$

$$P_2 = P_3 = \frac{0,8}{17,83 - 14,75} = \frac{0,8}{3,08} = 0,1298$$

$$P_4 = P_5 = \frac{0,8}{17,83 - 14,75} = \frac{0,8}{3,08} = 0,2596$$

$$P_6 = \frac{\tilde{b} - X^{(6)}}{\tilde{b} - \tilde{a}} = \frac{17,83 - 17,6}{17,83 - 14,75} = 0,0747$$

6) вважаємо $\alpha=0,05$. Число ступенів свободи $f=k-e-1=6-2-1=3$, так як залишилося після об'єднання $k=6$ інтервалів, а число параметрів $e=2$ (а і b).

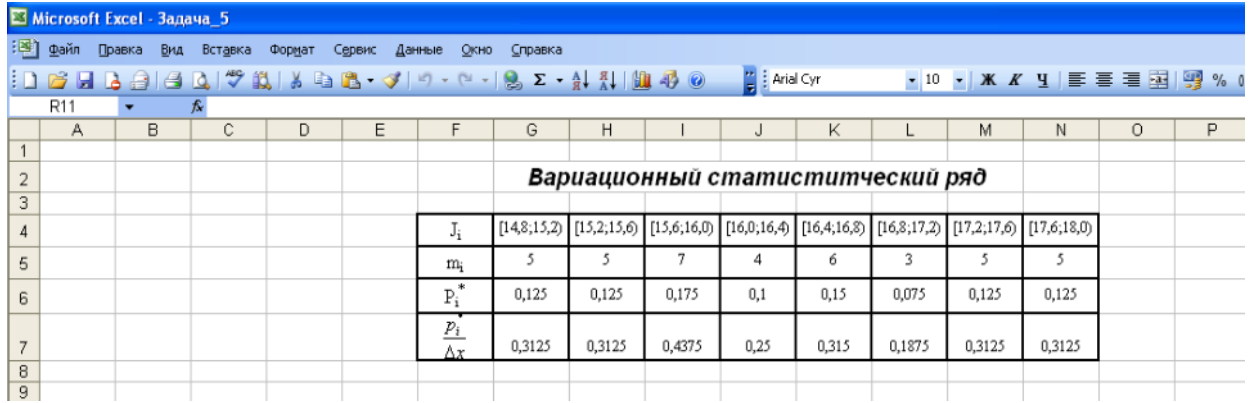
$$7) \chi^2_{\text{крит}} = \chi^2_{\text{табл}}(\alpha=0,05; f=3)=7,8 \quad (\chi^2_{\text{набл}}=3,1103) < (\chi^2_{\text{крит}}=7,8).$$

Висновок: вибірккові дані не суперечать гіпотезі про рівномірний розподіл продуктивності даної мартенівської печі.

Розв'язання завдання в середовищі EXCEL:



1. У завданні 5 вставляють лист 4 і будують варіаційний статистичний ряд (рис.5.14).



Microsoft Excel - Задача_5

Вариационный статистический ряд

J_i	[14,8;15,2)	[15,2;15,6)	[15,6;16,0)	[16,0;16,4)	[16,4;16,8)	[16,8;17,2)	[17,2;17,6)	[17,6;18,0)
m_i	5	5	7	4	6	3	5	5
P_i^*	0,125	0,125	0,175	0,1	0,15	0,075	0,125	0,125
$\bar{P}_i$								
Δx	0,3125	0,3125	0,4375	0,25	0,315	0,1875	0,3125	0,3125

Рис.5.14. Варіаційний статистичний ряд

2. Викликаючи «Мастер Диаграмм», «Гистограмма», будують гістограму відносних частот (рис.5.15).

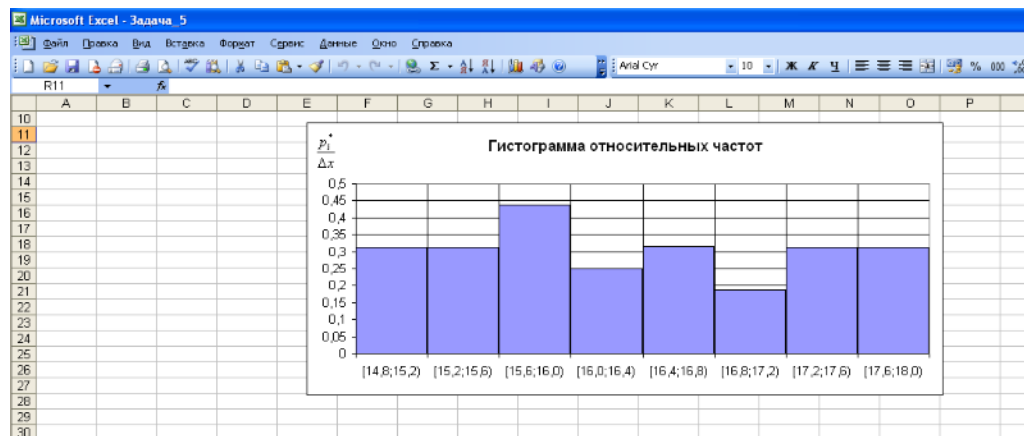


Рис.5.15. Гістограма відносних частот

3. За виглядом гістограми висловлюється гіпотеза H_0 : генеральна сукупність має рівномірний закон розподілу;

4. Рівномірний закон розподілу має два параметри a і b . Їх оцінки визначаються на підставі наступних співвідношень:

$$\bar{a} = \bar{m}_x - \sqrt{3} \times \sqrt{\bar{D}_x}$$

$$\bar{b} = \bar{m}_x + \sqrt{3} \times \sqrt{\bar{D}_x}$$

Для визначення оцінок $\tilde{m}_x$ і $\tilde{D}_x$ будують допоміжний ряд (рис.5.16).

Вспомогательный ряд									
J _{сер}	15	15,4	13,8	16,2	16,6	17	17,4	17,8	
m _i	5	5	7	4	6	3	5	5	

Рис.5.16. Допоміжний ряд для визначення оцінок розподілу

- у комірці N39 визначають оцінку математичного очікування $\tilde{m}_x$ за формулою

$$= \text{СУММПРОИЗВ}(H34:O34;H35:O35)/40$$

Для оцінки дисперсії складають ряд суми квадратів відповідних величин (рис.5.17).

H39	$\tilde{m}_x =$	16,35
I41	Сум. кв.	33,50
I44	Оц. Дисп.	0,86

Рис.5.17 Визначення оцінки дисперсії в середовищі Excel

- потім визначають значення оцінок параметрів $\tilde{a}$ і $\tilde{b}$ (рис.5.18) за формулами:

$$=H39-\text{КОРЕНЬ}(3)*\text{КОРЕНЬ}(I44)$$

$$=H39+\text{КОРЕНЬ}(3)*\text{КОРЕНЬ}(I44)$$

H47	$\tilde{a}_x =$	14,74
H49	$\tilde{b}_x =$	17,96

Рис. 5.18. Визначення оцінок параметрів $\tilde{a}$ і $\tilde{b}$

5. Для обчислення $\chi^2_{\text{набл}}$ складають таблицю, об'єднавши інтервали з $m_i < 5$ (рис.5.19).

J _i		m _i	p _i	np _i	(m _i -np _i) ²	(m _i -np _i) ² / np _i
X ⁽ⁱ⁾	X ⁽ⁱ⁺¹⁾					
14,8	15,2	5	0,12	4,98	0,0003	0,0001
15,2	15,6	5	0,12	4,98	0,0003	0,0001
15,6	16	7	0,12	4,98	4,0661	0,8159
16	16,8	10	0,25	9,97	0,0011	0,0001
16,8	17,6	8	0,25	9,97	3,8695	0,3882
17,6	18	5	0,12	4,98	0,0003	0,0001
		40	1,00			$\chi^2_{набл} = 1,2044$

Рис.5.19. Розрахункова таблиця $\chi^2_{набл}$

Ймовірності $p_i = P(X^{(i)} < X < X^{(i+1)})$ обчислюють за формулою:

$$p_i = \frac{X^{(i+1)} - X^{(i)}}{\tilde{b} - \tilde{a}}, \text{ причому, оскільки } \tilde{a} \leq X \leq \tilde{b}, \text{ то } X^{(1)} = \tilde{a}; X^{(6)} = \tilde{b}.$$

Спостережуване значення критерію заносять в комірку I70.

6. Число ступенів свободи $f = k - e - l = 5 - 1 - 1 = 3$; $k = 5$, так як значення 5, 6, 7 мають мале число спостережень (<5) і їх об'єднуємо зі значенням 4.

У середовищі EXCEL критичне значення $\chi^2_{крит}$ визначають по функції
 $=\text{ХИ2ОБР}(0,05;3)$

і критичне значення критерію заносять в комірку I73.

7. У комірку K75 вводять формулу

$=\text{ЕСЛИ}(I70 \leq I73; \text{"При заданном уровне значимости гипотеза } H_0 \text{ не противоречит выборочным значениям"}; \text{"гипотеза } H_0 \text{ отклоняется"})$

Результат розв'язання задачі ілюструється на рис.5.20.

J _i		m _i	p _i	np _i	(m _i -np _i) ²	(m _i -np _i) ² / np _i
X ⁽ⁱ⁾	X ⁽ⁱ⁺¹⁾					
14,8	15,2	5	0,12	4,98	0,0003	0,0001
15,2	15,6	5	0,12	4,98	0,0003	0,0001
15,6	16	7	0,12	4,98	4,0661	0,8159
16	16,8	10	0,25	9,97	0,0011	0,0001
16,8	17,6	8	0,25	9,97	3,8695	0,3882
17,6	18	5	0,12	4,98	0,0003	0,0001
		40	1,00			$\chi^2_{набл} = 1,2044$

$\chi^2_{набл} = 1,20$

$\chi^2_{крит} = 7,81$

При заданном уровне значимости гипотеза H₀ не противоречит выборочным значениям

Рис.5.20. Результат розв'язання задачі в середовищі EXCEL

Висновок: вибірові дані не суперечать гіпотезі про рівномірний розподіл продуктивності даної мартенівської печі.

Приклад 5. Згрупований ряд контролю якості продукції дав наступні результати:

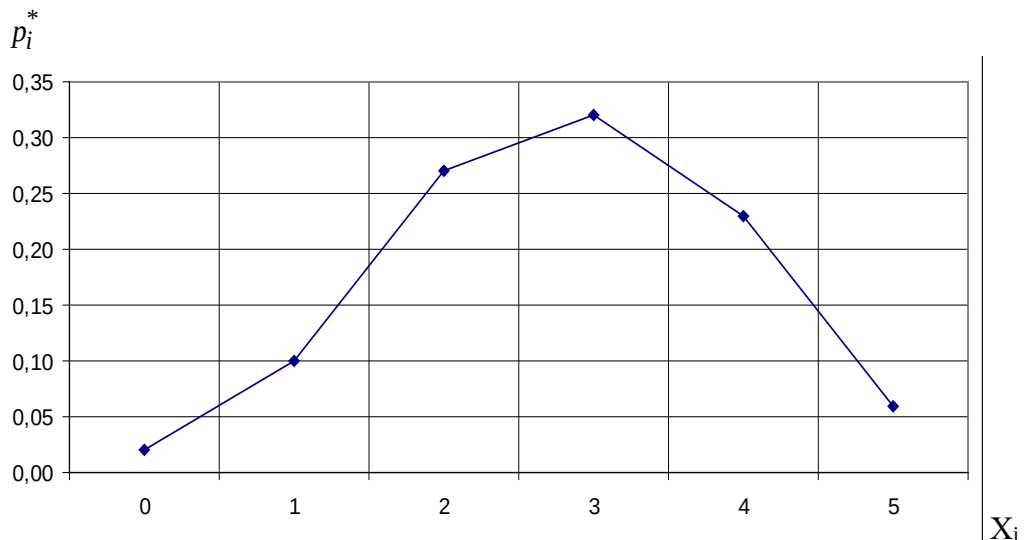
x_i	0	1	2	3	4	5
m_i	2	10	27	32	23	6

Сформулювати гіпотезу про теоретичному розподілі числа бракованих виробів і перевірити ступінь згоди теоретичного та емпіричного розподілів за критерієм χ^2 з рівнем значущості $\alpha = 0,01$.

1. Побудуємо дискретний варіаційний ряд, так як досліджувана ознака є дискретною випадковою величиною. Обчисливши частоти значень x_i , одержимо наступний статистичний ряд:

x_i	0	1	2	3	4	5
m_i	2	10	27	32	23	6
p_i^*	0,02	0,10	0,27	0,32	0,23	0,06

2. Далі будемо полігон відносних частот



Через порівняно невелике число експериментів можна припустити, що має місце біноміальний закон розподілу. Обчислимо:

$$m_x = \frac{0 \cdot 2 + 1 \cdot 10 + 2 \cdot 27 + 3 \cdot 32 + 4 \cdot 23 + 5 \cdot 6}{100} = 2,82 = \bar{X}$$

Висуваємо гіпотезу H_0 : число бракованих виробів має розподіл Бернуллі.

4. Розподіл Бернуллі залежить від одного параметра. Його оцінка $\bar{X} = 2,82$.

5. Обчислимо критерій $\chi^2_{\text{набл}}$, складемо розрахункову таблицю.

Теоретичні ймовірності обчислюємо за формулою

$$p_i = p(x_i) = C_n^{x_i} \cdot \left(\frac{\tilde{m}_x}{n}\right)^{x_i} \cdot \left(1 - \frac{\tilde{m}_x}{n}\right)^{n-x_i}.$$

X_i	m_i	P_i	np_i	$m_i - np_i$	$(m_i - np_i)^2$	$\frac{(m_i - np_i)^2}{np_i}$
1	12	0,17	16,61	-4,61	21,25	1,28
2	27	0,24	23,86	3,14	9,87	0,41
3	32	0,23	22,64	9,38	88,06	3,89
4	23	0,16	15,91	7,09	50,20	3,15
5	6	0,21	21,00	-15	225,00	10,71
Σ	$\Sigma m_i = 100$	$\Sigma P_i = 1$	$\Sigma np_i = 100$			$\chi^2_{набл} = 19,455$

6. Число ступенів свободи $f = k - e - 1 = 5 - 1 - 1 = 3$; $k = 5$, так як значення 0 має мале число спостережень (< 5) і його об'єднують із значенням 1.

7. У таблиці 5.1 значенням $\alpha = 0,01$ і $f = 3$ відповідає $\chi^2_{кр} = 11,34$.

Висновок: так як $\chi^2_{набл} > \chi^2_{крит} (19,455 > 11,34)$, то гіпотеза H_0 не підтвержується.

Застосування критерію χ^2 до інших розподілів відрізняється від наведених тільки способами обчислень величин p_i .

Для рівномірного розподілу: $P_1 = \frac{(X_1 - a)}{b - a}$; $P_i = \frac{X_i - X_{i-1}}{b - a}$; ($i = 2, \dots, k - 1$); $P_k = \frac{b - X_{k-1}}{b - a}$;

для показового розподілу: $P_i = e^{-\hat{\lambda} X_i^*} - e^{-\hat{\lambda} X_{i+1}^*}$; ($i = 1, 2, \dots, k$).

Розв'язання задачі в середовищі EXCEL:



1) У листі 2 завдання 5 формують дискретний статистичний ряд (рис.5. 21).

X_i	0	1	2	3	4	5
m_i	2	10	27	32	23	6
P_i	0,02	0,10	0,27	0,32	0,23	0,06

Рис.5. 21. Дискретний статистичний ряд

2) Викликаючи «Мастер Диаграмм», «Точечная Диаграмма», будують полігон *относительных частот*. Редагують графік таким чином, щоб він був максимально схожий на наведений на рис.5.22.

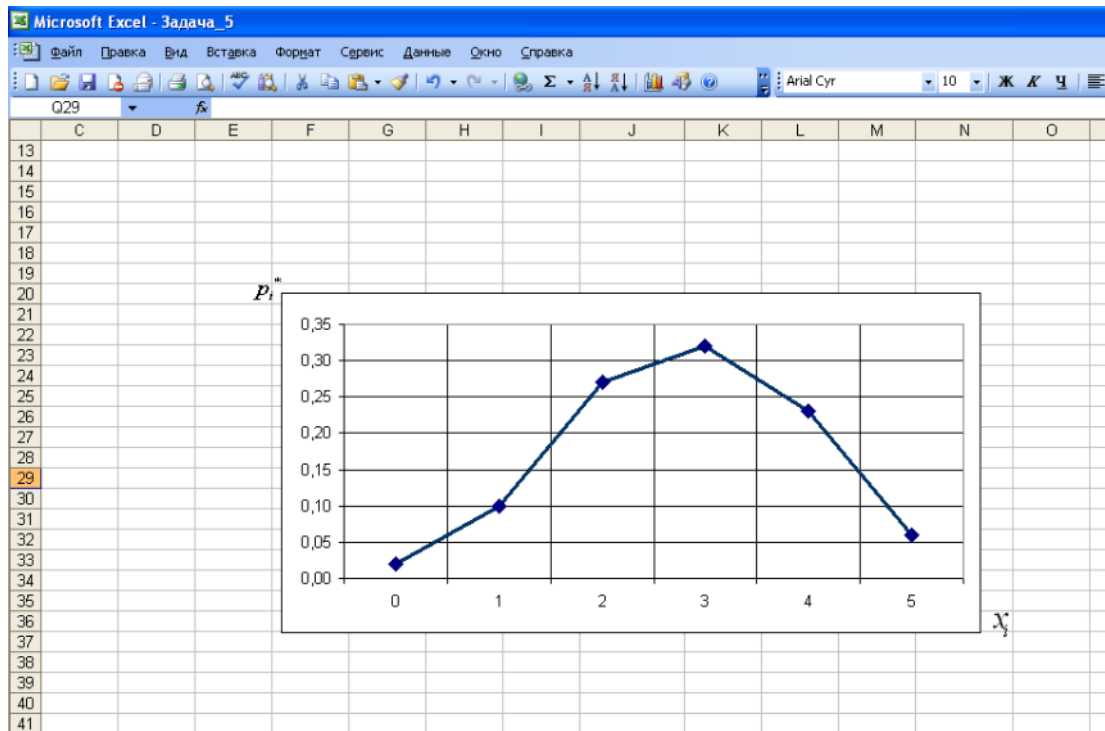


Рис. 5.22. Полігон відносних частот

- 3) Можна припустити, що має місце біноміальний закон розподілу через порівняно невеликі значення випробувань. У комірці J37 обчислюють оцінку математичного очікування. Для оцінки математичного очікування застосовують функцію

$$=\text{СУММПРОИЗВ}(G8:L8;G9:L9)/100$$

- 4) Розподіл Бернуллі залежить від одного параметра, його оцінка $\tilde{m}_x = 2,82 = \bar{X}$.

- 5) Обчислюють критерій $\chi^2_{набл}$, складають при цьому розрахункову таблицю (рис.5.23).

При цьому ймовірність появи події в одному експерименті обчислюється в комірці I40 за формулою:

$$\bar{p} = \frac{\tilde{m}_x}{n}.$$

Теоретичні ймовірності обчислюються в комірках J46: J50 на підставі застосування функції:

$$=\text{БИНОМРАСП}(H46;100;0,0282;\text{ЛОЖЬ}) ,$$

яка заноситься в комірку H46 і далі по стовпцю тиражується вниз (рис.5.23).

X_i	m_i	P_i	nP_i	$(m_i - np_i)$	$(m_i - np_i)^2$	$\frac{(m_i - np_i)^2}{np_i}$
1	12	0.17	16.61	-4.61	21.25	1.28
2	27	0.24	23.86	3.14	9.87	0.41
3	32	0.23	22.62	9.38	88.06	3.89
4	23	0.16	15.91	7.09	50.20	3.15
5	6	0.21	21.00	-15.00	225.00	10.71
	100	1.00	100.00		$\chi^2_{набл} = 19.455$	

Рис.5.23. Розрахункова таблиця обчислення $\chi^2_{набл.}$.

Спостережуване значення критерію заносять в комірку K61.

6) Число ступенів свободи $f = k - e - 1 = 5 - 1 - 1 = 3$; $k = 5$, так як значення 0 має мале число спостережень (< 5) і його об'єднують із значенням 1.

У середовищі EXCEL критичне значення $\chi^2_{крит}$ визначають по функції

$$=ХИ2ОБР(0,01;3)$$

і заносять його до комірки K63.

7) У комірку J67 вводять формулу

=ЕСЛИ(K60<=K63;"При заданном уровне значимости гипотеза H_0 не противоречит выборочным значениям"; "гипотеза H_0 отклоняется")

Результат розв'язання задачі ілюструється на рис.5.24.

X_i	m_i	P_i	nP_i	$(m_i - np_i)$	$(m_i - np_i)^2$	$\frac{(m_i - np_i)^2}{np_i}$
1	12	0.17	16.61	-4.61	21.25	1.28
2	27	0.24	23.86	3.14	9.87	0.41
3	32	0.23	22.62	9.38	88.06	3.89
4	23	0.16	15.91	7.09	50.20	3.15
5	6	0.21	21.00	-15.00	225.00	10.71
	100	1.00	100.00		$\chi^2_{набл} = 19.455$	

$\chi^2_{крит} = 11.34$

гипотеза H_0 отклоняется

Рис.5.24. Результат розв'язання задачі у середовищі EXCEL

Варіанти індивідуальних завдань.

Завдання 5.1.

Сформулювати гіпотезу про закон розподілу неперервної випадкової величини X по даній вибірці і перевірити її за критерієм χ^2 при рівні значущості $\alpha = 0,05$:

5.1.1.	I_i	[5;7)[7;9)[9;11)[11;13)[13;15)[15;17)[17;19)[19;21)[21;23)[23;25)[25;27)[27;29)[29;31)
	m_i	2 4 12 24 25 32 24 23 22 20 8 3 1
5.1.2.	I_i	[0;5)[5;10)[10;15)[15;20)[20;25)[25;30)[30;35)[35;40)[40;45)[45;50)
	m_i	7 11 15 24 49 41 26 17 7 3
5.1.3.	I_i	[-2;-1)[-1;0)[0;1)[1;2)[2;3)[3;4)[4;5)[5;6)[6;7)[7;8)[8;9)[9;10)
	m_i	11 13 14 10 12 9 12 15 10 11 13 10
5.1.4.	I_i	[-2;-1,5)[-1,5;-1)[-1;-0,5)[-0,5;0)[0;0,5)[0,5;1)[1;1,5)[1,5;2)[2;2,5)[2,5;3)
	m_i	15 20 16 18 17 19 21 15 19 20
5.1.5.	I_i	[5;6)[6;7)[7;8)[8;9)[9;10)[10;11)[11;12)[12;13)[13;14)[14;15)
	m_i	61 56 49 33 23 19 16 9 4 2
5.1.6.	I_i	[6;16)[16;26)[26;36)[36;46)[46;56)[56;66)[66;76)[76;86)
	m_i	35 16 5 8 8 7 6 5
5.1.7.	I_i	[5;10)[10;15)[15;20)[20;25)[25;30)[30;35)[35;40)[40;45)[45;50)
	m_i	7 8 15 18 23 19 14 10 6
5.1.8.	I_i	[2;4)[4;6)[6;8)[8;10)[10;12)[12;14)[14;16)[16;18)[18;20)[20;22)[22;24)
	m_i	6 9 26 25 30 26 21 24 20 8 5
5.1.9.	I_i	[1;3)[3;5)[5;7)[7;9)[9;11)[11;13)[13;15)[15;17)[17;19)
	m_i	15 26 25 24 26 27 24 20 13
5.1.10.	I_i	[8;18)[18;28)[28;38)[38;48)[48;58)[58;68)[68;78)[78;88)
	m_i	10 12 9 15 18 16 8 12
5.1.11.	I_i	[-4;-3)[-3;-2)[-2;-1)[-1;0)[0;1)[1;2)[2;3)[3;4)
	m_i	133 120 88 72 46 25 10 6
5.1.12.	I_i	[3;3,5)[3,5;4)[4;4,5)[4,5;5)[5;5,5)[5,5;6)[6;6,5)
	m_i	43 35 22 15 8 5 2
5.1.13.	I_i	[-10;0)[0;10)[10;20)[20;30)[30;40)[40;50)[50;60)
	m_i	48 47 49 40 40 46 30
5.1.14.	I_i	[4;7)[7;10)[10;13)[13;16)[16;19)[19;22)[22;25)[25;28)[28;31)
	m_i	3 4 13 23 22 29 29 16 11
5.1.15.	I_i	[0,2;0,4)[0,4;0,6)[0,6;0,8)[0,8;1,0)[1,0;1,2)[1,2;1,4)[1,4;1,6)[1,6;1,8)[1,8;2,0)[2,0;2,2)
	m_i	35 26 26 25 24 21 20 9 8 6
5.1.16.	I_i	[5;6)[6;7)[7;8)[8;9)[9;10)[10;11)[11;12)[12;13)[13;14)[14;15)
	m_i	61 56 49 33 23 19 16 9 4 2
5.1.17.	I_i	[6;16)[16;26)[26;36)[36;46)[46;56)[56;66)[66;76)[76;86)
	m_i	35 16 5 8 8 7 6 5
5.1.18.	I_i	[5;10)[10;15)[15;20)[20;25)[25;30)[30;35)[35;40)[40;45)[45;50)
	m_i	7 8 15 18 23 19 14 10 6

5.1.19.	$\frac{I_i}{m_i} \mid \begin{array}{cccccccccccc} [2;4)[4;6)[6;8)[8;10)[10;12)[12;14)[14;16)[16;18)[18;20)[20;22)[22;24) \\ 6 \quad 9 \quad 26 \quad 25 \quad 30 \quad 26 \quad 21 \quad 24 \quad 20 \quad 8 \quad 5 \end{array}$
5.1.20.	$\frac{I_i}{m_i} \mid \begin{array}{cccccccc} [1;3)[3;5)[5;7)[7;9)[9;11)[11;13)[13;15)[15;17)[17;19) \\ 15 \quad 26 \quad 25 \quad 24 \quad 26 \quad 27 \quad 24 \quad 20 \quad 13 \end{array}$
5.1.21.	$\frac{I_i}{m_i} \mid \begin{array}{ccccccccccccccc} [5;7)[7;9)[9;11)[11;13)[13;15)[15;17)[17;19)[19;21)[21;23)[23;25)[25;27)[27;29)[29;31) \\ 2 \quad 4 \quad 12 \quad 24 \quad 25 \quad 32 \quad 24 \quad 23 \quad 22 \quad 20 \quad 8 \quad 3 \quad 1 \end{array}$
5.1.22.	$\frac{I_i}{m_i} \mid \begin{array}{cccccccccc} [0;5)[5;10)[10;15)[15;20)[20;25)[25;30)[30;35)[35;40)[40;45)[45;50) \\ 7 \quad 11 \quad 15 \quad 24 \quad 49 \quad 41 \quad 26 \quad 17 \quad 7 \quad 3 \end{array}$
5.1.23.	$\frac{I_i}{m_i} \mid \begin{array}{cccccccccccc} [-2;-1)[-1;0)[0;1)[1;2)[2;3)[3;4)[4;5)[5;6)[6;7)[7;8)[8;9)[9;10) \\ 11 \quad 13 \quad 14 \quad 10 \quad 12 \quad 9 \quad 12 \quad 15 \quad 10 \quad 11 \quad 13 \quad 10 \end{array}$
5.1.24.	$\frac{I_i}{m_i} \mid \begin{array}{cccccccccccc} [-2;-1,5)[-1,5;-1)[-1;-0,5)[-0,5;0)[0;0,5)[0,5;1)[1,1,5)[1,5;2)[2,2,5)[2,5;3) \\ 15 \quad 20 \quad 16 \quad 18 \quad 17 \quad 19 \quad 21 \quad 15 \quad 19 \quad 20 \end{array}$
5.1.25.	$\frac{I_i}{m_i} \mid \begin{array}{cccccccc} [8;18)[18;28)[28;38)[38;48)[48;58)[58;68)[68;78)[78;88) \\ 10 \quad 12 \quad 9 \quad 15 \quad 18 \quad 16 \quad 8 \quad 12 \end{array}$
5.1.26.	$\frac{I_i}{m_i} \mid \begin{array}{cccccccc} [-4;-3)[-3;-2)[-2;-1)[-1;0)[0;1)[1;2)[2;3)[3;4) \\ 133 \quad 120 \quad 88 \quad 72 \quad 46 \quad 25 \quad 10 \quad 6 \end{array}$
5.1.27.	$\frac{I_i}{m_i} \mid \begin{array}{ccccccc} [3;3,5)[3,5;4)[4;4,5)[4,5;5)[5,5;5)[5,5;6)[6,6,5) \\ 43 \quad 35 \quad 22 \quad 15 \quad 8 \quad 5 \quad 2 \end{array}$
5.1.28.	$\frac{I_i}{m_i} \mid \begin{array}{ccccccc} [-10;0)[0;10)[10;20)[20;30)[30;40)[40;50)[50;60) \\ 48 \quad 47 \quad 49 \quad 40 \quad 40 \quad 46 \quad 30 \end{array}$
5.1.29.	$\frac{I_i}{m_i} \mid \begin{array}{ccccccccc} [4;7)[7;10)[10;13)[13;16)[16;19)[19;22)[22;25)[25;28)[28;31) \\ 3 \quad 4 \quad 13 \quad 23 \quad 22 \quad 29 \quad 29 \quad 16 \quad 11 \end{array}$
5.1.30.	$\frac{I_i}{m_i} \mid \begin{array}{ccccccccccccccc} [0,2;0,4)[0,4;0,6)[0,6;0,8)[0,8;1,0)[1,0;1,2)[1,2;1,4)[1,4;1,6)[1,6;1,8)[1,8;2,0)[2,0;2,2) \\ 35 \quad 26 \quad 26 \quad 25 \quad 24 \quad 21 \quad 20 \quad 9 \quad 8 \quad 6 \end{array}$
5.1.31.	$\frac{I_i}{m_i} \mid \begin{array}{ccccccccccccccc} [5;7)[7;9)[9;11)[11;13)[13;15)[15;17)[17;19)[19;21)[21;23)[23;25)[25;27)[27;29)[29;31) \\ 2 \quad 4 \quad 12 \quad 24 \quad 25 \quad 32 \quad 24 \quad 23 \quad 22 \quad 20 \quad 8 \quad 3 \quad 1 \end{array}$
5.1.32.	$\frac{I_i}{m_i} \mid \begin{array}{cccccccc} [0;5)[5;10)[10;15)[15;20)[20;25)[25;30)[30;35)[35;40)[40;45)[45;50) \\ 7 \quad 11 \quad 15 \quad 24 \quad 49 \quad 41 \quad 26 \quad 17 \quad 7 \quad 3 \end{array}$
5.1.33.	$\frac{I_i}{m_i} \mid \begin{array}{cccccccccccc} [-2;-1)[-1;0)[0;1)[1;2)[2;3)[3;4)[4;5)[5;6)[6;7)[7;8)[8;9)[9;10) \\ 11 \quad 13 \quad 14 \quad 10 \quad 12 \quad 9 \quad 12 \quad 15 \quad 10 \quad 11 \quad 13 \quad 10 \end{array}$
5.1.34.	$\frac{I_i}{m_i} \mid \begin{array}{cccccccccccc} [-2;-1,5)[-1,5;-1)[-1;-0,5)[-0,5;0)[0,0,5)[0,5;1)[1,1,5)[1,5;2)[2,2,5)[2,5;3) \\ 15 \quad 20 \quad 16 \quad 18 \quad 17 \quad 19 \quad 21 \quad 15 \quad 19 \quad 20 \end{array}$
5.1.35.	$\frac{I_i}{m_i} \mid \begin{array}{cccccccccc} [5;6)[6;7)[7;8)[8;9)[9;10)[10;11)[11;12)[12;13)[13;14)[14;15) \\ 61 \quad 56 \quad 49 \quad 33 \quad 23 \quad 19 \quad 16 \quad 9 \quad 4 \quad 2 \end{array}$
5.1.36.	$\frac{I_i}{m_i} \mid \begin{array}{ccccccc} [6;16)[16;26)[26;36)[36;46)[46;56)[56;66)[66;76)[76;86) \\ 35 \quad 16 \quad 5 \quad 8 \quad 8 \quad 7 \quad 6 \quad 5 \end{array}$
5.1.37.	$\frac{I_i}{m_i} \mid \begin{array}{cccccccc} [5;10)[10;15)[15;20)[20;25)[25;30)[30;35)[35;40)[40;45)[45;50) \\ 7 \quad 8 \quad 15 \quad 18 \quad 23 \quad 19 \quad 14 \quad 10 \quad 6 \end{array}$

5.1.38.	I_i	[2;4)[4;6)[6;8)[8;10)[10;12)[12;14)[14;16)[16;18)[18;20)[20;22)[22;24)
	m_i	6 9 26 25 30 26 21 24 20 8 5
5.1.39.	I_i	[1;3)[3;5)[5;7)[7;9)[9;11)[11;13)[13;15)[15;17)[17;19)
	m_i	15 26 25 24 26 27 24 20 13
5.1.40.	I_i	[8;18)[18;28)[28;38)[38;48)[48;58)[58;68)[68;78)[78;88)
	m_i	10 12 9 15 18 16 8 12
5.1.41.	I_i	[1;3)[3;5)[5;7)[7;9)[9;11)[11;13)[13;15)[15;17)[17;19)
	m_i	15 26 25 24 26 27 24 20 13
5.1.42.	I_i	[5;7)[7;9)[9;11)[11;13)[13;15)[15;17)[17;19)[19;21)[21;23)[23;25)[25;27)[27;29)[29;31)
	m_i	2 4 12 24 25 32 24 23 22 20 8 3 1
5.1.43.	I_i	[0;5)[5;10)[10;15)[15;20)[20;25)[25;30)[30;35)[35;40)[40;45)[45;50)
	m_i	7 11 15 24 49 41 26 17 7 3
5.1.44.	I_i	[-2;-1)[-1;0)[0;1)[1;2)[2;3)[3;4)[4;5)[5;6)[6;7)[7;8)[8;9)[9;10)
	m_i	11 13 14 10 12 9 12 15 10 11 13 10
5.1.45.	I_i	[-2;-1,5)[-1,5;-1)[-1;-0,5)[-0,5;0)[0;0,5)[0,5;1)[1;1,5)[1,5;2)[2;2,5)[2,5;3)
	m_i	15 20 16 18 17 19 21 15 19 20
5.1.46.	I_i	[8;18)[18;28)[28;38)[38;48)[48;58)[58;68)[68;78)[78;88)
	m_i	10 12 9 15 18 16 8 12
5.1.47.	I_i	[-4;-3)[-3;-2)[-2;-1)[-1;0)[0;1)[1;2)[2;3)[3;4)
	m_i	133 120 88 72 46 25 10 6
5.1.48.	I_i	[3;3,5)[3,5;4)[4;4,5)[4,5;5)[5;5,5)[5,5;6)[6;6,5)
	m_i	43 35 22 15 8 5 2
5.1.49.	I_i	[-10;0)[0;10)[10;20)[20;30)[30;40)[40;50)[50;60)
	m_i	48 47 49 40 40 46 30
5.1.50.	I_i	[4;7)[7;10)[10;13)[13;16)[16;19)[19;22)[22;25)[25;28)[28;31)
	m_i	3 4 13 23 22 29 29 16 11

Завдання 5.2.

За допомогою критерію χ^2 перевірити, чи узгоджуються вибіркові дані з гіпотезою про те, що дискретна випадкова величина X розподілена за законом Пуассона ($\alpha = 0,05$):

5.2.1.	X_i	0 1 2 3 4 5 <i>и больше</i>
	m_i	229 211 93 35 7 1

5.2.2.	X_i	0 1 2 3 4
	m_i	116 56 22 4 2

5.2.3.	X_i	0 1 2 3 4
	m_i	132 43 20 3 2

5.2.4.	X_i	0 1 2 3 4 5 6
	m_i	405 366 175 40 8 4 2

5.2.5.

X_i	0	1	2	3	4	5
m_i	505	336	125	24	8	2

5.2.6.

X_i	0	1	2	3	4	5	6	7
m_i	199	169	87	31	9	3	1	1

5.2.7.

X_i	0	1	2	3	4	5
m_i	65	50	35	30	12	8

5.2.8.

X_i	0	1	2	3	4	5
m_i	125	110	95	80	50	40

5.2.9.

X_i	0	1	2	3	4	5
m_i	65	50	35	30	12	8

5.2.10.

X_i	0	1	2	3	4	5
m_i	125	110	95	80	50	40

5.2.11.

X_i	0	1	2	3	4	5
m_i	253	151	55	17	6	2

5.2.12.

X_i	0	1	2	3	4	5
m_i	308	164	95	14	12	3

5.2.13.

X_i	0	1	2	3	4	5
m_i	185	64	21	16	5	2

5.2.14.

X_i	0	1	2	3	4	5
m_i	330	141	54	19	7	3

5.2.15.

X_i	0	1	2	3	4	5
m_i	233	128	61	32	13	5

5.2.16.

X_i	0	1	2	3	4	5
m_i	114	98	56	21	8	2

5.2.17.

X_i	0	1	2	3	4	5
m_i	96	84	51	23	11	4

5.2.18.

X_i	0	1	2	3	4	5
m_i	67	54	19	7	3	1

5.2.19.

X_i	0	1	2	3	4	5
m_i	67	54	32	19	5	2

5.2.20.

X_i	0	1	2	3	4	5
m_i	63	54	41	22	15	4

5.2.21.

X_i	0	1	2	3	4	5
m_i	17	19	11	6	4	1

5.2.22.

X_i	0	1	2	3	4	5
m_i	144	95	62	29	11	3

5.2.23.

X_i	0	1	2	3	4	5
m_i	49	25	12	5	2	1

5.2.24.

X_i	0	1	2	3	4	5
m_i	185	64	21	16	5	2

5.2.25.

X_i	0	1	2	3	4	5	6
m_i	405	366	175	40	8	4	2

Перевірити гіпотезу про біноміальний розподіл дискретної випадкової величини X за критерієм χ^2 при рівні значущості $\alpha = 0,01$:

5.2.26.

X_i	0	1	2	3	4	5
m_i	2	10	27	32	23	6

5.2.27.

X_i	0	1	2	3	4	5	6	7
m_i	2	3	10	22	26	20	12	5

5.2.28.

X_i	0	1	2	3	4	5
m_i	72	77	34	14	2	1

5.2.29.

X_i	0	1	2	3	4
m_i	60	29	4	3	4

5.2.30.

X_i	0	1	2	3	4	5
m_i	10	10	30	10	25	15

За допомогою критерію χ^2 перевірити, чи узгоджуються вибіркові дані з гіпотезою про те, що дискретна випадкова величина X розподілена за законом Пуассона ($\alpha = 0,05$):

5.2.31.

X_i	0	1	2	3	4	5
m_i	5	16	25	17	20	17

5.2.32.

X_i	0	1	2	3	4	5
m_i	1	3	5	18	13	10

5.2.33.

X_i	0	1	2	3	4	5
m_i	72	77	34	14	2	1

5.2.34.

X_i	0	1	2	3	4
m_i	60	29	4	3	4

5.2.35.

X_i	0	1	2	3	4	5
m_i	2	5	13	48	16	9

5.2.36.

X_i	0	1	2	3	4	5
m_i	35	64	29	14	6	4

5.2.37.

X_i	0	1	2	3	4	5
m_i	4	8	22	54	13	11

5.2.38.

X_i	0	1	2	3	4	5
m_i	9	14	21	58	23	15

5.2.39.

X_i	0	1	2	3	4	5
m_i	5	14	33	78	24	9

5.2.40.

X_i	0	1	2	3	4	5
m_i	5	12	32	77	24	8

5.2.41.

X_i	0	1	2	3	4	5
m_i	15	76	54	24	9	2

5.2.42.

X_i	0	1	2	3	4	5
m_i	23	58	123	67	31	14

5.2.43.

X_i	0	1	2	3	4	5
m_i	9	18	26	12	4	1

5.2.44.

X_i	0	1	2	3	4	5
m_i	26	42	57	23	14	2

5.2.45.

X_i	0	1	2	3	4	5
m_i	35	64	29	14	6	4

5.2.46.

X_i	0	1	2	3	4	5
m_i	14	47	62	55	25	9

5.2.47.

X_i	0	1	2	3	4	5
m_i	6	18	42	21	9	3

5.2.48.

X_i	0	1	2	3	4	5
m_i	49	25	12	5	2	1

5.2.49.

X_i	0	1	2	3	4	5
m_i	5	16	25	17	20	17

5.2.50.

X_i	0	1	2	3	4	5
m_i	2	5	13	48	16	9

Таблиця 5.1

Таблиця критичних точок розподілу Пірсона «хі-квадрат»

k / α	0,01	0,025	0,05	0,95	0,975	0,99
1	6,63	5,02	3,84	0,00	0,00	0,00
2	9,21	7,38	5,99	0,10	0,05	0,02
3	11,34	9,35	7,81	0,35	0,22	0,11
4	13,28	11,14	9,49	0,71	0,48	0,30
5	15,09	12,83	11,07	1,15	0,83	0,55
6	16,81	14,45	12,59	1,64	1,24	0,87
7	18,48	16,01	14,07	2,17	1,69	1,24
8	20,09	17,53	15,51	2,73	2,18	1,65
9	21,67	19,02	16,92	3,33	2,70	2,09
10	23,21	20,48	18,31	3,94	3,25	2,56

11	24,72	21,92	19,68	4,57	3,82	3,05
12	26,22	23,34	21,03	5,23	4,40	3,57
13	27,69	24,74	22,36	5,89	5,01	4,11
14	29,14	26,12	23,68	6,57	5,63	4,66
15	30,58	27,49	25,00	7,26	6,26	5,23
16	32,00	28,85	26,30	7,96	6,91	5,81
17	33,41	30,19	27,59	8,67	7,56	6,41
18	34,81	31,53	28,87	9,39	8,23	7,01
19	36,19	32,85	30,14	10,12	8,91	7,63
20	37,57	34,17	31,41	10,85	9,59	8,26
21	38,93	35,48	32,67	11,59	10,28	8,90
22	40,29	36,78	33,92	12,34	10,98	9,54
23	41,64	38,08	35,17	13,09	11,69	10,20
24	42,98	39,36	36,42	13,85	12,40	10,86
25	44,31	40,65	37,65	14,61	13,12	11,52
26	45,64	41,92	38,89	15,38	13,84	12,20
27	46,96	43,19	40,11	16,15	14,57	12,88
28	48,28	44,46	41,34	16,93	15,31	13,56
29	49,59	45,72	42,56	17,71	16,05	14,26
30	50,89	46,98	43,77	18,49	16,79	14,95
31	52,19	48,23	44,99	19,28	17,54	15,66
32	53,49	49,48	46,19	20,07	18,29	16,36
33	54,78	50,73	47,40	20,87	19,05	17,07
34	56,06	51,97	48,60	21,66	19,81	17,79
35	57,34	53,20	49,80	22,47	20,57	18,51
36	58,62	54,44	51,00	23,27	21,34	19,23
37	59,89	55,67	52,19	24,07	22,11	19,96
38	61,16	56,90	53,38	24,88	22,88	20,69
39	62,43	58,12	54,57	25,70	23,65	21,43
40	63,69	59,34	55,76	26,51	24,43	22,16
41	64,95	60,56	56,94	27,33	25,21	22,91
42	66,21	61,78	58,12	28,14	26,00	23,65
43	67,46	62,99	59,30	28,96	26,79	24,40
44	68,71	64,20	60,48	29,79	27,57	25,15
45	69,96	65,41	61,66	30,61	28,37	25,90
46	71,20	66,62	62,83	31,44	29,16	26,66
47	72,44	67,82	64,00	32,27	29,96	27,42
48	73,68	69,02	65,17	33,10	30,75	28,18
49	74,92	70,22	66,34	33,93	31,55	28,94
50	76,15	71,42	67,50	34,76	32,36	29,71

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$$

Таблица 5.2

x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)	x	Φ(x)
0,00	0,00000	0,50	0,19146	1,00	0,34134	1,50	0,43319	2,00	0,47725	3,00	0,49865
0,01	0,00399	0,51	0,19497	1,01	0,34375	1,51	0,43448	2,02	0,47831	3,05	0,49886
0,02	0,00798	0,52	0,19847	1,02	0,34614	1,52	0,43574	2,04	0,47932	3,10	0,49903
0,03	0,01197	0,53	0,20194	1,03	0,34849	1,53	0,43699	2,06	0,48030	3,15	0,49918
0,04	0,01595	0,54	0,20540	1,04	0,35083	1,54	0,43822	2,08	0,48124	3,20	0,49931
0,05	0,01994	0,55	0,20884	1,05	0,35314	1,55	0,43943	2,10	0,48214	3,25	0,49942
0,06	0,02392	0,56	0,21226	1,06	0,35543	1,56	0,44062	2,12	0,48300	3,30	0,49952
0,07	0,02790	0,57	0,21566	1,07	0,35769	1,57	0,44179	2,14	0,48382	3,35	0,49960
0,08	0,03188	0,58	0,21904	1,08	0,35993	1,58	0,44295	2,16	0,48461	3,40	0,49966
0,09	0,03586	0,59	0,22240	1,09	0,36214	1,59	0,44408	2,18	0,48537	3,45	0,49972
0,10	0,03983	0,60	0,22575	1,10	0,36433	1,60	0,44520	2,20	0,48610	3,50	0,49977
0,11	0,04380	0,61	0,22907	1,11	0,36650	1,61	0,44630	2,22	0,48679	3,55	0,49981
0,12	0,04776	0,62	0,23237	1,12	0,36864	1,62	0,44738	2,24	0,48745	3,60	0,49984
0,13	0,05172	0,63	0,23565	1,13	0,37076	1,63	0,44845	2,26	0,48809	3,65	0,49987
0,14	0,05567	0,64	0,23891	1,14	0,37286	1,64	0,44950	2,28	0,48870	3,70	0,49989
0,15	0,05962	0,65	0,24215	1,15	0,37493	1,65	0,45053	2,30	0,48928	3,75	0,49991
0,16	0,06356	0,66	0,24537	1,16	0,37698	1,66	0,45154	2,32	0,48983	3,80	0,49993
0,17	0,06749	0,67	0,24857	1,17	0,37900	1,67	0,45254	2,34	0,49036	3,85	0,49994
0,18	0,07142	0,68	0,25175	1,18	0,38100	1,68	0,45352	2,36	0,49086	3,90	0,49995
0,19	0,07535	0,69	0,25490	1,19	0,38298	1,69	0,45449	2,38	0,49134	3,95	0,49996
0,20	0,07926	0,70	0,25804	1,20	0,38493	1,70	0,45543	2,40	0,49180	4,00	0,49997
0,21	0,08317	0,71	0,26115	1,21	0,38686	1,71	0,45637	2,42	0,49224	4,05	0,49997
0,22	0,08706	0,72	0,26424	1,22	0,38877	1,72	0,45728	2,44	0,49266	4,10	0,49998
0,23	0,09095	0,73	0,26730	1,23	0,39065	1,73	0,45818	2,46	0,49305	4,15	0,49998
0,24	0,09483	0,74	0,27035	1,24	0,39251	1,74	0,45907	2,48	0,49343	4,20	0,49999
0,25	0,09871	0,75	0,27337	1,25	0,39435	1,75	0,45994	2,50	0,49379	4,25	0,49999
0,26	0,10257	0,76	0,27637	1,26	0,39617	1,76	0,46080	2,52	0,49413	4,30	0,49999
0,27	0,10642	0,77	0,27935	1,27	0,39796	1,77	0,46164	2,54	0,49446	4,35	0,49999
0,28	0,11026	0,78	0,28230	1,28	0,39973	1,78	0,46246	2,56	0,49477	4,40	0,49999
0,29	0,11409	0,79	0,28524	1,29	0,40147	1,79	0,46327	2,58	0,49506	4,45	0,50000
0,30	0,11791	0,80	0,28814	1,30	0,40320	1,80	0,46407	2,60	0,49534	4,50	0,50000

0,31	0,12172	0,81	0,29103	1,31	0,40490	1,81	0,46485	2,62	0,49560	4,55	0,50000
0,32	0,12552	0,82	0,29389	1,32	0,40658	1,82	0,46562	2,64	0,49585	4,60	0,50000
0,33	0,12930	0,83	0,29673	1,33	0,40824	1,83	0,46638	2,66	0,49609	4,65	0,50000
0,34	0,13307	0,84	0,29955	1,34	0,40988	1,84	0,46712	2,68	0,49632	4,70	0,50000
0,35	0,13683	0,85	0,30234	1,35	0,41149	1,85	0,46784	2,70	0,49653	4,75	0,50000
0,36	0,14058	0,86	0,30511	1,36	0,41309	1,86	0,46856	2,72	0,49674	4,80	0,50000
0,37	0,14431	0,87	0,30785	1,37	0,41466	1,87	0,46926	2,74	0,49693	4,85	0,50000
0,38	0,14803	0,88	0,31057	1,38	0,41621	1,88	0,46995	2,76	0,49711	4,90	0,50000
0,39	0,15173	0,89	0,31327	1,39	0,41774	1,89	0,47062	2,78	0,49728	4,95	0,50000
0,40	0,15542	0,90	0,31594	1,40	0,41924	1,90	0,47128	2,80	0,49744	5,00	0,50000
0,41	0,15910	0,91	0,31859	1,41	0,42073	1,91	0,47193	2,82	0,49760		
0,42	0,16276	0,92	0,32121	1,42	0,42220	1,92	0,47257	2,84	0,49774		
0,43	0,16640	0,93	0,32381	1,43	0,42364	1,93	0,47320	2,86	0,49788		
0,44	0,17003	0,94	0,32639	1,44	0,42507	1,94	0,47381	2,88	0,49801		
0,45	0,17364	0,95	0,32894	1,45	0,42647	1,95	0,47441	2,90	0,49813		
0,46	0,17724	0,96	0,33147	1,46	0,42785	1,96	0,47500	2,92	0,49825		
0,47	0,18082	0,97	0,33398	1,47	0,42922	1,97	0,47558	2,94	0,49836		
0,48	0,18439	0,98	0,33646	1,48	0,43056	1,98	0,47615	2,96	0,49846		
0,49	0,18793	0,99	0,33891	1,49	0,43189	1,99	0,47670	2,98	0,49856		

Задача № 6
Перевірка статистичних гіпотез. Критерій згоди λ (Критерій Колмогорова)
Короткий теоретичний матеріал

За критеріями згоди визначають, чи спричинені розбіжності між теоретичною кривою і статистичним розподілом тільки випадковими обставинами, пов'язаними з обмеженим числом спостережень, або вони є суттєвими і пов'язані з тим, що обрана крива погано вирівнює даний статистичний розподіл. Крім критерію χ^2 - Пірсона на практиці найбільш поширений є критерій Колмогорова. Зауважимо, що у критерії згоди Пірсона порівнюються теоретичні і експериментальні ряди розподілу (для дискретної випадкової величини) або щільність розподілу ймовірностей (для неперервної випадкової величини).

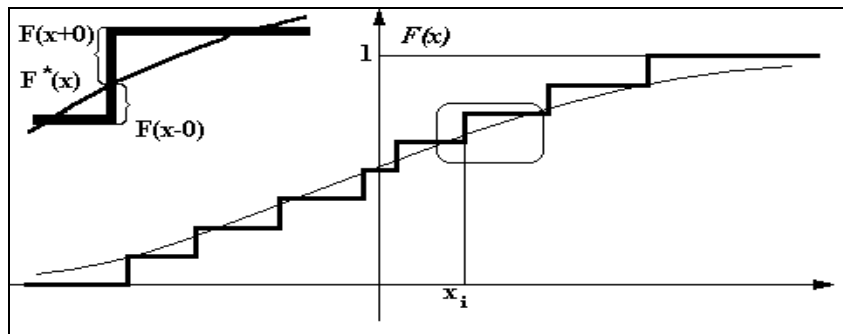
Особливість критерію Колмогорова полягає в тому, що тут порівнюються теоретична $F(x)$ і емпірична (статистична) $F^*(x)$ функції розподілу. Тоді у якості запобіжної розбіжності між теоретичним і емпіричним розподілами розглядають максимальне значення абсолютної величини різниці між емпіричною функцією розподілу $F^*(x)$ і відповідною теоретичною функцією розподілу

$$D = \max_{-\infty < x < \infty} |F^*(x) - F(x)|, \text{ званої } \textit{статистикою критерію Колмогорова}.$$

Для дискретної випадкової ознаки методика перевірки статистичних гіпотез за критерієм Колмогорова має деякі особливості.

Розглянемо їх.

За заданою вибіркою $(x_1, x_2, \dots, x_n)$ будується емпірична функція розподілу $F^*(x)$. Цілком очевидно, що у точках x_i така функція має стрибок, рівний відносній частоті w_i . Отже, у кожній експериментальній точці x_i слід брати два значення - границю зліва $F^*(x_i - 0)$ і границю справа $F^*(x_i + 0)$.



В експериментальних точках x_i підраховують значення теоретичної функції розподілу $F(x_i)$ за передбачуваний формулі. Цілком очевидно, що теоретичні та емпіричні значення не будуть збігатися. Якщо вони розрізняються незначно, то гіпотезу про передбачуваний закон розподілу слід прийняти. Якщо відмінність значна, то гіпотезу треба відкидати.

Методика проведення досліджень:

1. За вибіровими даними знаходиться емпірична функція розподілу $F^*(x)$.
2. Будуються графіки функцій $F(x)$ и $F^*(x)$.
3. Визначається максимальна величина модуля різниці їх ординат

$$D = \max_{-\infty < x < \infty} |F^*(x) - F(x)|.$$

4. Обчислюється величина $\lambda_{набл} = D\sqrt{n}$, де n – об'єм вибірки.

5. По заданому α у таблиці критичних точок критерію λ знаходиться $\alpha_{\text{крит.}}$.

Висновок: якщо $\lambda_{\text{набл}} \leq \lambda_{\text{кр}}$, гіпотеза H_0 приймається, якщо $\lambda_{\text{набл}} > \lambda_{\text{кр}}$, гіпотеза H_0 не приймається.

Зауваження.

1. Критерій Колмогорова застосовуємо у тому випадку, коли параметри теоретичного розподілу відомі заздалегідь, якщо ж вони визначаються за вибіркою, то зростає ймовірність прийняття невірної гіпотези.

2. Критерій згоди Колмогорова в своєму класичному вигляді є більш потужним, ніж критерій χ^2 і може бути використаний для перевірки гіпотези про відповідність емпіричного розподілу будь-якого теоретичного безперервного розподілу $F(x)$ з заздалегідь відомими параметрами.

Дослідження безперервної випадкової ознаки

Приклад 1.

У результаті зважування 100 заготовок отримано наступний емпіричний розподіл

I_i	$[0,5; 2,5)$	$[2,5; 4,5)$	$[4,5; 6,5)$	$[6,5; 8,5)$	$[8,5; 10,5)$
m_i	20	19	20	22	19
p_i^*	0,2	0,19	0,2	0,22	0,19

Потрібно при рівні значущості $\alpha = 0,05$ перевірити гіпотезу про рівномірний закон розподілу ваги заготовок у інтервалі $(0; 10)$.

Так як параметри рівномірного розподілу відомі ($a = 0$; $b = 10$), застосуємо критерій Колмогорова для перевірки узгодженості теоретичного та емпіричного розподілів.

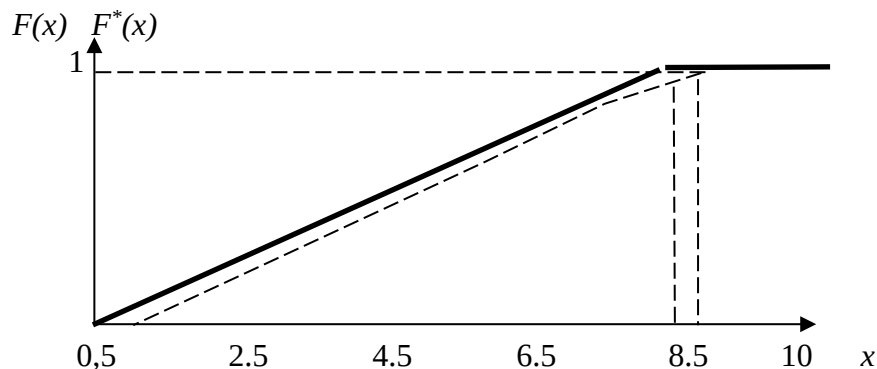
Теоретична функція рівномірного розподілу на інтервалі $(0; 10)$ має вигляд

$$F(x) = \begin{cases} 0; & x < 0, \\ \frac{x}{10}; & 0 \leq x \leq 10, \\ 1; & x > 10. \end{cases}$$

1. Знаходимо емпіричну функцію розподілу $F^*(x)$:

$$F^*(0,5)=0; F^*(2,5)=0,2; F^*(4,5)=0,39; F^*(6,5)=0,59; F^*(8,5)=0,81; F^*(10,5)=1.$$

2. Побудуємо графіки функцій $F(x)$ и $F^*(x)$



3. Визначаємо максимальне відхилення між $F(x)$ і $F^*(x)$

$$D = \max_{-\infty < x < \infty} |F^*(x) - F(x)| = \max\{0,05; 0,05; 0,06; 0,04; 0,05\} = 0,06$$

4. $\lambda_{\text{набл}} = D\sqrt{n} = 0,06 \cdot \sqrt{100} = 0,6.$

5. $\lambda_{\text{крит}} = \lambda_{\text{табл}}(0,05) = 1,358$ (табл.6.1).

Висновок: оскільки $\lambda_{\text{набл}} < \lambda_{\text{кр}} (0,6 < 1,358)$, гіпотеза H_0 про рівномірний розподіл генеральної сукупності приймається.

Розв'язання задачі у середовищі EXCEL:



1. На листі 1 завдання 6 будуюмо інтервальний статистичний ряд (рис.6.1).

Интервальный статистический ряд						
левые границы интервалов	0,5	2,5	4,5	6,5	8,5	10,5
длина интервалов	2					
f_i	[0,5;2,5)	[2,5;4,5)	[4,5;6,5)	[6,5;8,5)	[8,5;10,5)	
m_i	20	19	20	22	19	100
F_i^*	0,2	0,19	0,2	0,22	0,19	1

Рис.6.1. Интервальный статистический ряд

Знаходимо емпіричну функцію розподілу $F^*(x)$:

Эмпирическая функция распределения			
	0	при	$x=0,5$
	0,2	при	$x=2,5$
	0,39	при	$x=4,5$
	0,59	при	$x=6,5$
$F^*(x)=$	0,81	при	$x=8,5$
	1	при	$x=10,5$

Вспомогательная таблица для построения эмпирической функции распределения						
0	0,2	0,39	0,59	0,81	1	1
$x=0,5$	$x=2,5$	$x=4,5$	$x=6,5$	$x=8,5$	$x=10,5$	$x=12,5$

Рис. 6.2. Аналітична формула емпіричної функції розподілу

2. Будуюмо графіки функцій розподілу $F(x)$ и $F^*(x)$

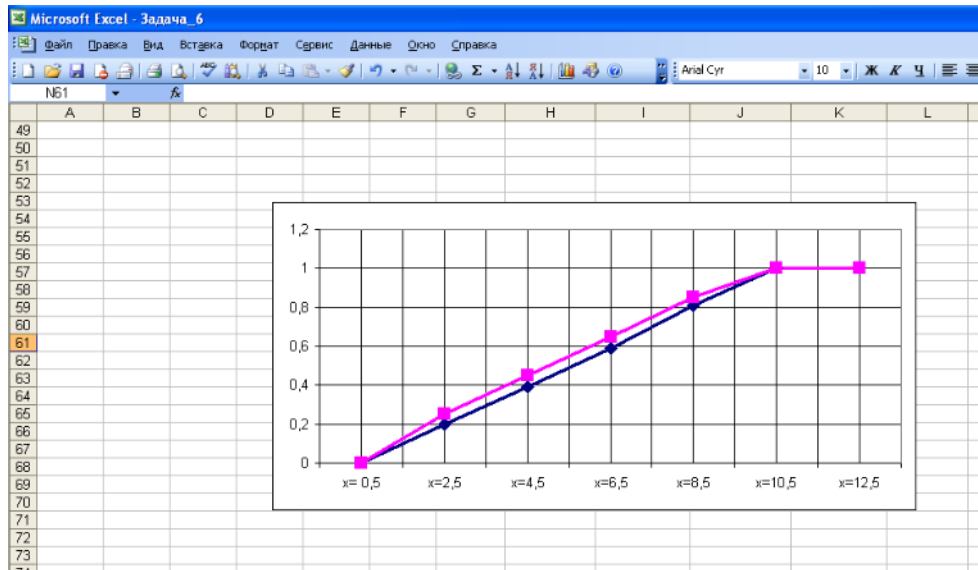


Рис. 6.3. Графіки функцій розподілу

3. Визначаємо максимальне відхилення між $F(x)$ и $F^*(x)$.

Для цього у комірки F52: J52 вводять відповідні різниці і у комірку F54 за допомогою функції

$$=\text{МАКС}(F52:J52)$$

визначають максимальне значення з виділеного діапазону (рис.6.4).

Microsoft Excel - Задача_6												
Файл Правка Вид Вставка Формат Сервис Данные Окно Справка												
G68												
A	B	C	D	E	F	G	H	I	J	K	L	M
37												
38	Вспомогательная таблица для построения эмпирической функции распределения											
39												
40												
41												
42												
43												
44												
45	Вспомогательная таблица для построения теоретической функции распределения											
46												
47												
48												
49												
50												
51												
52												
53												
54												
55												
56												
57												
58												

Рис.6.4. Визначення максимального значення різниці функцій Розподілу

4. У комірці F60 за формулою $\lambda_{набл} = D\sqrt{n}$ обчислюють спостережуване значення критерію Колмогорова.

5. У комірку F62 вносять $\lambda_{крит}$.

6. У комірку B65 вводять функцію

=ЕСЛИ(F60<=F62;"При заданном уровне значимости гипотеза H_0 не противоречит выборочным значениям";"гипотеза H_0 отклоняется")

Відповідні операції відображені на рис.6.5.

Вспомогательная таблица для построения теоретической функции распределения							
0	0,25	0,45	0,65	0,85	1	1	
x=0,5	x=2,5	x=4,5	x=6,5	x=8,5	x=10,5	x=12,5	
D=	0,05	0,06	0,06	0,04	0		
max D=	0,06						
$\lambda_{набл} =$	0,6						
$\lambda_{крит} =$	1,358						
При заданном уровне значимости гипотеза H_0 о равномерном законе распределения принимается							

Рис. 6.5. Формування висновку щодо гіпотези H_0

Дослідження дискретної випадкової ознаки

Приклад 2.

Для задачі 2 завдання 5 виконати перевірку статистичних гіпотез за критерієм Колмогорова.

- Для виконаного експерименту будуюмо таблицю абсолютних частот і теоретичних частот. Теоретичні частоти для закону Пуассона обчислюємо за співвідношенням: $\tilde{m}_i = n \cdot p(x_i)$, де

$$p_i = p(x_i) = \frac{2^{x_i}}{x_i!} e^{-2}; \quad (e^{-2} = 0,1353).$$

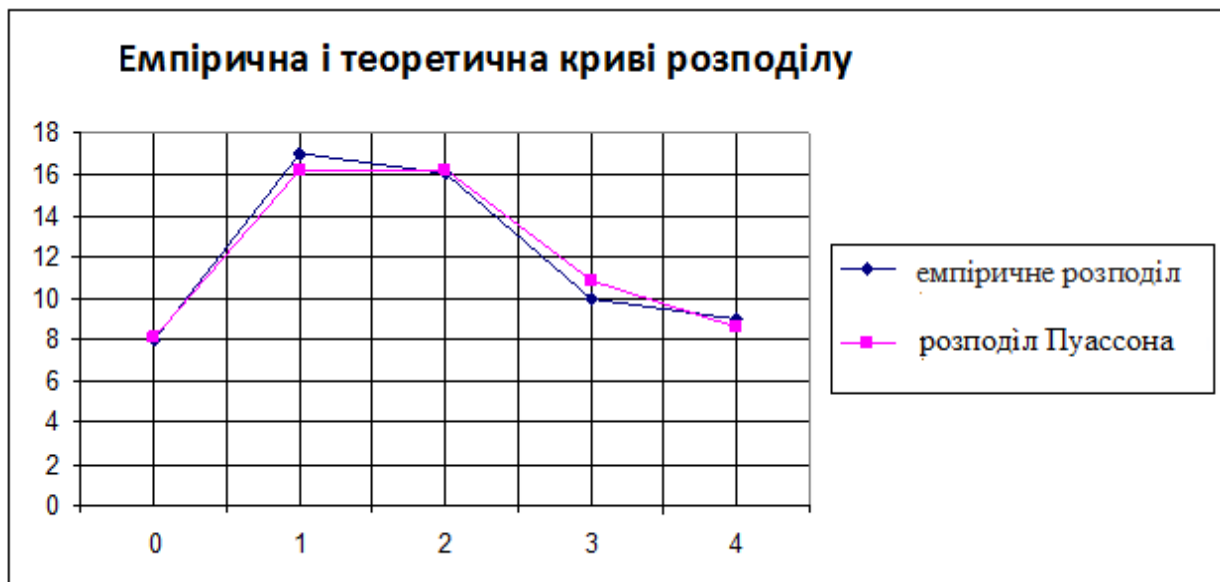
Зауважимо, що для нормального закону розподілу теоретичні частоти обчислюються по співвідношенню: $\tilde{m}_i = n \cdot h_i \cdot \varphi(t)$, де h_i – довжина розмаху i -го інтервалу, $\varphi(t)$ – функція

Гаусса, яка має вигляд:
$$\varphi(t) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{t^2}{2}}, \quad \text{при цьому} \quad t = \frac{x - \bar{x}}{\sigma}.$$

У п'ятому і шостому стовпчиках таблиці вказуємо накопичені частоти

X_i	m_i	P_i	$\tilde{m}_i$	m'_i	$\tilde{m}'_i$	$ m'_i - \tilde{m}'_i $
0	8	0,1353	8,13	8	8,13	0,13
1	17	0,2707	16,24	25	24,37	0,63
2	16	0,2707	16,24	41	40,61	0,39
3	10	0,1804	10,82	51	51,43	0,43
4	9	0,1429	8,57	60	60	0,00
и бо- лее						
	$\sum m_i = 60$	$\sum p_i = 1$	$\sum n_i = 60$			$D_{\max} = 0,63$

2. Для наочності побудуємо емпіричну і теоретичну (розподіл Пуассона) криві розподілу.



3. Розрахунки накопичених частот і різниць між ними показали, що в 2-ій групі спостерігається максимальна розбіжність (різниця) $D = 0,63$.
4. Зауважимо, що критерій Колмогорова λ оснований на визначенні максимальної розбіжності між накопиченими частотами емпіричного і теоретичного розподілу (D) і розраховується за формулою

$$\lambda = D / \sqrt{n}$$

Розрахувавши λ , по таблиці значень критерію Колмогорова $P(\lambda)$ (табл. 6.2), визначають ймовірність, з якою можна стверджувати, що відхилення емпіричних частот від теоретичних випадкові.

Зауважимо, що при $P(\lambda) = 1$ відбувається повний збіг частот, при $P(\lambda) = 0$ - повна розбіжність. Для розглянутого прикладу отримуємо:

$$\lambda = 0,63 / \sqrt{60} = 0,081.$$

5. По таблиці додатків 2 видно, що найбільш близьке до цим розрахунковим λ є значення $\lambda = 0,3$. Таким чином, з імовірністю рівною 1 можна стверджувати справедливості висунутої гіпотези про розподіл випадкової величини за законом Пуассона у генеральній сукупності. Деякі розбіжності між емпіричним і теоретичним розподілами носять випадковий характер.

Розв'язання задачі у середовищі EXCEL:



1. У листі 2 завдання 6 будемо таблицю абсолютних і теоретичних частот (рис.6.6).

Microsoft Excel - Задача_6

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

S25

	D	E	F	G	H	I	J	K	L	M	N	O	P
1													
2													
3													
4													
5													
6													
7													
8													
9													
10													
11													
12													
13													
14													
15													
16													

Таблица абсолютных и теоретических частот

X_i	m_i	P_i	$\tilde{m}_i$	m'_i	$\tilde{m}'_i$	$ m'_i - \tilde{m}'_i $
0	8	0,1353	8,13	8	8,13	0,13
1	17	0,2707	16,24	25	24,37	0,63
2	16	0,2707	16,24	41	40,61	0,39
3	10	0,1804	10,82	51	51,43	0,43
4	9	0,1429	8,57	60	60	0
и более						
$\Sigma m_i =$		$\Sigma p_i = 1$	$\Sigma \tilde{m}_i =$			$D_{\max} = 0,63$
= 60			= 60			

Рис. 6.6. Таблица частот

2. Потім будемо криві розподілу (рис.6.7), *викликаючи «Мастер Диаграмм», «График».*

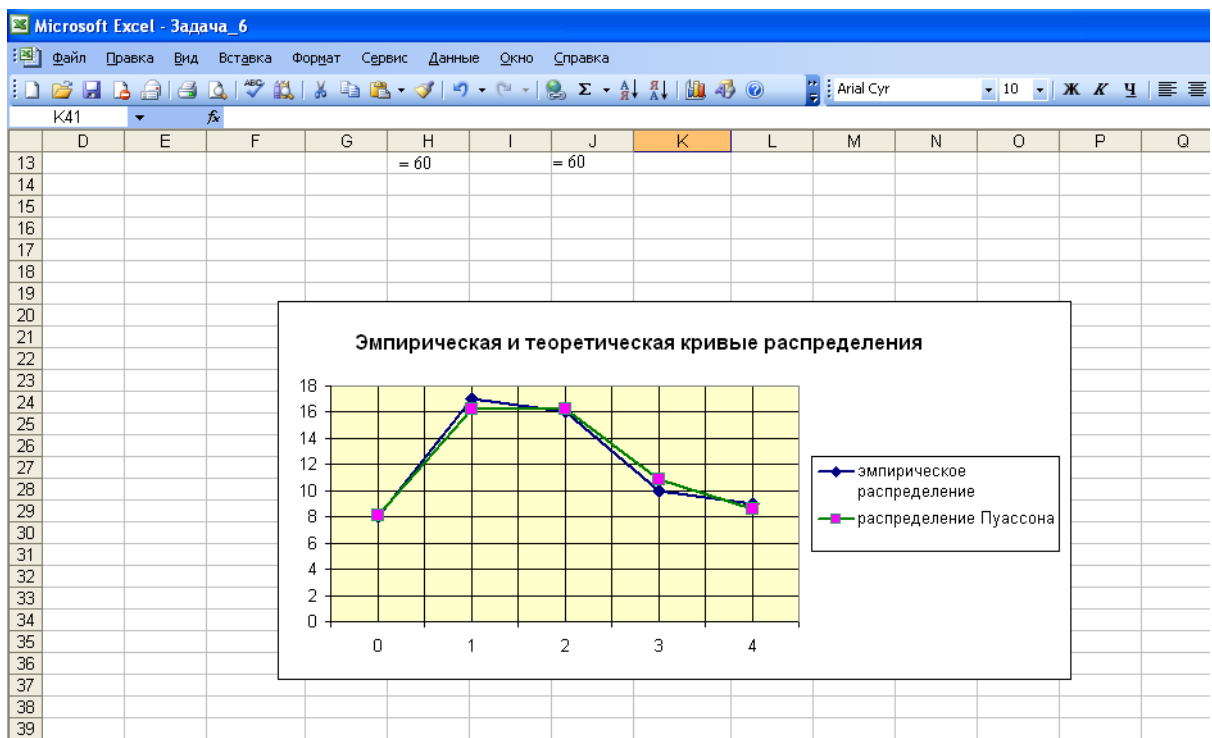


Рис. 6.7. Криві розподілу

3. У комірки I40 і L40 заносимо значення об'єму вибірки і максимальну розбіжність D , а потім у комірку I42 записуємо значення критерію Колмогорова λ (рис.6. 8).

	D	E	F	G	H	I	J	K	L	M
37										
38										
39										
40					n=	60		D=	0,63	
41										
42					λ =	0,08				
43										
44										
45										
46										
47										

Рис.6.8 Розрахунок значення критерію Колмогорова λ

4. По таблиці додатків 2 визначаємо найбільш близьке до розрахункового λ значення, яке дорівнює $\lambda = 0,3$. Таким чином, з імовірністю, рівною 1, можна стверджувати справедливості висунутої гіпотези про розподіл випадкової величини за законом Пуассона в генеральній сукупності. Деякі розбіжності між емпіричним і теоретичним розподілами носять випадковий характер.

Завдання 1.

У результаті зважування 100 заготовок отримано наступний емпіричний розподіл Потрібно при рівні значущості $\alpha = 0,05$ перевірити гіпотезу про рівномірний закон розподілу ваги заготовок у інтервалі (0;10).

1)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	18	21	19	19	23
p_i^*	0,21	0,18	0,2	0,22	0,19

2)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	19	22	18	18	23
p_i^*	0,2	0,18	0,2	0,22	0,2

3)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	21	20	19	17	23
p_i^*	0,17	0,18	0,24	0,22	0,19

4)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	16	17	19	21	27
p_i^*	0,2	0,19	0,24	0,22	0,15

5)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	18	23	18	17	24
p_i^*	0,24	0,18	0,2	0,2	0,18

6)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	23	17	19	18	23
p_i^*	0,18	0,21	0,22	0,2	0,19

7)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	17	24	18	16	25
p_i^*	0,2	0,19	0,17	0,23	0,21

8)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	18	20	18	17	27
p_i^*	0,25	0,17	0,16	0,2	0,22

9)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	23	17	21	18	21
p_i^*	0,19	0,18	0,21	0,22	0,2

10)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	20	21	17	19	24
p_i^*	0,2	0,17	0,24	0,2	0,19

11)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	22	17	16	23	22
p_i^*	0,19	0,17	0,22	0,24	0,18

12)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	20	21	19	17	23
p_i^*	0,18	0,17	0,24	0,21	0,2

13)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	17	17	25	20	21
p_i^*	0,18	0,19	0,23	0,25	0,15

14)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	18	24	18	17	23
p_i^*	0,18	0,2	0,24	0,19	0,19

15)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	22	21	16	15	26
p_i^*	0,21	0,19	0,21	0,2	0,19

16)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	16	17	26	19	22
p_i^*	0,19	0,18	0,24	0,2	0,19

17)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	22	14	19	22	23
p_i^*	0,24	0,18	0,17	0,19	0,22

18)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	18	26	16	19	21
p_i^*	0,19	0,18	0,24	0,19	0,2

19)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
-------	------------	------------	------------	------------	-------------

	m_i	17	23	18	20	22
	p_i^*	0,18	0,18	0,24	0,2	0,2
20)	I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
	m_i	22	20	20	15	23
	p_i^*	0,2	0,22	0,2	0,18	0,2
21)	I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
	m_i	21	21	19	19	20
	p_i^*	0,18	0,22	0,22	0,18	0,2
22)	I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
	m_i	23	18	19	17	23
	p_i^*	0,2	0,19	0,24	0,24	0,13
23)	I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
	m_i	20	24	19	18	19
	p_i^*	0,26	0,17	0,18	0,2	0,19
24)	I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
	m_i	22	21	17	19	21
	p_i^*	0,2	0,18	0,23	0,23	0,16
25)	I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
	m_i	24	17	19	16	24
	p_i^*	0,2	0,18	0,2	0,2	0,22
26)	I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
	m_i	17	17	26	17	23
	p_i^*	0,17	0,18	0,22	0,2	0,23
27)	I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
	m_i	22	21	22	14	21
	p_i^*	0,24	0,18	0,21	0,14	0,23
28)	I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
	m_i	18	24	19	17	22
	p_i^*	0,22	0,17	0,21	0,22	0,18
29)	I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
	m_i	24	18	19	20	19
	p_i^*	0,21	0,18	0,22	0,2	0,19
30)	I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
	m_i	19	21	17	17	26
	p_i^*	0,25	0,16	0,19	0,22	0,18
31)	I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
	m_i	18	21	19	19	23

p_i^*	0,21	0,18	0,2	0,22	0,19
---------	------	------	-----	------	------

32)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	19	22	18	18	23
p_i^*	0,2	0,18	0,2	0,22	0,2

33)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	21	20	19	17	23
p_i^*	0,17	0,18	0,24	0,22	0,19

34)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	16	17	19	21	27
p_i^*	0,2	0,19	0,24	0,22	0,15

35)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	18	23	18	17	24
p_i^*	0,24	0,18	0,2	0,2	0,18

36)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	23	17	19	18	23
p_i^*	0,18	0,21	0,22	0,2	0,19

37)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	17	24	18	16	25
p_i^*	0,2	0,19	0,17	0,23	0,21

38)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	18	20	18	17	27
p_i^*	0,25	0,17	0,16	0,2	0,22

39)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	23	17	21	18	21
p_i^*	0,19	0,18	0,21	0,22	0,2

40)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	20	21	17	19	24
p_i^*	0,2	0,17	0,24	0,2	0,19

41)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	22	20	20	15	23
p_i^*	0,2	0,22	0,2	0,18	0,2

42)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
m_i	21	21	19	19	20
p_i^*	0,18	0,22	0,22	0,18	0,2

43)

I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
-------	------------	------------	------------	------------	-------------

44)	m_i	23	18	19	17	23
	p_i^*	0,2	0,19	0,24	0,24	0,13
45)	I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
	m_i	20	24	19	18	19
	p_i^*	0,26	0,17	0,18	0,2	0,19
25)	I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
	m_i	22	21	17	19	21
	p_i^*	0,2	0,18	0,23	0,23	0,16
46)	I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
	m_i	24	17	19	16	24
	p_i^*	0,2	0,18	0,2	0,2	0,22
47)	I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
	m_i	17	17	26	17	23
	p_i^*	0,17	0,18	0,22	0,2	0,23
48)	I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
	m_i	22	21	22	14	21
	p_i^*	0,24	0,18	0,21	0,14	0,23
50)	I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
	m_i	18	24	19	17	22
	p_i^*	0,22	0,17	0,21	0,22	0,18
	I_i	[0,5; 2,5[	[2,5; 4,5[	[4,5; 6,5[	[6,5; 8,5[	[8,5; 10,5[
	m_i	24	18	19	20	19
	p_i^*	0,21	0,18	0,22	0,2	0,19

Завдання 2. Для задачі 2 завдання 5 (Перевірка статистичних гіпотез. Критерій згоди χ^2 (критерій Пірсона)) виконати перевірку статистичних гіпотез за критерієм Колмогорова.

Критичні точки критерію λ (Колмогорова)

Табл.6.1

λ	0,1	0,05	0,01	0,001
$\lambda_{кр}$	1,224	1,358	1,627	1,950

Значення критерію Колмогорова $P(\lambda)$

Табл.6.2

λ	P	λ	P
0,30	1	0,80	0,5441
0,35	0,9997	0,85	0,4653
0,40	0,9972	0,90	0,3927
0,45	0,9874	0,95	0,3275
0,50	0,9639	1,0	0,2700
0,55	0,9228	1,1	0,1777
0,60	0,8643	1,2	0,1122
0,65	0,7920	1,3	0,0681
0,70	0,7112	1,4	0,0397
0,75	0,6272	1,5	0,0222

Задача № 7

Дисперсійний аналіз

Короткий теоретичний матеріал

Дисперсійний аналіз дозволяє досліджувати відмінність між групами даних, визначати, чи мають ці розбіжності випадковий характер або викликані конкретними обставинами. Наприклад, якщо продажі фірми в одному з регіонів знизилися, то за допомогою дисперсійного аналізу можна з'ясувати, чи випадково зниження оборотів в цьому регіоні в порівнянні з іншими, і при необхідності провести організаційні зміни. При виконанні експерименту в різних умовах дисперсійний аналіз допоможе визначити, наскільки впливають зовнішні фактори на вимірювання, або відхилення носять випадковий характер. Якщо на виробництві для поліпшення якості продукції змінюють режим процесів, то дисперсійний аналіз дозволяє оцінити результати впливу даного чинника.

Крім того, починаючи дослідження погано вивченою складної системи, слід виділити суттєві фактори з великого числа незалежних факторів, скоротивши тим самим загальну кількість досліджуваних факторів. Так, наприклад, при виконанні на автоматичній лінії деякої операції обробки паралельно на декількох верстатах важливо для правильної побудови подальшої обробки знати, чи є однотипним середні розміри деталей, тобто визначити вплив одного фактора - верстата.

Якщо обробка ведеться кількома операторами, то нас цікавить вплив двох факторів - верстата і оператора.

Науково обгрунтоване рішення подібних задач при деяких припущеннях становить предмет дисперсійного аналізу.

Основоположником дисперсійного аналізу вважають відомого англійського математика Р.Фішера (1900-1962г.г.), який ввів у статистику самі терміни «дисперсія» і «дисперсійний аналіз».

Дисперсійний аналіз - статистичний метод, призначений для оцінки впливу різних факторів на результат експерименту, а також для подальшого планує-вання аналогічних експериментів.

Може виникнути питання: чому метод, який використовується для порівняння середніх значень, називається дисперсійним аналізом? Вся справа в тому, що при встановленні різниці між середніми значеннями ми насправді порівнюємо дисперсії аналізованих сукупностей.

Залежно від кількості факторів, включених до аналізу, розрізняють однофакторний, двохфакторну і багатфакторний аналізи.

Так, наприклад, при виконанні деякої операції паралельно на декількох верстатах важливо знати, чи є однотипними середні розміри деталей, що дозволяє визначити вплив *одного фактора*.

Якщо ж обробка деталей ведеться кількома фахівцями, то нас буде цікавити вплив *двох факторів* - верстата і фахівця.

Суть дисперсійного аналізу полягає у тому, що загальну дисперсію досліджуваної ознаки розділяють на окремі компоненти, які обумовлені впливом певних конкретних факторів:

$$S_{\text{обиц}} = S_{\text{факт}} + S_{\text{ост}},$$

де $S_{\text{обиц}}$ — сума квадратів відхилень від середнього для всієї вибірки,

$S_{\text{факт}}$ — для досліджуваного фактора (між групова дисперсія)

$S_{\text{ост}}$ — для неврахованих факторів (внутрігрупова дисперсія).

Цю рівність називають основним рівнянням дисперсійного аналізу.

Для однофакторного дисперсійного аналізу: $S_{\text{общ}} = S_A + S_{\text{ост}}$.

Для двофакторного: $S_{\text{общ}} = S_A + S_B + S_{\text{ост}}$,

якщо для кожної комбінації рівнів факторів є лише одне вимірне значення і

$$S_{\text{общ}} = S_A + S_B + S_{AB} + S_{\text{ост}},$$

якщо для кожної комбінації рівнів факторів є багаторазово виміряні значення.

Тут S_{AB} — сума квадратів відхилень, викликаних взаємодією факторів A і B .

Слід зазначити, що для коректності отриманих результатів дисперсійного аналізу необхідно виконання наступних умов:

- результати спостережень повинні бути **незалежними випадковими величинами, що мають нормальний розподіл** (це дозволяє оцінити значимість оцінок дисперсій);
- **дисперсії генеральних сукупностей** повинні бути рівні (хоча і невідомі). Це гарантує відтворюваність дослідів з рівною точністю, що дає, в свою чергу, можливість правильно оцінити залишкову дисперсію.

Тільки при виконанні цих умов можна застосувати апарат дисперсійного аналізу!

Однофакторний дисперсійний аналіз

Однофакторний дисперсійний аналіз (ANOVA – *analysis of variance*) використовується для порівняння середніх значень для трьох і більше вибірок. *Фактором* називається *незалежна* змінна, вплив якої вивчається на залежну змінну.

Наприклад, фактором може бути рівень освіти, вид діяльності, вікова група респондентів, ступінь лояльності до торгової марки і т.д. Аналіз заснований на розрахунку F – **статистики** (статистика Фішера), яка представляє собою відношення двох дисперсій: між групувальною і внутрігруповою. F – тест в однофакторному дисперсійному аналізі встановлює, значимо чи відрізняються середні декількох незалежних вибірок. Він замінює t - тест для незалежних вибірок при наявності більше двох вибірок і дає той же результат у випадку двох вибірок.

Необхідною умовою для проведення дисперсійного аналізу є те, щоб незалежна змінна була категоріальною, а залежна - метричною.

Набір даних у ANOVA складається з k - незалежних одновимірних вибірок, елементи яких виміряні у однакових одиницях (дол., кг., бали). Припустимі різні обсяги (розміри) вибірок.

Процедура виконання однофакторного дисперсійного аналізу:

1. Визначення незалежних і залежних змінних
2. Розкладання повної дисперсії ($S_{\text{общ}}$)
3. Вимірювання ефекту (η^2)
4. Перевірка значущості (F)
5. Подання результату

1 етап. Підготовка даних для аналізу виглядає наступним чином:

	Незалежна змінна - фактор (Напр., вид діяльності) (Кількість вибірок $k = 4$)			
	Вибірка 1 - (економісти)	Вибірка 2 - (інженери)	Вибірка 3 - (філологи)	Вибірка k - (хіміки)
Залежна	$X_{1,1}$	$X_{2,1}$	$X_{3,1}$	$X_{k,1}$
Залежна	$X_{1,2}$	$X_{2,2}$	$X_{3,2}$	$X_{k,2}$
Залежна	$X_{1,3}$	$X_{2,3}$	$X_{3,3}$	$X_{k,3}$
Залежна	$X_{1,4}$	$X_{2,4}$	$X_{3,4}$	$X_{k,4}$
Залежна	$X_{1,5}$	$X_{2,5}$		$X_{k,5}$
Залежна		$X_{2,6}$		$X_{k,6}$
Залежна		$X_{2,7}$		
Об'єм $n =$ $n_1 + n_2 + n_3 + \dots + n_k$	$n_1 = 5$	$n_2 = 7$	$n_3 = 4$	$n_k = 6$
Середнє	X_1	X_2	X_3	X_k
Ст. відхилення	σ_1	σ_2	σ_3	σ_k

Нульова гіпотеза у однофакторному дисперсійному аналізі стверджує, що всі середні значення з різних генеральних сукупностей (які представлені вибірковими середніми) рівні між собою.

$H_0 : \mu_1 = \mu_k$ (всі рівні), (або $X_1 = X_2 = \dots = X_k$), або $H_0 : S_{\text{факт}} = S_{\text{ост}}$.

Альтернативна гіпотеза стверджує, що хоча б два будь-яких середніх не рівні між собою.

$H_1 : \mu_1 \neq \mu_k$ (хоча б дві нерівні). (або $X_1 \neq X_k$), або $H_1 : S_{\text{факт}} > S_{\text{ост}}$.

F – тест складається у розрахунку F - статистики та порівнянні її з табличним значенням (аналогічно з t - тестом).

Оскільки нульова гіпотеза стверджує, що середні всіх генеральних сукупностей рівні, необхідно оцінити це середнє значення за всіма вибірками, тобто розрахувати загальну середню. **Загальна середня є середньою всіх значень з усіх вибірок.**

Якщо розміри вибірок не рівні, то середнє розраховується як середньозважене

$$\bar{X} = \frac{\sum_{i=1}^k n_i \cdot X_i}{n} \quad \text{з урахуванням розміру вибірок.}$$

2 етап. Для вивчення відмінностей між залежними змінними проводиться розкладання повної дисперсії: $S_{\text{общ}} = S_{\text{факт}} + S_{\text{ост}}$, де $S_{\text{факт}}$ – міжгрупова варіація и $S_{\text{ост}}$ - внутрігрупова варіація.

Міжгрупова варіація ($S_{\text{факт}}$) показує, наскільки вибіркові середні відрізняються між собою. Вона дорівнює нулю, якщо середні рівні і тим більше, чим сильніше розрізняються середні. Розрахунок міжгрупової дисперсії (варіації):

$$S_{\text{факт}} = \frac{n \sum_{i=1}^k (\bar{X}_i - \bar{X})^2}{k-1}, \quad \bar{X}_i = \frac{\sum_{j=1}^n x_{ij}}{n}, \quad \bar{X} = \frac{\sum_{i=1}^k \bar{x}_i}{k}.$$

Внутрігрупова варіація ($S_{\text{ост}}$) показує, наскільки відрізняються між собою значення по кожній вибірці

$$S_{\text{ост}} = \frac{\sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2}{n-k}.$$

Очевидно, що: $S_{\text{факт}} > S_{\text{ост}}$.

В протилежному випадку немає необхідності застосовувати критерій Фішера, тому що в цьому випадку фактор впливає незначно.

3 етап. Ефект впливу незалежної змінної на залежну змінну розраховується через кореляційне відношення

η^2 (*ета-квадрат*), яке розраховується за формулою:

$$\eta^2 = \frac{S_{\text{факт}} \cdot (k-1)}{S_{\text{факт}} \cdot (k-1) + S_{\text{ост}} \cdot (n-k)}$$

Значення кореляційного відношення знаходиться в межах від 0 до 1.

Воно дорівнює 0, коли всі вибіркові середні рівні, тобто незалежна змінна не впливає на залежну, і, навпаки, вплив збільшується зі зростанням цього значення. Іншими словами, величина η^2 являє собою міру варіації залежної змінної, викликану впливом на неї незалежною змінною.

4 етап фактично зводиться до процедури статистичної перевірки гіпотези про рівність середніх (наявності відмінностей) шляхом розрахунку F - статистики:

$$F = \frac{S_{\text{факт}}}{S_{\text{ост}}}$$

5 етап. Для того, щоб зробити остаточний висновок, необхідно звернутися до F - таблиці, що містить критичні значення F - статистики при істинній нульовій гіпотезі. Щоб знайти критичне значення, необхідно врахувати кількість ступенів свободи (*df - degree freedom*) і відповідний рівень перевірки (за замовчуванням 5%).

Ступінь свободи для груповий варіації становить « $k-1$ », а для внутрішньогрупової варіації « $n-k$ ».

F - тест полягає в порівнянні F - статистики, розрахованої за наявними даними з критичним значенням F - таблиці. Результат є значущим, якщо $F_{\text{стат}} > F_{\text{критич}}$, поскільки це говорить про наявність істотних відмінностей між середніми значеннями по групах.

Приклад 1.

Поставки продукції для деякої компанії здійснюються трьома по-постачальниками («Азимут», «Елен» і «Охаміт») в різний час: денні години, нічні зміни і під час перезміни. Цілком очевидно, що контроль за якістю продукції в денний час вище, ніж в інший час. Зібрані дані з оцінками якості (в балах), і необхідно дізнатися, чи є відмінність в якості продукції, яка поставляється в різний час?

	Денна зміна	Нічна зміна	Перезмінка
«Азимут»	77,06	93,12	77,05
«Елен»	81,14	88,13	78,11
«Охаміт»	82,02	81,18	79,91

1. Обчислимо оцінку математичного очікування кожного з варіантів поставки продукції:

$$\bar{X}_1 = \frac{77,06 + 81,14 + 82,02}{3} = 80,07,$$

$$\bar{X}_2 = \frac{93,12 + 88,13 + 81,18}{3} = 87,48,$$

$$\bar{X}_3 = \frac{77,05 + 78,11 + 79,91}{3} = 78,36.$$

2. Обчислимо оцінку математичного очікування спостережуваних величин:

$$\bar{X} = \frac{80,07 + 87,48 + 78,36}{3} = 81,97.$$

3. Обчислимо оцінку груповий дисперсії:

$$D_{\text{факт}} = \frac{3 \cdot ((80,07 - 81,97)^2 + (87,48 - 81,97)^2 + (78,36 - 81,97)^2)}{2} = 70,47.$$

4. Обчислимо оцінку внутрішню групової дисперсії:

$$D_{\text{ост}} = \frac{1}{(9 - 3)} \cdot ((77,06 - 80,07)^2 + (81,14 - 80,07)^2 + (82,02 - 80,07)^2 + \\ + (93,12 - 87,48)^2 + (88,13 - 87,48)^2 + (81,18 - 87,48)^2 + \\ + (77,05 - 78,36)^2 + (78,11 - 78,36)^2 + (79,91 - 78,36)^2) = 15,02$$

5. Обчислимо ефект впливу незалежної змінної на залежну змінну через кореляційне відношення η^2 (ета-квадрат):

$$\eta^2 = \frac{70,47 \cdot (3 - 1)}{(70,47 \cdot (3 - 1)) + (15,02 \cdot (9 - 3))} = 0,61.$$

Отримане кореляційне відношення показує, що 61% варіації залежної змінної пояснюється зміною незалежної змінної і 39% залежною змінною пояснюється іншими факторами, що діють на вибірку вибірково.

6. Обчислимо F - статистику:

$$F_{\text{набл.}} = \frac{70,47}{15,02} = 4,69.$$

7. При рівні значущості $\alpha=0,05$ по таблиці критичних точок розподілу Фішера зі ступенями свободи визначимо критичне значення критерію

$$k_1 = k - 1 = 3 - 1 = 2; k_2 = (n - k) = (9 - 3) = 6$$

$$F_{\text{крит}} = F_{\text{табл}}(\alpha; k_1 = k - 1; k_2 = (n - k)) = F_{\text{табл}}(0,05; 2; 6) = 5,14.$$

7. Оскільки $F_{\text{набл}} = 4,69 < 5,14 = F_{\text{кр}}$, то немає підстав відкинути нульову гіпотезу H_0 . Таким чином, групові середні відрізняються незначно, тобто фактор A впливає незначно. Іншими словами можна відзначити: результати розрахунку показують, що $F_{\text{стат}} < F_{\text{критич}}$ ($4,691 < 5,14$), отже, відмінність в якості продукції, що поставляється в різний час відсутня. Можна вважати доведеним той факт, що якість продукції, що поставляється не залежить від часу поставки і є однаковим в різний час.

Розв'язання задачі в середовищі EXCEL:



В системі електронних таблиць Microsoft Excel є набір інструментів для аналізу даних, званий **пакет аналізу**, який може бути використаний для вирішення складних статистичних задач. Для використання одного з цих інструментів необхідно вказати вхідні дані і вибрати параметри; аналіз буде проведено за допомогою відповідної статистичної макрофункції, і результати будуть представлені в вихідному діапазоні.

В меню **Сервис** необхідно обрати команду **Анализ данных**. Якщо така команда відсутня в меню **Сервис**, то необхідно встановити в Microsoft Excel пакет **анализа данных**. Установка такого пакета проводиться таким чином. В меню **Сервис** необхідно обрати команду **Надстройки**. Якщо у списку надбудов немає пакета **анализа данных**, то необхідно натиснути кнопку “Обзор” і задати диск, каталог і ім'я файлу для надбудови “Пакет анализа”, Або запустити програму установки Microsoft Excel. При цьому необхідно встановити прапорець “Пакет анализа” (надбудови, встановлені у Microsoft Excel, залишаються доступними, доки не будуть вилучені).

Методика проведення однофакторного дисперсійного аналізу у середовищі EXCEL

1. Для використання інструментів аналізу вхідний діапазон має бути поданий у вигляді рядків або стовпців на листі (рис.7.1). При цьому сукупність комірок, що містять аналізовані дані, називається **вхідним діапазоном**.

The screenshot shows the Microsoft Excel interface with a data table. The table has columns for shifts and average values. The data is as follows:

	Дневная смена	Ночная смена	Пересемка
«Азимут»	77,06	93,12	77,05
«Эпел»	81,14	88,13	78,11
«Охотин»	82,02	81,18	79,91

Рис. 7.1. Формування вхідного діапазону для аналізу даних

2. У меню **Сервис** вибирають команду **Анализ данных**. У списку інструментів статистичного аналізу вибирають **Однофакторный дисперсионный анализ** (рис.7.2).

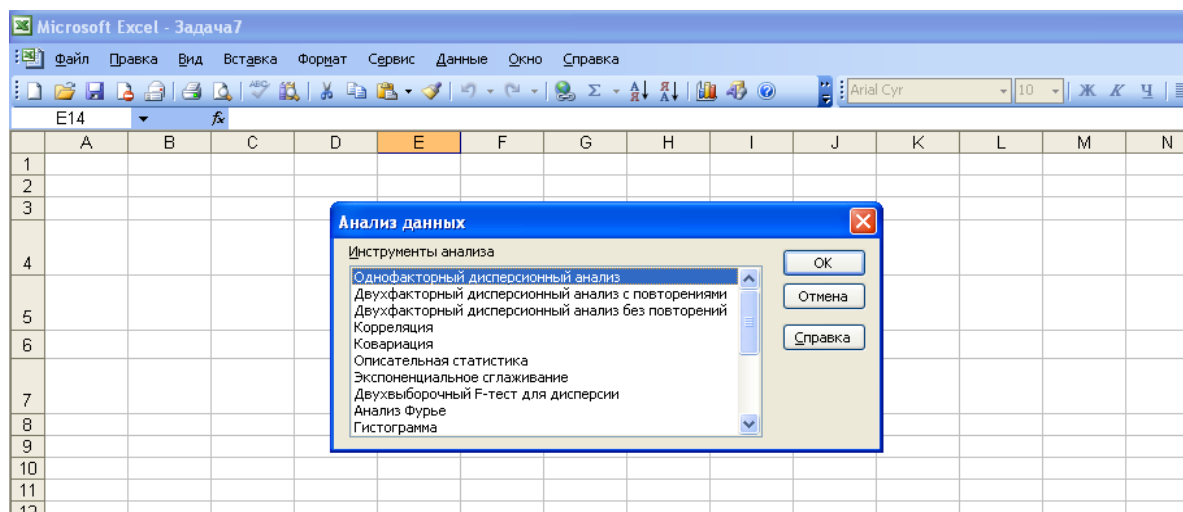


Рис.7.2. Вибір інструменту аналізу

3. У діалоговому вікні *однофакторний дисперсійний аналіз* (рис.7.3) вказують вхідний інтервал, спосіб групування, вихідний інтервал, Метки в первой строке / Метки в первом столбце/, альфа (рівень значущості).

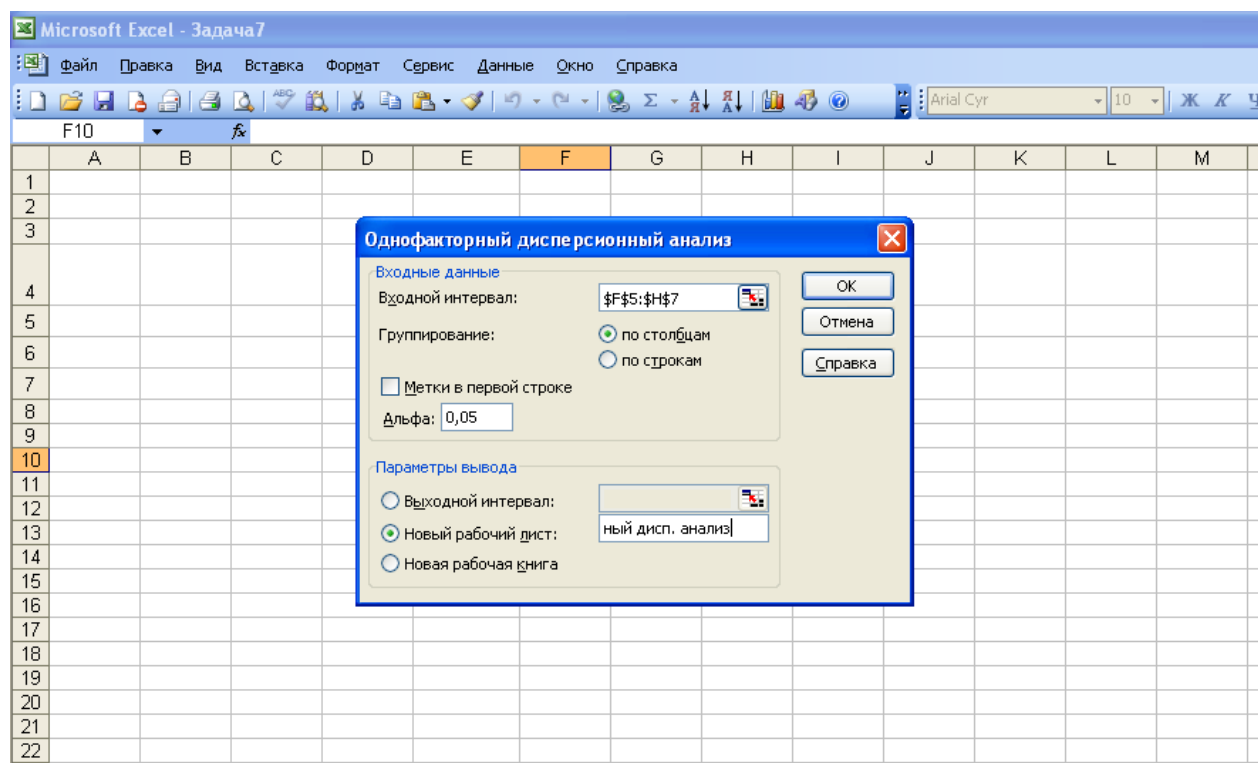


Рис.7.3. Діалогове вікно однофакторного дисперсійного аналізу

Вхідний діапазон - це посилання на комірки, що містять аналізовані дані. Посилання має складатися як мінімум з двох суміжних діапазонів даних, організованих у вигляді стовпців або рядків. Вхідний інтервал можна задати за допомогою миші, або набрати на клавіатурі.

Групування. Виберіть пункт “по столбцам” або “по строкам” залежно від розташування даних у вхідному діапазоні. *Метки в первой строке/ Метки в первом столбце.* Встановіть перемикач у положення “Метки в первой строке”, якщо перший рядок у вхідному діапазоні містить назви стовпців. Встановіть перемикач у положення “Метки в первом столбце”, якщо назви рядків знаходяться у першому стовпці вхідного діапазону. Якщо вхідний діапазон не містить міток, то необхідні заголовки у вихідному діапазоні будуть створені автоматично.

Вихідний діапазон. Введіть посилання на осередок, розташовану в лівому верхньому кутку вихідного діапазону. Розміри вихідної області будуть розраховані автоматично, і відповідне повідомлення з'явиться на екрані в тому випадку, якщо вихідний діапазон займає місце існуючих даних або його розміри перевищують розміри листа.

Новий лист. Встановіть перемикач, щоб відкрити новий лист в книзі і вставити результати аналізу, починаючи з комірки A1. Якщо в цьому є необхідність, введіть ім'я нового аркуша в полі, розташованому навпроти відповідного положення перемикача.

Нова книга. Встановіть перемикач, щоб відкрити нову книгу і вставити результати аналізу в комірку A1 на першому листі у цієї книзі.

4. В результаті обробки даних створюється новий лист з ім'ям *однофакторный дисп. анализ* з результатами, які представлені на рис.7.4.

	A	B	C	D	E	F	G	H	I
1	Однофакторный дисперсионный анализ								
2									
3	ИТОГИ								
4	Группы	Счет	Сумма	Среднее	Дисперсия				
5	Столбец 1	3	240,22	80,07333	7,003733333				
6	Столбец 2	3	262,43	87,47667	35,96103333				
7	Столбец 3	3	235,07	78,35667	2,090533333				
8									
9									
10	Дисперсионный анализ								
11	Источник вариации	SS	df	MS	F	P-Значение	F критическое		
12	Между группами	140,9307	2	70,46534	4,691923777	0,059327883	5,14325285		
13	Внутри групп	90,1106	6	15,01843					
14									
15	Итого	231,0413	8						
16									
17									
18									
19									
20									
21									
22									
23									

Рис.7.4. Результаты однофакторного дисперсионного анализа

Аналіз таблиці ИТОГИ:

“Счет” – число повторностей. “Сумма” – сума значень показника по рядках. “Дисперсия” – частинна дисперсія показника.

Таблиця ANOVA представляє результати дисперсійного аналізу однофакторного комплексу, у якому перша колонка "Источник вариации" містить найменування дисперсій. Графа "SS" - це сума квадратів відхилень, "df" - ступінь свободи, графа "MS" - середній квадрат, "F" - критерій фактичного F - розподілу. "Р - значение" - ймовірність того, що дисперсія, відтворена рівнянням, дорівнює дисперсії залишків. Визначає ймовірність того, що отримана кількісна визначеність взаємозв'язку між факторами і результатом може вважатися випадковою. "F - критическое" - це значення F - теоретичного, яке згодом порівнюється з F - фактичним.

5 Аналіз результатів.

Результати розрахунку показують, що $F_{\text{стат}} < F_{\text{критич}}$ ($4,691 < 5,14$), отже, відмінність у якості продукції, що поставляється у різний час відсутня. Крім того, *P-значение* (ймовірність істинності нульової гіпотези про рівність середніх) перевищує 0,05, тобто вона не може бути відхилена. Можна вважати доведеним той факт, що якість продукції, що поставляється не залежить від часу поставки і є однаковою у різний час.

6. Розрахунок емпіричного кореляційного відношення.

У комірку C24 вводять формулу $=B12/B15$. При цьому отримують відповідне значення кореляційного відношення (рис. 7.5).

The screenshot shows an Excel spreadsheet titled "Microsoft Excel - Задача 7". The formula bar shows $=B12/B15$ for cell C24. The spreadsheet contains two tables. The first table is an ANOVA summary, and the second is a detailed ANOVA table.

	A	B	C	D	E	F	G	H	I	J	K
1	Однофакторный дисперсионный анализ										
2											
3	ИТОГИ										
4	Группы	Очет	Сумма	Среднее	Дисперсия						
5	Столбец 1	3	240,22	80,07333	7,003733333						
6	Столбец 2	3	262,43	87,47667	35,96103333						
7	Столбец 3	3	235,07	78,35667	2,090533333						
8											
9											
10	Дисперсионный анализ										
11	Источник вариации	SS	df	MS	F	P-Значение	F критическое				
12	Между группами	140,9307	2	70,46534	4,691923777	0,059327883	5,14325285				
13	Внутри групп	90,1106	6	15,01843							
14											
15	Итого	231,0413	8								
16											
17											
18											
19											
20											
21											
22											
23											
24	$\eta^2 =$		0,61								
25											
26											
27											

Рис.7.5. Розрахунок кореляційного відношення

Отримане кореляційне відношення показує, що 61% варіації залежної змінної пояснюється зміною незалежної змінної і 39% залежною змінною пояснюється іншими факторами, що діють на вибірку вибірково.

Приклад 2. Потрібно оцінити вплив рівня реклами всередині магазину на обсяги продажів. Є такі дані по 10 торгових точках:

Торгова точка	Рівень реклами		
	високий	середній	низький
	Продажі, тис. грн		
1	10	8	5
2	9	8	7
3	10	7	6
4	8	9	4
5	9	6	5
6	8	4	2
7	9	5	3
8	7	5	2
9	7	6	1
10	6	4	2

1. Обчислимо оцінку математичного очікування кожного з варіантів поставки продукції:

$$\bar{X}_1 = \frac{10 + 9 + 10 + 8 + 9 + 8 + 9 + 7 + 7 + 6}{10} = 8,3.$$

$$\bar{X}_2 = \frac{8 + 8 + 7 + 9 + 6 + 4 + 5 + 5 + 6 + 4}{10} = 6,2.$$

$$\bar{X}_3 = \frac{5 + 7 + 6 + 4 + 5 + 2 + 3 + 2 + 1 + 2}{10} = 3,7.$$

2. Обчислимо оцінку математичного очікування спостережуваних величин:

$$\bar{X} = \frac{8,3 + 6,2 + 3,7}{3} = 6,07.$$

3. Обчислимо оцінку міжгрупової дисперсії:

$$D_{\text{факт}} = \frac{10 \cdot ((8,3 - 6,07)^2 + (6,2 - 6,07)^2 + (3,7 - 6,07)^2)}{2} = 53,03.$$

3. Обчислимо оцінку внутрігрупової дисперсії:

$$D_{\text{ост}} = \frac{1}{(30 - 3)} \cdot ((10 - 8,3)^2 + (9 - 8,3)^2 + (10 - 8,3)^2 + (8 - 8,3)^2 + (9 - 8,3)^2 + (8 - 8,3)^2 + (9 - 8,3)^2 + (7 - 8,3)^2 + (7 - 8,3)^2 + (6 - 8,3)^2 + (8 - 6,2)^2 + (8 - 6,2)^2 + (7 - 6,2)^2 + (9 - 6,2)^2 + (6 - 6,2)^2 + (4 - 6,2)^2 + (5 - 6,2)^2 + (5 - 6,2)^2 + (6 - 6,2)^2 + (4 - 6,2)^2 + (5 - 3,7)^2 + (7 - 3,7)^2 + (6 - 3,7)^2 + (4 - 3,7)^2 + (5 - 3,7)^2 + (2 - 3,7)^2 + (3 - 3,7)^2 + (2 - 3,7)^2 + (1 - 3,7)^2 + (2 - 3,7)^2) = 2,96.$$

4. Обчислимо ефект впливу незалежної змінної на залежну змінну через кореляційне відношення η^2 (ета-квадрат):

$$\eta^2 = \frac{53,03 \cdot (3-1)}{(53,03 \cdot (3-1)) + (2,96 \cdot (30-3))} = 0,57.$$

Отримане кореляційне відношення показує, що 57% варіації залежної змінної пояснюється зміною незалежної змінної і 43% залежною змінною пояснюється іншими факторами, що діють на вибірку вибірково.

6. Обчислимо F – статистику:

$$F_{\text{набл.}} = \frac{53,03}{2,96} = 17,94.$$

7. При рівні значущості $\alpha=0,05$ за таблицею критичних точок розподілу Фішера зі ступеня ми свободи визначимо критичне значення критерію

$$k_1 = k - 1 = 3 - 1 = 2; k_2 = (n - k) = (30 - 3) = 27.$$

$$F_{\text{крит}} = F_{\text{табл}}(\alpha; k_1 = k - 1; k_2 = (n - k)) = F_{\text{табл}}(0,05; 2; 27) = 3,35.$$

8. Результати аналізу показують, що різниця у обсягах продажів у магазинах з різним рівнем реклами, є значущою (суттєвою). Про це свідчить значення $F: 17,943 > 3,354$. Можна вважати доведеним той факт, що реклама впливає на обсяги продажів продукції.

Отже, нульова гіпотеза відхиляється і приймається альтернативна про те, що рівень реклами впливає на обсяги продажів, причому спостерігається пряма залежність, тобто більш високому рівню реклами відповідають більш високі обсяги продажів.

Розв'язання задачі у середовищі EXCEL:



Методика проведення однофакторного дисперсійного аналізу у середовищі EXCEL аналогічна раніше розглянутого прикладу.

1.В результаті обробки даних створюється новий лист з ім'ям **приклад 2** з результатами, які представлені на рис. 7.6.

Microsoft Excel - Задача7						
Файл Правка Вид Вставка Формат Сервис Данные Окно Справка						
D23 fx Arial Cyr						
	A	B	C	D	E	F
1	Однофакторный дисперсионный анализ					
2						
3	ИТОГИ					
4	Группы	Счет	Сумма	Среднее	Дисперсия	
5	Столбец 1	10	83	8,3	1,788888889	
6	Столбец 2	10	62	6,2	3,066666667	
7	Столбец 3	10	37	3,7	4,011111111	
8						
9						
10	Дисперсионный анализ					
11	Источник вариации	SS	df	MS	F	P-Значение
12	Между группами	106,066667	2	53,03333	17,94360902	1,10362E-05
13	Внутри групп	79,8	27	2,955556		3,354130829
14						
15	Итого	185,866667	29			
16						
17						
18						
19						
20						

Рис. 7.6. Результати однофакторного дисперсійного аналізу

2. Аналіз результатів.

Результати аналізу показують, що різниця у обсягах продажів у магазинах з різним рівнем реклами, є значущою (суттєвою). Про це свідчить значення $F: 17,943 > 3,354$, а також мала ймовірність прийняття нульової гіпотези (P – значення = $1,10 \text{ E-}0,05$). Отже, нульова гіпотеза відхиляється і приймається альтернативна про те, що рівень реклами впливає на обсяги продажів, причому спостерігається пряма залежність, тобто більш високому рівню реклами відповідають більш високі обсяги продажів.

3. Розрахунок емпіричного кореляційного відношення.

У комірку C24 вводять формулу = B12 / B15. При цьому отримують відповідне значення кореляційного відношення (рис.7.7).

	A	B	C	D	E	F	G	H	I
1	Однофакторный дисперсионный анализ								
2									
3	ИТОГИ								
4	Группы	Счет	Сумма	Среднее	Дисперсия				
5	Столбец 1	10	83	8,3	1,788888889				
6	Столбец 2	10	62	6,2	3,066666667				
7	Столбец 3	10	37	3,7	4,011111111				
8									
9									
10	Дисперсионный анализ								
11	Источник вариации	SS	df	MS	F	P-Значение	F критическое		
12	Между группами	106,066667	2	53,03333	17,94360902	1,10362E-05	3,354130829		
13	Внутри групп	79,8	27	2,955556					
14									
15	Итого	185,866667	29						
16									
17									
18									
19									
20									
21									
22									
23									
24		η^2	=	0,57					
25									
26									
27									

Рис.7.7. Розрахунок кореляційного відношення

Отримане кореляційне відношення показує, що 57% варіації залежної змінної пояснюється зміною незалежної змінної і 43% залежною змінною пояснюється іншими факторами, що діють на вибірку вибірково.

Двофакторний дисперсійний аналіз

Як впливає з назви, завдання розглянутого раніше *однофакторного дисперсійного аналізу* полягає в з'ясуванні впливу якогось одного фактора на цікаву для нас кількісну змінну. Однак дуже рідко той чи інший процес визначається тільки одним фактором. Навпаки - зазвичай спостерігається одночасний вплив багатьох чинників. Задача дослідника - виявити, які фактори впливають на досліджуване явище, а які - можна виключити з розгляду. Двохфакторний дисперсійний аналіз (англ. *Two-way analysis of variance*, або *two-way ANOVA*) дозволяє встановити одночасний вплив двох факторів, а також взаємодію між цими факторами. При наявності більше двох факторів говорять про багатофакторний дисперсійний аналіз (англ. *Multifactor ANOVA*).

Слід зазначити, що принципової різниці між багатфакторним і однофакторним дисперсійним аналізом немає. Багатфакторний аналіз не змінює загальну логіку дисперсійного аналізу, а лише дещо ускладнює її, оскільки, крім врахування впливу на залежну змінну кожного з факторів окремо, слід оцінювати і їх спільну дію. Таким чином, те нове, що вносить в аналіз даних багатфакторний дисперсійний аналіз, стосується в основному можливості оцінити міжфакторну взаємодію. Проте, як і раніше, залишається можливість оцінювати вплив кожного фактора окремо. У цьому сенсі процедура багатфакторного дисперсійного аналізу (у варіанті її комп'ютерного використання), безсумнівно, більш економічна, оскільки всього за один запуск вирішує відразу два завдання: оцінюється вплив кожного з факторів і їх взаємодію.

Розглянемо одну з реалізацій двофакторного дисперсійного аналізу, яка найчастіше застосовується на практиці.

Для проведення аналізу необхідно мати результати спостережень над досліджуваною ознакою, які представляються у вигляді таблиці спостережень.

A B	A ₁	A ₂	...	A _k	$\bar{y}_{*j}$
B ₁	y ₁₁	y ₂₁	...	y _{k1}	$\bar{y}_{*1}$
B ₂	y ₁₂	y ₂₂	...	y _{k2}	$\bar{y}_{*2}$
...	...	...	...	...	...
B _n	y _{1n}	y _{2n}	...	y _{kn}	$\bar{y}_{*n}$
$\bar{y}_{i*}$	$\bar{y}_{1*}$	$\bar{y}_{2*}$	...	$\bar{y}_{k*}$	$\bar{y}$

Тут y_{ij} – значення ознаки

A - фактор з k рівнями: $A_1, A_2, \dots, A_k$.

B - фактор з n рівнями: $B_1, B_2, \dots, B_n$.

$$\tilde{y}_{*j} = \frac{1}{k} \sum_{i=1}^k y_{ij} \text{ – середнє значення по рядках.}$$

$$\tilde{y}_{i*} = \frac{1}{n} \sum_{j=1}^n y_{ij} \text{ – середнє значення по стовпцях/}$$

$$\bar{y} = \frac{1}{n} \sum_{j=1}^n \bar{y}_{*j} = \frac{1}{k} \sum_{i=1}^k \bar{y}_{i*} = \frac{1}{kn} \sum_{i=1}^k \sum_{j=1}^n y_{ij} \text{ - загальне середнє значення.}$$

Очевидно, що $\bar{y}_{*1}$ є оцінкою математичного очікування тобто

$$\tilde{m}_1^B = \bar{y}_{*1}; \tilde{m}_2^B = \bar{y}_{*2}; \dots; \tilde{m}_n^B = \bar{y}_{*n}$$

Аналогічно $\tilde{m}_1^A = \bar{y}_{1*}; \tilde{m}_2^A = \bar{y}_{2*}; \dots; \tilde{m}_k^A = \bar{y}_{k*}$

Необхідно перевірити гіпотезу про рівність математичних очікувань генеральних сукупностей, тобто для фактора A

$$H_0^A : m_1^A = m_2^A = \dots = m_k^A;$$

$$H_1^A : m_1^A \neq m_2^A \dots \neq m_k^A$$

Для фактора В

$$H_0^B : m_1^B = m_2^B = \dots = m_n^B;$$

$$H_1^B : m_1^B \neq m_2^B \neq \dots \neq m_n^B$$

Ці гіпотези ми перевіряємо за відомими оцінками Безпосередня перевірка цих гіпотез не раціональна. Р. Фішер запропонував замість цих гіпотез перевіряти однорідність оцінок дисперсій а саме: розкид значень $\tilde{m}_i^A$ ($i = 1, 2, \dots, k$) можна виміряти за допомогою факторної дисперсії, яку позначають $\tilde{D}_A$ розкид значень $\tilde{m}_j^B$ ($j = 1, 2, \dots, n$) можна виміряти за допомогою факторної дисперсії, яку позначають $\tilde{D}_B$. Чим сильніше розрізняються між собою середні величини, тим більше факторна дисперсія. Крім впливу факторів на вихідну величину впливають випадкові фактори (помилки досвіду). Розкид значень за рахунок випадкових причин вимірюється дисперсією, яка називається залишковою і позначається $\tilde{D}_{\text{ост}}$.

Отже нульові і протилежні гіпотези для факторів А і В можна записати так:

$$H_0^A : m_1^A = m_2^A = \dots = m_k^A \Leftrightarrow D_A = D_{\text{ост}} - \text{фактор А впливає не значимо};$$

$$H_1^A : m_1^A \neq m_2^A \neq \dots \neq m_k^A \Leftrightarrow D_A > D_{\text{ост}} - \text{фактор А впливає значимо};$$

$$H_0^B : m_1^B = m_2^B = \dots = m_n^B \Leftrightarrow D_B = D_{\text{ост}} - \text{фактор В впливає не значимо};$$

$$H_1^B : m_1^B \neq m_2^B \neq \dots \neq m_n^B \Leftrightarrow D_B > D_{\text{ост}} - \text{фактор В впливає значимо}.$$

$$\text{Перевіримо гіпотези: } H_0^A : D_A = D_{\text{ост}}; H_0^B : D_B = D_{\text{ост}};$$

$$H_1^A : D_A > D_{\text{ост}}; H_1^B : D_B > D_{\text{ост}}.$$

Однорідність дисперсій перевіряється за критерієм Фішера:

$$F_{\text{набл}}^A = \frac{\tilde{D}_A}{\tilde{D}_{\text{ост}}}; F_{\text{набл}}^B = \frac{\tilde{D}_B}{\tilde{D}_{\text{ост}}}, \text{ де}$$

$$\tilde{D}_A = \frac{n \sum_{i=1}^k (\bar{y}_{i*} - \bar{y})^2}{k-1}; \tilde{D}_B = \frac{k \sum_{j=1}^n (\bar{y}_{*j} - \bar{y})^2}{n-1};$$

$$\tilde{D}_{\text{ост}} = \frac{\sum_{i=1}^n \sum_{j=1}^k (y_{ij} - \bar{y}_{i*} - \bar{y}_{*j} + \bar{y})^2}{(k-1)(n-1)}$$

Критичне значення визначається за таблицею Фішера:

$$F_{\text{крит}}^A = F_{\text{табл}}^A(\alpha; f_1 = k-1; f_2 = (k-1)(n-1));$$

$$F_{\text{крит}}^B = F_{\text{табл}}^B(\alpha; f_1 = n-1; f_2 = (k-1)(n-1)).$$

Якщо $F_{\text{набл}} \leq F_{\text{крит}}$, дисперсії неоднорідні, а це значить, що фактор впливає значимо.

Критерій Фішера застосовується, якщо $\tilde{D}_{\text{факт}} > \tilde{D}_{\text{ост}}$; якщо $\tilde{D}_{\text{факт}} < \tilde{D}_{\text{ост}}$, то без перевірки робиться висновок про незначущість впливу факторів.

Приклад 1.

У таблиці представлені дані про врожайність (ц / га) чотирьох сортів пшениці (чотири рівні фактора А), досягнутої при використанні п'яти типів добрив (п'ять рівнів фактора В). Дані отримані на 20 ділянках однакового розміру і аналогічного ґрунтового покриття.

Необхідно визначити, чи впливає сорт і тип добрива на урожайність пшениці.

Фактор В – тип добрів	Фактор А – сорт пшениці				
А В	А ₁	А ₂	А ₃	А ₄	$\bar{y}_{i*}$
В ₁	19	25	17	21	20,5
В ₂	22	19	19	18	19,5
В ₃	26	23	22	25	24
В ₄	18	26	20	23	21,75
В ₅	21	22	21	24	22
$\bar{y}_{*j}$	21,2	23	19,8	22,2	$\bar{y} = 21,55$

При рівні значущості $\alpha=0,05$ перевірити чи є значимим вплив факторів А і В на вихідну величину Y, тобто перевірити гіпотези:

$$H_0^A : m_1^A = m_2^A = m_3^A = m_4^A \Leftrightarrow D_A = D_{ocm} \Leftrightarrow \text{фактор А впливає не значимо};$$

$$H_1^A : m_1^A \neq m_2^A \neq m_3^A \neq m_4^A \Leftrightarrow D_A > D_{ocm} \Leftrightarrow \text{фактор А впливає значимо};$$

$$H_0^B : m_1^B = m_2^B = m_3^B = m_4^B = m_5^B \Leftrightarrow D_B = D_{ocm} \Leftrightarrow \text{фактор В впливає не значимо};$$

$$H_1^B : m_1^B \neq m_2^B \neq m_3^B \neq m_4^B \neq m_5^B \Leftrightarrow D_B > D_{ocm} \Leftrightarrow \text{фактор В впливає значимо};$$

На першому етапі обчислимо оцінки математичних очікувань:

А - фактор з k рівнями: А₁, А₂, А₃, А₄ (k=4).

В – фактор з n рівнями: В₁, В₂, В₃, В₄, В₅ (n=5) .

$$\tilde{y}_{*j} = \frac{1}{k} \sum_{i=1}^k y_{ij} \text{ – середнє значення по рядках}$$

$$\tilde{y}_{*1} = \frac{1}{k} \sum_{i=1}^k y_{ij} = \frac{19+25+17+21}{4} = 20,5.$$

$$\tilde{y}_{*2} = \frac{1}{k} \sum_{i=1}^k y_{ij} = \frac{22+19+19+18}{4} = 19,5.$$

$$\tilde{y}_{*3} = \frac{1}{k} \sum_{i=1}^k y_{ij} = \frac{26+23+22+25}{4} = 24.$$

$$\tilde{y}_{*4} = \frac{1}{k} \sum_{i=1}^k y_{ij} = \frac{18+26+20+23}{4} = 21,75.$$

$$\tilde{y}_{*5} = \frac{1}{k} \sum_{i=1}^k y_{ij} = \frac{21+22+21+24}{4} = 22.$$

$$\tilde{y}_{i*} = \frac{1}{n} \sum_{j=1}^n y_{ij} - \text{середнє значення за стовпцями}$$

$$\tilde{y}_{1*} = \frac{1}{k} \sum_{i=1}^k y_{ij} = \frac{19+22+26+18+21}{5} = 21,2.$$

$$\tilde{y}_{2*} = \frac{1}{k} \sum_{i=1}^k y_{ij} = \frac{25+19+23+26+22}{5} = 23,0.$$

$$\tilde{y}_{3*} = \frac{1}{k} \sum_{i=1}^k y_{ij} = \frac{17+19+22+20+21}{5} = 19,8.$$

$$\tilde{y}_{4*} = \frac{1}{k} \sum_{i=1}^k y_{ij} = \frac{21+18+25+23+24}{5} = 22,2.$$

$$\bar{y} = \frac{1}{n} \sum_{j=1}^n \bar{y}_{*j} = \frac{1}{k} \sum_{i=1}^k \bar{y}_{i*} = \frac{1}{kn} \sum_{i=1}^k \sum_{j=1}^n y_{ij} - \text{загальне середнє значення.}$$

$$\bar{y} = \frac{1}{9} (20,5 + 19,5 + 24 + 21,75 + 22 + 21,2 + 23 + 19,8 + 22,2) = 21,55.$$

Очевидно, що $\bar{y}_{*1}$ є оцінкою математичного очікування тобто

$$\tilde{m}_1^B = \bar{y}_{*1}; \tilde{m}_2^B = \bar{y}_{*2}; \dots; \tilde{m}_5^B = \bar{y}_{*5}$$

Аналогічно

$$\tilde{m}_1^A = \bar{y}_{1*}; \tilde{m}_2^A = \bar{y}_{2*}; \dots; \tilde{m}_4^A = \bar{y}_{4*}$$

Далі обчислимо оцінки дисперсій:

$$\tilde{D}_A = \frac{n \sum_{i=1}^k (\bar{y}_{i*} - \bar{y})^2}{k-1} = \frac{4[(20,5 - 21,55)^2 + (19,5 - 21,55)^2 + (24 - 21,55)^2 + (21,75 - 21,55)^2 + (22 - 21,55)^2]}{5-1} = 11,55;$$

$$\tilde{D}_B = \frac{k \sum_{j=1}^n (\bar{y}_{*j} - \bar{y})^2}{n-1} = \frac{5[(21,2 - 21,55)^2 + (23 - 21,55)^2 + (19,8 - 21,55)^2 + (22,2 - 21,55)^2]}{4-1} = 9,52;$$

$$\begin{aligned}\tilde{D}_{ост} &= \frac{\sum_{i=1}^n \sum_{j=1}^k (y_{ij} - \bar{y}_{i*} - \bar{y}_{*j} + \bar{y})^2}{(k-1)(n-1)} = \frac{1}{(5-1)(4-1)} [(19-20,5-21,2+21,55)^2 + \\ &+ (25-20,5-23+21,55)^2 + (17-20,5-19,8+21,55)^2 + (21-20,5-22,2+21,55)^2 + \\ &+ (22-19,5-21,2+21,55)^2 + (19-19,5-23+21,55)^2 + (19-19,5-19,8+21,55)^2 + (18-19,5-22,2+21,55)^2 + \\ &+ (26-24-21,2+21,55)^2 + (23-24-23+21,55)^2 + (22-24-19,8+21,55)^2 + (25-24-22,2+21,55)^2 + \\ &+ (18-24-21,2+21,55)^2 + (26-2,75-23+21,55)^2 + (20-21,75-19,8+21,55)^2 + (23-21,75-22,2+21,55)^2 + \\ &+ (21-22-21,2+21,55)^2 + (22-22-23+21,55)^2 + (21-22-19,8+21,55)^2 + (24-22-22,2+21,55)^2] = \\ &= 5,68;\end{aligned}$$

Однорідність дисперсій перевіряється за критерієм Фішера.

Для цього обчислимо F - статистику:

$$F_{набл}^A = \frac{\tilde{D}_A}{\tilde{D}_{ост}} = \frac{11,55}{5,68} = 2,03;$$

$$F_{набл}^B = \frac{\tilde{D}_B}{\tilde{D}_{ост}} = \frac{9,52}{5,68} = 1,67;$$

При рівні значущості $\alpha=0,05$ за таблицею критичних точок розподілу Фішера визначимо критичне значення критерію

$$F_{крит}^A = F_{табл}^A(\alpha=0,05; f_1=k-1=4-1=3; f_2=(k-1)(n-1)=3 \cdot 4=12)=3,49;$$

$$F_{крит}^B = F_{табл}^B(\alpha=0,05; f_1=n-1=5-1=4; f_2=(k-1)(n-1)=12)=3,26;$$

$$(F_{набл}^A=2,03) < (F_{крит}^A=3,49) \Leftrightarrow \text{фактор А впливає значимо};$$

$$(F_{набл}^B=1,67) < (F_{крит}^B=3,26) \Leftrightarrow \text{фактор В впливає не значимо}.$$

Висновок: Як видно з результатів, розрахункове значення величини F -статистики для фактора А (тип добрива) **$F = 1,67$** , а критична область утворюється правостороннім інтервалом $(3,49; +\infty)$. Так як **$F = 1,67$** не влучає у критичну область, гіпотеза H_0^A **приймається**, тобто вважається, що в цьому експерименті **тип добрива не вплинув на врожайність**. Розрахункове значення величини F -статистики для фактора В (сорт пшениці) **$F = 2,03$** , а критична область утворюється правостороннім інтервалом $(3,259; +\infty)$. Оскільки **$F = 2,03$** не влучає у критичну область, гіпотеза H_0^B **приймається**, тобто вважається, що у даному експерименті **сорт пшениці також не вплинув на врожайність**.

Рішення задачі у середовищі EXCEL:



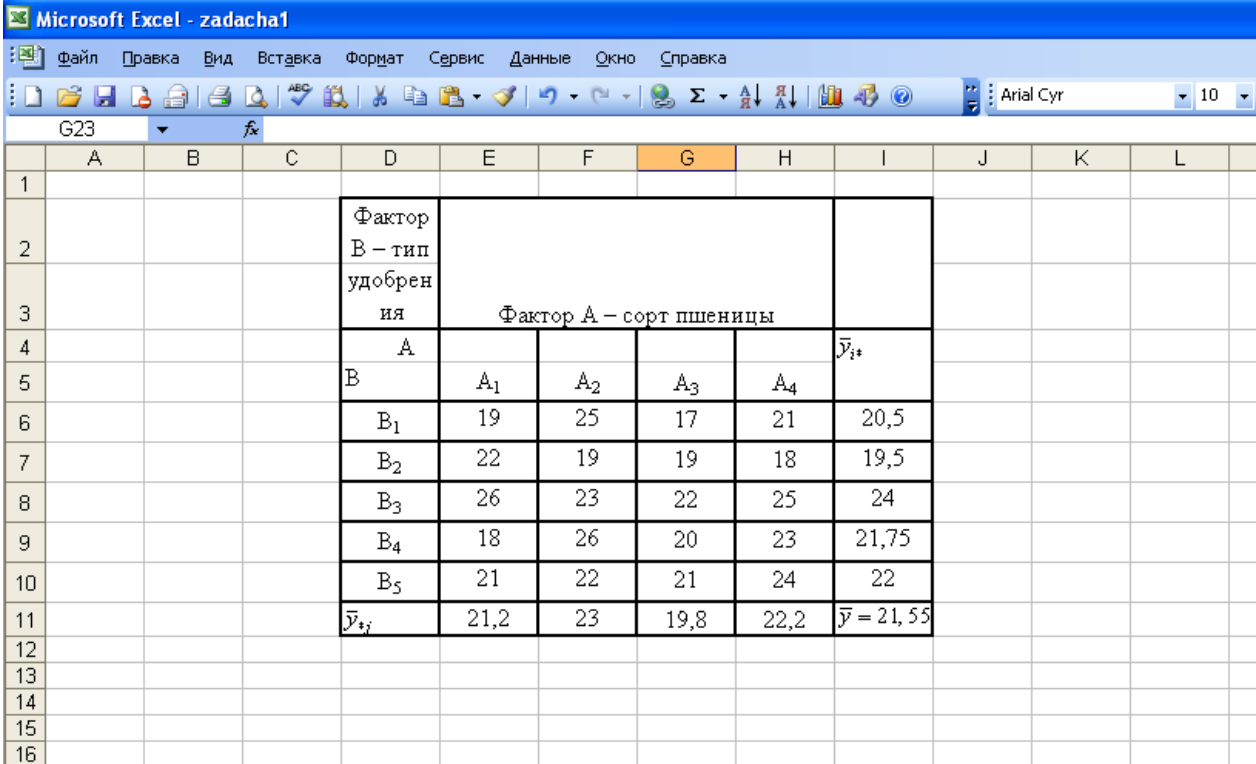
В системі електронних таблиць Microsoft Excel є набір інструментів для аналізу даних, званий **пакет аналізу**, який може бути використаний для вирішення складних статистичних задач. Для використання одного з цих інструментів необхідно вказати вхідні дані і вибрати параметри;

аналіз буде проведено за допомогою відповідної статистичної макрофункції, і результати будуть представлені у вихідному діапазоні.

У меню **Сервис** необхідно обрати команду **Анализ данных**. Якщо така команда відсутня у меню **Сервис**, то необхідно встановити у Microsoft Excel пакет *анализа данных*. Установка такого пакета здійснюється таким чином. В меню **Сервис** необхідно обрати команду **Надстройки**. Якщо у списку надбудов немає пакета *анализа данных*, то необхідно натиснути кнопку “Обзор” і задати диск, каталог і ім'я файлу для надбудови “Пакет анализа”, або запустити програму установки Microsoft Excel. При цьому необхідно встановити прапорець “Пакет анализа” (Надбудови, встановлені у Microsoft Excel, залишаються доступними, поки не будуть видалені).

Методика проведення двофакторного дисперсійного аналізу у середовищі EXCEL

1. Дані для задач двофакторного дисперсійного аналізу зазвичай записують у вигляді прямокутної таблиці, кожен стовпець якої - дані за рівнями першого фактора (A), рядки - дані за рівнями другого фактора (B). Значення X_{ij} – елементи випадкової величини, отримане при i -му рівні фактора B і при j -му рівні фактора A. (рис.7.8). При цьому сукупність комірок, що містять аналізовані дані, називається *вхідним діапазоном*.



	A	B	C	D	E	F	G	H	I	J	K	L
1												
2				Фактор B – тип удобрения	Фактор A – сорт пшеницы							
3												
4				A					$\bar{y}_{i\cdot}$			
5				B	A ₁	A ₂	A ₃	A ₄				
6				B ₁	19	25	17	21	20,5			
7				B ₂	22	19	19	18	19,5			
8				B ₃	26	23	22	25	24			
9				B ₄	18	26	20	23	21,75			
10				B ₅	21	22	21	24	22			
11				$\bar{y}_{\cdot j}$	21,2	23	19,8	22,2	$\bar{y} = 21,55$			
12												
13												
14												
15												
16												

Рис. 7.8. Формування вхідного діапазону для аналізу даних

2. У меню **Сервис** вибирають команду **Анализ данных**.

У списку інструментів статистичного аналізу вибирають **Двухфакторный дисперсионный анализ без повторений** (рис. 7.9).

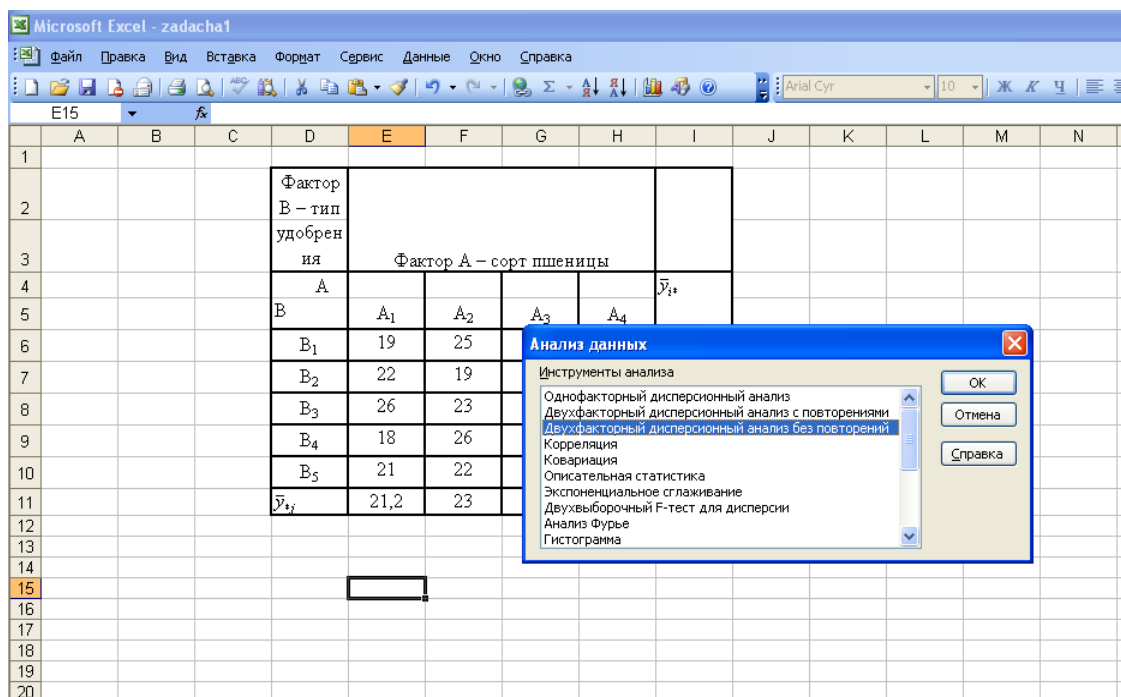


Рис. 7.9. Вибір інструмента аналізу

3. У діалоговому вікні *двухфакторный дисперсионный анализ* без повторень (рис.7.10) вказують вхідний інтервал, параметри виведення, альфа (рівень значущості).

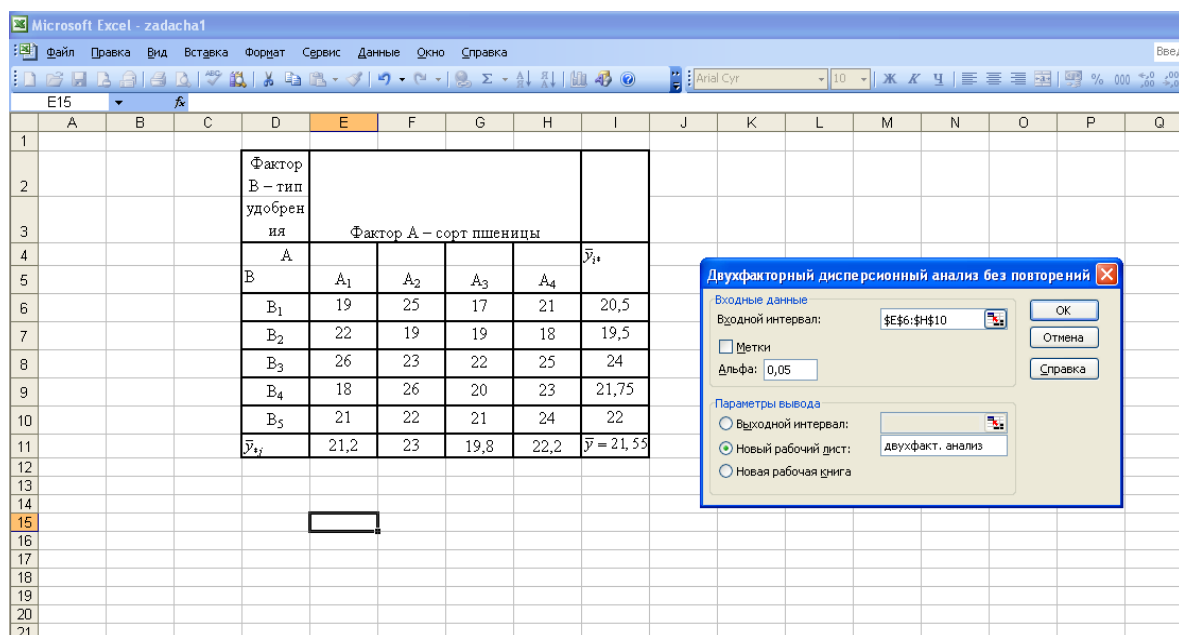


Рис. 7.10. Діалогове вікно двофакторного дисперсійного аналізу

Входной диапазон – це посилання на комірки, що містять аналізовані дані. Посилання має складатися як мінімум з двох суміжних діапазонів даних, організованих у вигляді стовпців і рядків. Вхідний інтервал можна задати за допомогою миші, або набрати на клавіатурі.

Альфа. Вкажіть рівень значущості (як правило - 0,05).

Параметры вывода. Введіть посилання на комірку (*выходной интервал*), розташовану в лівому верхньому кутку вихідного діапазону. Розміри вихідної області будуть розраховані автоматично, і відповідне повідомлення з'явиться на екрані в тому випадку, якщо вихідний діапазон займає місце існуючих даних або його розміри перевищують розміри листа

Новый лист. Встановіть перемикач, щоб відкрити новий лист в книзі і вставити результати аналізу, починаючи з комірки A1. Якщо в цьому є необхідність, введіть ім'я нового листа в полі, розташованому навпроти відповідного положення перемикача.

Новая книга. Встановіть перемикач, щоб відкрити нову книгу і вставити результати аналізу в комірку A1 на першому листі в цій книзі.

4. В результаті обробки даних створюється новий лист з ім'ям **двухфакт. анализ** з результатами, які представлені на рис. 7.11.

	A	B	C	D	E	F	G	H	I
1	Двухфакторный дисперсионный анализ без повторений								
2									
3	ИТОГИ	Счет	Сумма	Среднее	Дисперсия				
4	Строка 1	4	82	20,5	11,6666667				
5	Строка 2	4	78	19,5	3				
6	Строка 3	4	96	24	3,33333333				
7	Строка 4	4	87	21,75	12,25				
8	Строка 5	4	88	22	2				
9									
10	Столбец 1	5	106	21,2	9,7				
11	Столбец 2	5	115	23	7,5				
12	Столбец 3	5	99	19,8	3,7				
13	Столбец 4	5	111	22,2	7,7				
14									
15									
16	Дисперсионный анализ								
17	Источник вариации	SS	df	MS	F	P-Значение	F критическое		
18	Строки	46,2	4	11,55	2,03225806	0,153662099	3,259166727		
19	Столбцы	28,55	3	9,51666667	1,6744868	0,225106114	3,490294821		
20	Погрешность	68,2	12	5,68333333					
21									
22	Итого	142,95	19						
23									
24									
25									
26									
27									
28									
29									
30									
31									
32									
33									
34									

Рис. 7.11. Результаты двофакторного дисперсійного аналізу

Аналіз таблиці **ДВУХФАКТОРНЫЙ ДИСПЕРССИОННЫЙ АНАЛИЗ БЕЗ ПОВТОРЕНИЙ**.

Результати роботи процедури «Двухфакторный дисперсионный анализ» виводяться в двох таблицях.

Аналіз таблиці **ИТОГИ**.

“Счет” – число повторностей. “Сумма” – сума значень показника по рядках. “Среднее” – оцінка математичного очікування. “Дисперсия” – частинна дисперсія показника.

Аналіз таблиці **ДИСПЕРСИОННЫЙ АНАЛИЗ**.

Таблиця представляє результати дисперсійного аналізу двофакторного комплексу, в якому перша колонка “Источник вариации” містить найменування дисперсій. Графа “SS” - це сума квадратів відхилень, “df” - ступінь свободи, графа “MS” - середній квадрат, “F” - критерій

фактичного F - розподілу. "Р - значення" - ймовірність того, що дисперсія, відтворена рівнянням, дорівнює дисперсії залишків. Визначає ймовірність того, що отримана кількісна визначеність взаємозв'язку між факторами і результатом може вважатися випадковою. "F - критичне" - це значення F - теоретичного, яке згодом порівнюється з F - фактичним.

5. Аналіз результатів.

Як видно з результатів, розрахункове значення величини F -статистики для фактора А (тип добрива) $F = 1,67$, а критична область утворюється правостороннім інтервалом $(3,49; +\infty)$. Так як $F=1,67$ не влучає у критичну область, гіпотеза H_0^A **приймається**, тобто вважається, що в цьому експерименті **тип добрива не вплинув на врожайність**. Розрахункове значення величини F -статистики для фактора В (сорт пшениці) $F = 2,03$, а критична область утворюється правостороннім інтервалом $(3,259; +\infty)$. Оскільки $F=2,03$ не влучає у критичну область, гіпотеза H_0^B **приймається**, тобто вважається, що в даному експерименті **сорт пшениці також не вплинув на врожайність**.

Приклад 2.

Мають такі показники: Y - продуктивність праці; A - концентрація виробництва;
 B - тип обладнання.

Таблиця спостережень:

$\begin{matrix} B \\ A \end{matrix}$	B_1	B_2	$\bar{y}_{i*}$
A_1	25	30	27,5
A_2	50	60	55
A_3	80	90	85
$\bar{y}_{*j}$	51,7	60	$\bar{y} = 55,8$

При рівні значущості $\alpha=0,05$ перевірити чи є значимим вплив факторів А і В на вихідну величину Y , тобто перевірити гіпотези:

$$H_0^A : m_1^A = m_2^A = m_3^A \Leftrightarrow D_A = D_{\text{ост}} \Leftrightarrow \text{фактор А впливає не значимо};$$

$$H_1^A : m_1^A \neq m_2^A \neq m_3^A \Leftrightarrow D_A > D_{\text{ост}} \Leftrightarrow \text{фактор А впливає значимо};$$

$$H_0^B : m_1^B = m_2^B \Leftrightarrow D_B = D_{\text{ост}} \Leftrightarrow \text{фактор В впливає не значимо};$$

$$H_1^B : m_1^B \neq m_2^B \Leftrightarrow D_B > D_{\text{ост}} \Leftrightarrow \text{фактор В впливає значимо};$$

Оцінки математичних очікувань попередньо обчислені і занесені в таблицю спостережень. Обчислимо оцінки дисперсій.

Однорідність дисперсій перевіряється за критерієм Фішера, для цього обчислимо F - статистику:

$$F_{\text{набл}}^A = \frac{\tilde{D}_A}{\tilde{D}_{\text{ост}}} = \frac{1654,2}{4,2} = 393,9;$$

$$F_{\text{набл}}^B = \frac{\tilde{D}_B}{\tilde{D}_{\text{ост}}} = \frac{103,35}{4,2} = 24,6;$$

При рівні значущості $\alpha=0,05$ по таблиці критичних точок розподілу Фішера визначимо критичне значення критерію

$$F^A_{\text{крит}} = F^A_{\text{табл}}(\alpha=0,05; f_1=k-1=3-1=2; f_2=(k-1)(n-1)=2 \cdot 1=2)=19;$$

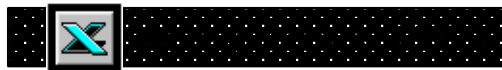
$$F^B_{\text{крит}} = F^B_{\text{табл}}(\alpha=0,05; f_1=n-1=2-1=1; f_2=(k-1)(n-1)=2)=18,51;$$

$$(F^A_{\text{набл}}=393,9) > (F^A_{\text{крит}}=19) \Leftrightarrow \text{фактор А впливає значимо};$$

$$(F^B_{\text{набл}}=24,6) > (F^B_{\text{крит}}=18,51) \Leftrightarrow \text{фактор В впливає значимо}.$$

Висновок: на продуктивність праці в цій групі підприємств значимо впливає і концентрація виробництва і тип обладнання.

Розв'язання задачі в середовищі EXCEL:



- Дані для задач двофакторного дисперсійного аналізу зазвичай записують у вигляді прямокутної таблиці (рис.7.12). При цьому сукупність комірок, що містять аналізовані дані, називається **вхідним діапазоном**.

	B	B ₁	B ₂	$\bar{y}_i$
A				
A ₁	25	30	27,5	
A ₂	50	60	55	
A ₃	80	90	85	
$\bar{y}_j$	51,7	60	$\bar{y} = 55,8$	

Рис. 7.12. Формування вхідного діапазону для аналізу даних

- У меню **Сервис** вибирають команду **Анализ данных**. У списку інструментів статистичного аналізу вибирають **Двухфакторный дисперсионный анализ без повторений** (рис.7.13).

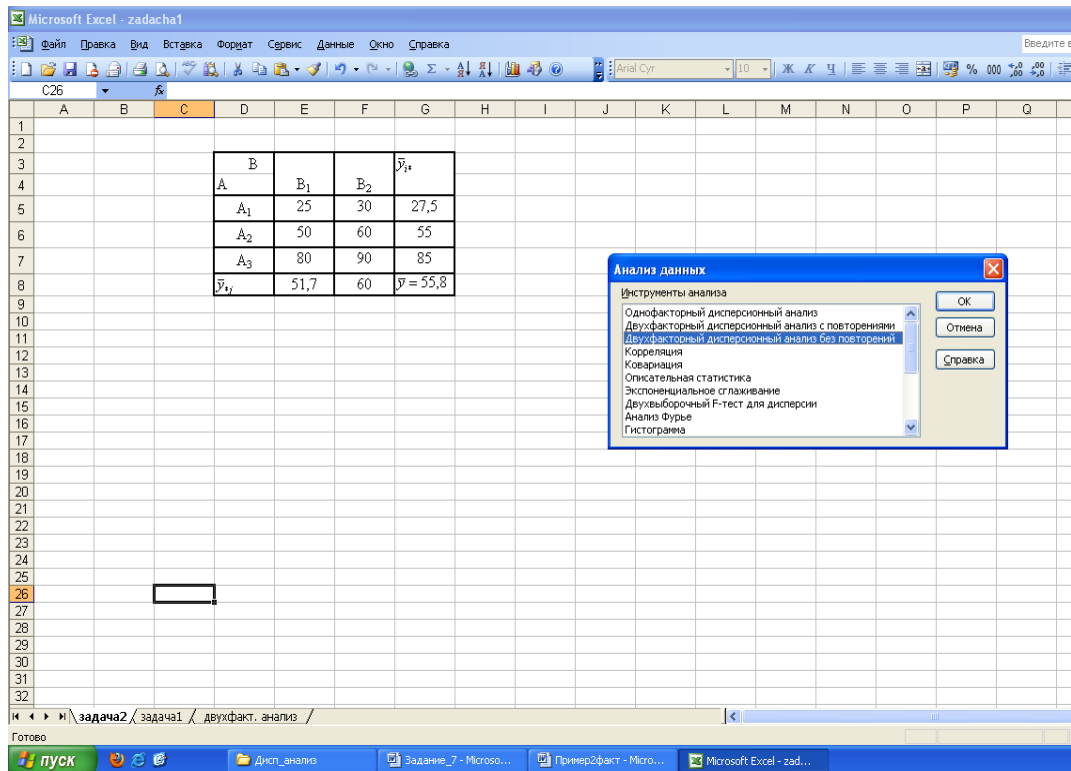


Рис. 7.13. Вибір інструменту аналізу

3. У діалоговому вікні *двухфакторный дисперсионный анализ без повторений* вказують вхідний інтервал, вихідний інтервал, альфа (рівень значущості) (рис. 7.14).

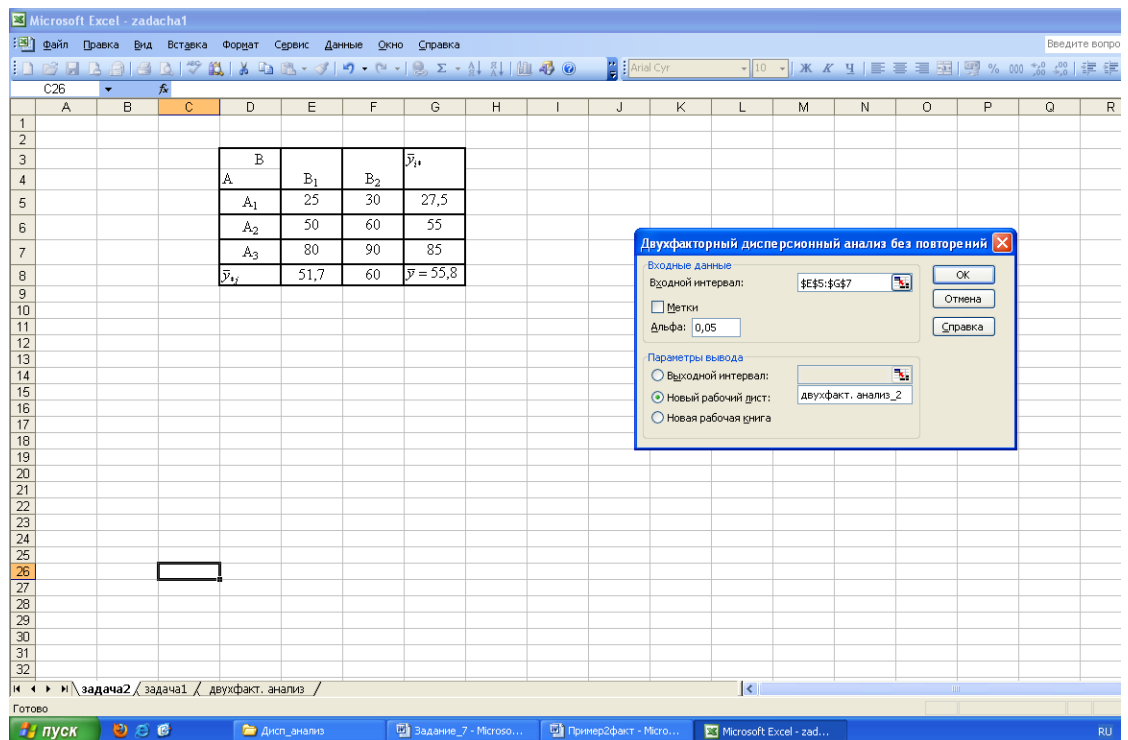


Рис. 7.14. Діалогове вікно двофакторного дисперсійного аналізу

Входной диапазон – це посилання на комірки, що містять аналізовані дані. Посилання має складатися як мінімум з двох суміжних діапазонів даних, організованих у вигляді стовпців і рядків. Вхідний інтервал можна задати за допомогою миші, або набрати на клавіатурі.

Альфа. Вкажіть рівень значущості (як правило- 0,05).

Параметры вывода. Введіть посилання на комірку (вихідний інтервал), розташовану в лівому верхньому кутку вихідного діапазону. Розміри вихідної області будуть розраховані автоматично, і відповідне повідомлення з'явиться на екрані в тому випадку, якщо вихідний діапазон займає місце існуючих даних або його розміри перевищують розміри листа.

Новый лист. Встановіть перемикач, щоб відкрити новий лист в книзі і вставити результати аналізу, починаючи з комірки A1. Якщо в цьому є необхідність, введіть ім'я нового листа в полі, розташованому навпроти відповідного положення перемикача.

Новая книга. Встановіть перемикач, щоб відкрити нову книгу і вставити результати аналізу в комірку A1 на першому листі в цій книзі.

4. В результаті обробки даних створюється новий лист з ім'ям **двухфакторный анализ** з результатами, які представлені на рис. 7.15.

	A	B	C	D	E	F	G	H	I
1	Двухфакторный дисперсионный анализ без повторений								
2									
3	ИТОГИ	Счет	Сумма	Среднее	Дисперсия				
4	Строка 1	2	55	27,5	12,5				
5	Строка 2	2	110	55	50				
6	Строка 3	2	170	85	50				
7									
8	Столбец 1	3	155	51,66667	758,3333				
9	Столбец 2	3	180	60	900				
10									
11									
12	Дисперсионный анализ								
13	Источник вариации	SS	df	MS	F	P-Значение	F критическое		
14	Строки	3308,333	2	1654,167	397	0,002512563	19		
15	Столбцы	104,1667	1	104,1667	25	0,037749551	18,51282051		
16	Погрешность	8,333333	2	4,166667					
17									
18	Итого	3420,833	5						
19									
20									
21									
22									
23									
24									
25									
26									
27									
28									
29									
30									
31									
32									
33									
34									

Рис. 7.15. Результаты двофакторного дисперсійного аналізу

Аналіз таблиці ДВУХФАКТОРНИЙ ДИСПЕРСИОННИЙ АНАЛІЗ БЕЗ ПОВТОРЕНИЙ.

Результати роботи процедури «Двухфакторный дисперсионный анализ» виводяться в двох таблицях.

Аналіз таблиці ИТОГИ.

“Счет” – число повторностей. “Сумма” – сума значень показника по рядках. “Среднее” – оцінка математичного очікування. “Дисперсия” – частинна дисперсія показника.

Аналіз таблиці ДИСПЕРСИОННИЙ АНАЛІЗ.

Таблиця представляє результати дисперсійного аналізу двофакторного комплексу, в якому перша колонка “Источник вариации” містить найменування дисперсій. Графа “SS” - це сума квадратів відхилень, “df” - ступінь свободи, графа “MS” - середній квадрат, “F” - критерій фактичного F - розподілу. “P - значення” - ймовірність того, що дисперсія, відтворена рівнянням, дорівнює дисперсії залишків. Визначає ймовірність того, що отримана кількісна визначеність взаємозв'язку між факторами і результатом може вважатися випадковою. “F - критичне” - це те значення «F - теоретичного», яке згодом порівнюється з «F - фактичним».

5. Аналіз результатів.

Як видно з результатів, розрахункове значення величини F-статистики для фактора А (концентрація виробництва) $F = 393,6$, а критична область утворюється правостороннім інтервалом $(19; +\infty)$. Оскільки $F=393,9$ попадає в критичну область, гіпотеза H_0^A **не береться**, тобто вважається, що в цьому експерименті **концентрація виробництва мала вплив на продуктивність праці**.

Розрахункове значення величини F-статистики для фактора В (тип обладнання) $F = 24,6$, а критична область утворюється правостороннім інтервалом $(18,51; +\infty)$. Оскільки $F=24,6$ попадає у критичну область, гіпотезу H_0^B не приймається, тобто. вважається, що в даному експерименті тип обладнання також вплинув на продуктивність праці.

Таким чином отримують:

$$(F_{\text{набл}}^A=393,9) > (F_{\text{крит}}^A=19) \Leftrightarrow \text{фактор А впливає значимо};$$

$$(F_{\text{набл}}^B=24,6) > (F_{\text{крит}}^B=18,51) \Leftrightarrow \text{фактор В впливає значимо}.$$

Варіанти індивідуальних завдань.

ОДНОФАКТОРНИЙ ДИСПЕРСІЙНИЙ АНАЛІЗ

Задача. На заводі встановлено чотири лінії з випуску облицювальної плитки. З кожної лінії випадковим чином протягом зміни відібрано 6 плиток і зроблені виміри їх товщини (мм). Відхилення від номінального розміру наведені в табл. Потрібно на рівні значимості 5% встановити залежність випуску якісних плиток від лінії випуску (фактор А).

Варіант 1.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,6	0,2	0,4	0,5	0,8	0,2
2	0,2	0,2	0,4	0,3	0,3	0,6
3	0,8	0,6	0,2	0,4	0,9	1,1
4	0,7	0,7	0,3	0,3	0,2	0,8

Варіант 2.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,5	0,1	0,4	0,8	0,8	0,3
2	0,2	0,4	0,4	0,3	0,6	0,6
3	0,9	0,6	0,1	0,5	0,9	1,0
4	0,7	0,8	0,3	0,2	0,2	0,4

Варіант 3.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,2	0,2	0,4	0,4	0,8	0,2
2	0,2	0,5	0,4	0,7	0,7	0,6
3	0,9	0,6	0,1	0,3	0,9	1,2
4	0,5	0,7	0,3	0,3	0,3	0,9

Варіант 4.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,7	0,3	0,4	0,5	0,8	0,2
2	0,3	0,9	0,4	0,3	0,3	0,9
3	0,9	0,3	0,3	0,3	0,9	1,0
4	0,5	0,7	0,3	0,5	0,2	0,8

Варіант 5.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,2	0,2	0,4	0,4	0,8	0,9
2	0,1	0,2	0,1	0,3	0,2	0,6
3	0,9	0,9	0,1	0,4	0,9	0,8
4	0,6	0,5	0,2	0,3	0,2	0,9

Варіант 6.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,4	0,2	0,4	0,5	0,8	0,1
2	0,2	0,4	0,4	0,3	0,3	0,9
3	0,5	0,6	0,6	0,4	0,8	0,8
4	0,3	0,7	0,3	0,3	0,2	0,9

Варіант 7.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,1	0,2	0,1	0,5	0,8	0,2
2	0,3	0,2	0,4	0,5	0,3	0,9
3	0,5	0,4	0,2	0,2	0,9	1,2
4	0,6	0,7	0,4	0,3	0,4	0,8

Варіант 8.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,3	0,2	0,4	0,7	0,8	0,1
2	0,2	0,5	0,4	0,5	0,3	0,8
3	0,7	0,6	0,1	0,4	1,0	0,9
4	0,5	0,7	0,4	0,2	0,2	0,7

Варіант 9.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,4	0,1	0,4	0,7	0,8	0,1
2	0,3	0,2	0,5	0,3	0,7	0,6
3	0,6	0,6	0,1	0,9	0,9	0,8
4	0,7	0,6	0,3	0,2	0,2	0,9

Варіант 10.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,1	0,1	0,4	0,5	0,8	0,9
2	0,4	0,2	0,4	0,3	0,3	0,8
3	0,3	0,6	0,3	0,4	0,9	1,0
4	0,7	0,6	0,3	0,3	0,3	0,9

Варіант 11.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,5	0,4	0,4	0,5	0,6	0,1
2	0,2	0,4	0,4	0,3	0,3	0,5
3	0,8	0,2	0,3	0,7	0,9	1,2
4	0,3	0,4	0,3	0,5	0,1	0,8

Варіант 12.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,3	0,6	0,4	0,6	0,8	0,1
2	0,1	0,2	0,5	0,1	0,3	0,7
3	0,8	0,7	0,1	0,4	0,9	1,0
4	0,5	0,4	0,3	0,2	0,2	0,9

Варіант 13.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,2	0,2	0,3	0,5	0,9	0,2
2	0,2	0,7	0,9	0,3	0,9	0,6
3	0,4	0,5	0,2	0,4	0,9	0,9
4	0,6	0,7	0,4	0,3	0,1	0,7

Варіант 14.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,5	0,1	0,4	0,5	0,9	0,3
2	0,3	0,2	0,7	0,3	0,1	0,9
3	0,9	0,5	0,1	0,4	0,9	1,2
4	0,9	1,1	0,3	0,7	0,2	0,8

Варіант 15.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,8	0,2	0,5	0,5	0,9	0,1
2	0,2	0,1	0,4	0,3	0,6	0,9
3	0,5	0,6	0,2	0,4	0,9	1,1
4	0,7	0,9	0,3	0,1	0,2	0,8

Варіант 16.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,3	1,2	0,6	0,5	0,9	0,1
2	0,2	0,2	0,4	0,2	0,3	0,9
3	0,8	0,6	0,2	0,4	0,9	0,8
4	0,7	0,9	0,3	0,4	0,2	0,8

Варіант 17.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,6	0,1	0,3	0,5	0,3	0,2
2	0,2	0,4	0,4	0,7	0,8	0,6
3	0,9	0,6	1,2	0,4	0,9	1,0
4	0,7	0,5	0,3	0,2	0,2	0,9

Варіант 18.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,2	0,2	0,5	0,5	0,8	0,2
2	0,3	0,1	0,4	0,3	0,3	0,7
3	0,9	0,6	0,3	0,4	0,8	1,1
4	0,7	0,6	0,3	0,4	0,2	1,0

Варіант 19.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,5	0,1	0,4	0,6	0,8	0,2
2	0,3	0,2	0,4	0,8	0,9	0,6
3	0,5	0,6	0,3	0,9	0,9	0,8
4	0,6	0,9	0,3	0,2	0,2	0,8

Варіант 20.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,9	0,1	0,3	0,5	0,9	0,2
2	0,1	0,7	0,4	0,6	0,3	0,6
3	0,9	0,6	0,2	0,4	0,9	1,1
4	0,7	0,9	0,4	0,5	0,9	0,8

Варіант 21.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,6	0,6	0,4	0,5	0,9	0,1
2	0,2	0,2	0,4	0,4	0,3	0,9
3	0,8	0,7	0,2	0,2	0,9	1,0
4	0,7	0,9	0,3	0,5	0,1	0,8

Варіант 22.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,9	0,1	0,3	0,5	0,8	0,1
2	0,4	0,2	0,9	0,3	0,1	0,6
3	0,9	0,9	0,3	0,9	0,9	1,1
4	0,6	0,7	0,4	0,3	0,2	0,8

Варіант 23.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,1	0,2	0,6	0,5	0,9	0,2
2	0,9	0,2	0,7	0,3	0,3	0,9
3	0,3	0,6	0,5	0,4	0,9	0,7
4	0,7	0,9	0,3	0,3	0,1	0,8

Варіант 24.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,5	1,2	0,7	0,5	0,8	0,9
2	0,6	0,9	0,4	0,3	1,1	0,6
3	0,4	0,9	0,2	0,4	0,9	1,2
4	0,8	0,7	0,3	1,3	0,2	0,8

Варіант 25.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,6	0,1	0,1	0,7	0,8	1,2
2	0,1	0,2	0,4	0,8	0,3	1,1
3	0,8	0,6	0,3	0,4	0,8	1,1
4	0,7	0,5	0,3	0,3	0,2	0,8

Варіант 26.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,6	0,2	0,4	0,7	0,8	0,9
2	0,2	0,2	0,4	0,3	0,1	0,6
3	0,9	0,6	0,2	0,4	0,9	1,0
4	1,2	0,8	0,3	0,1	1,2	0,8

Варіант 27.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,7	0,1	0,4	0,6	0,9	0,1
2	0,2	0,3	0,4	0,3	1,3	0,7
3	0,7	0,1	0,2	0,5	0,9	1,1
4	0,9	0,7	0,8	0,3	0,8	0,8

Варіант 28.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,1	0,2	0,5	0,8	0,8	0,7
2	0,2	0,3	0,4	1,3	0,3	0,6
3	0,1	0,6	0,1	0,4	0,3	1,1
4	0,9	0,7	0,3	0,5	0,1	0,8

Варіант 29.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,5	0,2	0,9	0,5	0,8	1,2
2	0,2	0,3	0,9	0,3	0,3	0,9
3	0,8	0,1	0,2	0,4	0,2	1,1
4	0,7	0,1	0,3	0,1	0,2	0,8

Варіант 30.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,1	0,2	0,1	0,5	0,8	1,2
2	0,6	0,2	0,4	0,6	0,3	0,6
3	0,7	0,6	0,3	1,2	0,9	1,1
4	0,5	0,7	0,3	0,5	0,2	0,8

Варіант 31.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,6	0,2	0,4	0,5	0,8	0,2
2	0,2	0,2	0,4	0,3	0,3	0,6
3	0,8	0,6	0,2	0,4	0,9	1,1
4	0,7	0,7	0,3	0,3	0,2	0,8

Варіант 32.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,5	0,1	0,4	0,8	0,8	0,3
2	0,2	0,4	0,4	0,3	0,6	0,6
3	0,9	0,6	0,1	0,5	0,9	1,0
4	0,7	0,8	0,3	0,2	0,2	0,4

Варіант 33.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,2	0,2	0,4	0,4	0,8	0,2
2	0,2	0,5	0,4	0,7	0,7	0,6
3	0,9	0,6	0,1	0,3	0,9	1,2
4	0,5	0,7	0,3	0,3	0,3	0,9

Варіант 34.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,7	0,3	0,4	0,5	0,8	0,2
2	0,3	0,9	0,4	0,3	0,3	0,9
3	0,9	0,3	0,3	0,3	0,9	1,0
4	0,5	0,7	0,3	0,5	0,2	0,8

Варіант 35.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,2	0,2	0,4	0,4	0,8	0,9
2	0,1	0,2	0,1	0,3	0,2	0,6
3	0,9	0,9	0,1	0,4	0,9	0,8
4	0,6	0,5	0,2	0,3	0,2	0,9

Варіант 36.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,4	0,2	0,4	0,5	0,8	0,1
2	0,2	0,4	0,4	0,3	0,3	0,9
3	0,5	0,6	0,6	0,4	0,8	0,8
4	0,3	0,7	0,3	0,3	0,2	0,9

Варіант 37.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,1	0,2	0,1	0,5	0,8	0,2
2	0,3	0,2	0,4	0,5	0,3	0,9
3	0,5	0,4	0,2	0,2	0,9	1,2
4	0,6	0,7	0,4	0,3	0,4	0,8

Варіант 38.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,3	0,2	0,4	0,7	0,8	0,1
2	0,2	0,5	0,4	0,5	0,3	0,8
3	0,7	0,6	0,1	0,4	1,0	0,9
4	0,5	0,7	0,4	0,2	0,2	0,7

Варіант 39.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,4	0,1	0,4	0,7	0,8	0,1
2	0,3	0,2	0,5	0,3	0,7	0,6
3	0,6	0,6	0,1	0,9	0,9	0,8
4	0,7	0,6	0,3	0,2	0,2	0,9

Варіант 40.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,1	0,1	0,4	0,5	0,8	0,9
2	0,4	0,2	0,4	0,3	0,3	0,8
3	0,3	0,6	0,3	0,4	0,9	1,0
4	0,7	0,6	0,3	0,3	0,3	0,9

Варіант 41.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,9	0,1	0,3	0,5	0,9	0,2
2	0,1	0,7	0,4	0,6	0,3	0,6
3	0,9	0,6	0,2	0,4	0,9	1,1
4	0,7	0,9	0,4	0,5	0,9	0,8

Варіант 42.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,6	0,6	0,4	0,5	0,9	0,1
2	0,2	0,2	0,4	0,4	0,3	0,9
3	0,8	0,7	0,2	0,2	0,9	1,0
4	0,7	0,9	0,3	0,5	0,1	0,8

Варіант 43.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,9	0,1	0,3	0,5	0,8	0,1
2	0,4	0,2	0,9	0,3	0,1	0,6
3	0,9	0,9	0,3	0,9	0,9	1,1
4	0,6	0,7	0,4	0,3	0,2	0,8

Варіант 44.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,1	0,2	0,6	0,5	0,9	0,2
2	0,9	0,2	0,7	0,3	0,3	0,9
3	0,3	0,6	0,5	0,4	0,9	0,7
4	0,7	0,9	0,3	0,3	0,1	0,8

Варіант 45.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,5	1,2	0,7	0,5	0,8	0,9
2	0,6	0,9	0,4	0,3	1,1	0,6
3	0,4	0,9	0,2	0,4	0,9	1,2
4	0,8	0,7	0,3	1,3	0,2	0,8

Варіант 46.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,6	0,1	0,1	0,7	0,8	1,2
2	0,1	0,2	0,4	0,8	0,3	1,1
3	0,8	0,6	0,3	0,4	0,8	1,1
4	0,7	0,5	0,3	0,3	0,2	0,8

Варіант 47.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,6	0,2	0,4	0,7	0,8	0,9
2	0,2	0,2	0,4	0,3	0,1	0,6
3	0,9	0,6	0,2	0,4	0,9	1,0
4	1,2	0,8	0,3	0,1	1,2	0,8

Варіант 48.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,7	0,1	0,4	0,6	0,9	0,1
2	0,2	0,3	0,4	0,3	1,3	0,7
3	0,7	0,1	0,2	0,5	0,9	1,1
4	0,9	0,7	0,8	0,3	0,8	0,8

Варіант 49.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,1	0,2	0,5	0,8	0,8	0,7
2	0,2	0,3	0,4	1,3	0,3	0,6
3	0,1	0,6	0,1	0,4	0,3	1,1
4	0,9	0,7	0,3	0,5	0,1	0,8

Варіант 50.

Лінія з випуску плитки (фактор А)	Номер випробування					
	1	2	3	4	5	6
1	0,5	0,2	0,9	0,5	0,8	1,2
2	0,2	0,3	0,9	0,3	0,3	0,9
3	0,8	0,1	0,2	0,4	0,2	1,1
4	0,7	0,1	0,3	0,1	0,2	0,8

ДВОФАКТОРНИЙ ДИСПЕРСІЙНИЙ АНАЛІЗ

Задача 1. Визначити чи значимий при $\alpha = 0,025$ вплив концентрації виробництва (фактор A з рівнями: A_1 – дрібні підприємства, A_2 – середні, A_3 – великі) і типу обладнання (фактор B з рівнями: B_1 – прогресивне, B_2 – звичайне) на продуктивність праці (вихідна величина U в рублях, зменшена в 100 разів), в наступних групах підприємств, якщо відомо, що передумова дисперсійного аналізу виконуються, а таблиці спостережень мають такий вигляд:

1.

$\begin{matrix} B \\ A \end{matrix}$	B_1	B_2
A_1	28	32
A_2	53	61
A_3	84	81

2.

$\begin{matrix} B \\ A \end{matrix}$	B_1	B_2
A_1	33	37
A_2	53	65
A_3	93	91

3.

$\begin{matrix} B \\ A \end{matrix}$	B_1	B_2
A_1	27	33
A_2	52	60
A_3	85	80

4.

$\begin{matrix} B \\ A \end{matrix}$	B_1	B_2
A_1	23	34
A_2	58	69
A_3	88	93

5.

$\begin{matrix} B \\ A \end{matrix}$	B_1	B_2
A_1	30	33
A_2	61	53
A_3	82	94

6.

$\begin{matrix} B \\ A \end{matrix}$	B_1	B_2
A_1	36	27
A_2	64	51
A_3	90	83

7.

$\begin{matrix} B \\ A \end{matrix}$	B_1	B_2
A_1	25	38
A_2	59	66
A_3	86	94

8.

$\begin{matrix} B \\ A \end{matrix}$	B_1	B_2
A_1	24	23
A_2	46	67
A_3	75	73

9.

$\begin{matrix} B \\ A \end{matrix}$	B_1	B_2
A_1	24	38
A_2	56	69
A_3	91	87

10.

$\begin{matrix} B \\ A \end{matrix}$	B_1	B_2
A_1	34	23
A_2	56	57
A_3	95	83

11.

$\begin{matrix} B \\ A \end{matrix}$	B_1	B_2
A_1	38	26
A_2	62	55
A_3	95	81

12.

$\begin{matrix} B \\ A \end{matrix}$	B_1	B_2
A_1	44	26
A_2	46	67
A_3	75	93

13.

A \ B	B ₁	B ₂
A ₁	56	33
A ₂	34	47
A ₃	87	93

14.

A \ B	B ₁	B ₂
A ₁	53	43
A ₂	66	57
A ₃	85	73

15.

A \ B	B ₁	B ₂
A ₁	64	43
A ₂	46	37
A ₃	75	93

16.

A \ B	B ₁	B ₂
A ₁	58	24
A ₂	36	47
A ₃	86	93

17.

A \ B	B ₁	B ₂
A ₁	44	33
A ₂	66	47
A ₃	85	63

18.

A \ B	B ₁	B ₂
A ₁	24	29
A ₂	46	47
A ₃	98	63

19.

A \ B	B ₁	B ₂
A ₁	64	43
A ₂	46	57
A ₃	75	93

20.

A \ B	B ₁	B ₂
A ₁	24	33
A ₂	46	37
A ₃	65	73

21.

A \ B	B ₁	B ₂
A ₁	24	21
A ₂	36	37
A ₃	65	63

22.

A \ B	B ₁	B ₂
A ₁	64	33
A ₂	36	27
A ₃	65	73

23.

A \ B	B ₁	B ₂
A ₁	24	37
A ₂	76	27
A ₃	35	33

24.

A \ B	B ₁	B ₂
A ₁	26	29
A ₂	36	47
A ₃	65	53

25.

A \ B	B ₁	B ₂
A ₁	64	43
A ₂	26	37
A ₃	55	63

26.

A \ B	B ₁	B ₂
A ₁	64	43
A ₂	55	77
A ₃	65	43

27.

A \ B	B ₁	B ₂
A ₁	54	43
A ₂	76	67
A ₃	85	93

28.

A \ B	B ₁	B ₂
A ₁	27	34
A ₂	42	67
A ₃	56	73

29.

A \ B	B ₁	B ₂
A ₁	44	33
A ₂	67	27
A ₃	87	95

30.

A \ B	B ₁	B ₂
A ₁	64	34
A ₂	65	37
A ₃	85	33

31.

32.

33.

34.

$\begin{smallmatrix} B \\ A \end{smallmatrix}$	B_1	B_2
A_1	28	32
A_2	53	61
A_3	84	81

35.

$\begin{smallmatrix} B \\ A \end{smallmatrix}$	B_1	B_2
A_1	33	37
A_2	53	65
A_3	93	91

36.

$\begin{smallmatrix} B \\ A \end{smallmatrix}$	B_1	B_2
A_1	27	33
A_2	52	60
A_3	85	80

38.

$\begin{smallmatrix} B \\ A \end{smallmatrix}$	B_1	B_2
A_1	23	34
A_2	58	69
A_3	88	93

$\begin{smallmatrix} B \\ A \end{smallmatrix}$	B_1	B_2
A_1	30	33
A_2	61	53
A_3	82	94

$\begin{smallmatrix} B \\ A \end{smallmatrix}$	B_1	B_2
A_1	36	27
A_2	64	51
A_3	90	83

$\begin{smallmatrix} B \\ A \end{smallmatrix}$	B_1	B_2
A_1	25	38
A_2	59	66
A_3	86	94

$\begin{smallmatrix} B \\ A \end{smallmatrix}$	B_1	B_2
A_1	24	23
A_2	46	67
A_3	75	73

39.

$\begin{smallmatrix} B \\ A \end{smallmatrix}$	B_1	B_2
A_1	24	38
A_2	56	69
A_3	91	87

42.

$\begin{smallmatrix} B \\ A \end{smallmatrix}$	B_1	B_2
A_1	34	23
A_2	56	57
A_3	95	83

40.

41.

$\begin{smallmatrix} B \\ A \end{smallmatrix}$	B_1	B_2
A_1	38	26
A_2	62	55
A_3	95	81

$\begin{smallmatrix} B \\ A \end{smallmatrix}$	B_1	B_2
A_1	44	26
A_2	46	67
A_3	75	93

43.

$\begin{smallmatrix} B \\ A \end{smallmatrix}$	B_1	B_2
A_1	64	43
A_2	46	57
A_3	75	93

44.

$\begin{smallmatrix} B \\ A \end{smallmatrix}$	B_1	B_2
A_1	24	33
A_2	46	37
A_3	65	73

45.

$\begin{smallmatrix} B \\ A \end{smallmatrix}$	B_1	B_2
A_1	24	21
A_2	36	37
A_3	65	63

46.

$\begin{smallmatrix} B \\ A \end{smallmatrix}$	B_1	B_2
A_1	64	33
A_2	36	27
A_3	65	73

47.

$\begin{smallmatrix} B \\ A \end{smallmatrix}$	B_1	B_2
A_1	24	37
A_2	76	27
A_3	35	33

48.

$\begin{smallmatrix} B \\ A \end{smallmatrix}$	B_1	B_2
A_1	26	29
A_2	36	47
A_3	65	53

49.

$\begin{smallmatrix} B \\ A \end{smallmatrix}$	B_1	B_2
A_1	64	43
A_2	26	37
A_3	55	63

50.

$\begin{smallmatrix} B \\ A \end{smallmatrix}$	B_1	B_2
A_1	64	43
A_2	55	77
A_3	65	43

Задача 2. При рівні значущості $\alpha = 0,05$ визначити вплив факторів A і B і їх спільний вплив на досліджувану ознаку.

Варіант 1.

Досліджується вплив на зносостійкість деталей наступних факторів: A - матеріал для виготовлення деталей (застосовували три види сталі) і B - технологія виготовлення деталей (дві технології). Результати експерименту наведені у таблиці:

Фактор B	Фактор A		
	A_1	A_2	A_3
B_1	10	8	12
B_2	12	11	14

Варіант 2.

Досліджується вплив на врожайність ячменю наступних факторів: A : посів здійснюється після чорного пару - A_1 ; після коренеплодів - A_2 ; після колосових культур - A_3 ; B - сортність ячменю (три сорти). Результати досліджень наведені у таблиці:

Фактор B	Фактор A		
	A_1	A_2	A_3
B_1	32,4	38,6	42,3
B_2	38,9	40,5	50,2
B_3	34,9	30,8	32,3

Варіант 3.

Досліджується вплив на міцність чавуну двох факторів: A - вміст кремнію у чавуні, а саме A_1 - 0,24%; A_2 - 0,42%; A_3 - 0,52%; B - температурний режим плавлення (два режими). Результати досліджень наведені у таблиці:

Фактор B	Фактор A		
	A_1	A_2	A_3
B_1	44,5	41,8	51,8
B_2	33,4	38,9	30,6

Варіант 4.

Досліджується вплив на врожайність кукурудзи двох чинників: А - внесення добрив у ґрунт (три добрива); В - глибина поливу землі (три глибини поливу). Результати досліджень наведені у таблиці:

Фактор В	Фактор А		
	А ₁	А ₂	А ₃
В ₁	32,6	28,2	38,9
В ₂	44,2	36,4	46,8
В ₃	38,5	40,1	47,8

Варіант 5.

Досліджується вплив факторів А і В на число виготовлених втулок зі ста взятих, які відповідають нормам стандарту: А - використовували дві технології виготовлення; В - заготовки надходили з трьох заводів. Результати досліджень наведені у таблиці:

Фактор В	Фактор А	
	А ₁	А ₂
В ₁	90	96
В ₂	83	100
В ₃	75	86

Варіант 6 .

Досліджується вплив факторів А і В на продуктивність праці підприємства певної галузі промисловості: А - фондозабезпеченість (три рівня); В - рівень оплати праці працівникам (два рівня) Результатом дослідження наведені в таблиці:

Фактор В	Фактор А		
	А ₁	А ₂	А ₃
В ₁	9,5	4,2	3,24
В ₂	7,8	13,6	9,7
В ₃	12,1	22,5	15,8

Варіант 7.

Експериментально досліджувався вплив на зносостійкість деталей факторів А і В: фактор А - тип сталі (три типи); фактор В - технологія виготовлення (дві технології).
Результати експерименту наведені в таблиці:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	10	12	14
B_2	8	11	13

Варіант 8.

Експериментально досліджувався вплив факторів А і В на опріснення морської води: фактор А - тип опріснювача (три типи); фактор В - три різних лабораторії. Результати експерименту (відсоток солі в опрісненій воді) наведені в таблиці:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	3,6	2,92	3,25
B_2	4,25	3,8	3,64
B_3	3,46	3,8	4,12

Варіант 9.

Експериментально досліджувався вплив факторів А і В на врожайність цукрових буряків в ц / га: фактор А - добрива (три види); фактор - сума температур за період вегетації (три рівня).
Результати досліджень наведені в таблиці:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	360	402	433
B_2	340	412	455
B_3	395	399	375

Варіант 10.

Досліджувався вплив факторів А і В на пружність стали в умовних одиницях: фактор А -% нікелю (три рівня); фактор В -% марганцю (два рівня) Результати досліджень наведені в таблиці:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	36,4	41,2	36,5
B_2	39,2	39,7	40,8

Варіант 11.

Досліджувався вплив факторів А і В на твердість сплаву: фактор А -% нікелю (три рівня); фактор В -% хрому (три рівня). Результати досліджень наведені в таблиці, де твердість подається в умовних одиницях:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	53,2	55,4	68,3
B_2	67,2	77,2	77,9
B_3	70,2	69,2	75,5

Варіант 12.

Досліджувався вплив на продуктивність праці факторів А і В: фактор А - текучість кадрів,% (три рівня); фактор В - рівень оплати праці, тис.грн. в рік (два рівня). Результати досліджень наведені в таблиці:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	15,62	10,83	12,44
B_2	16,52	13,24	6,81

Варіант 13.

Досліджувався вплив факторів А і В на приріст врожаю картоплі з гектара: фактор А - внесення в ґрунт добрива (три рівні за кількістю у відсотках); фактор В - полив ґрунту (Два рівня проникнення вологи на глибину ґрунту). Результати досліджень наведені в таблиці:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	4,25	7,44	6,81
B_2	5,76	8,28	6,32

Варіант 14.

Досліджувався вплив факторів А і В на масу корів в кг: фактор А - три види соковитих кормів; фактор В - два види грубих кормів з добавкою певного відсотка вітамінів. Результати досліджень наведені в таблиці:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	455,6	489,6	331,4
B_2	446,2	600,0	443,8

Варіант 15.

Досліджувався вплив факторів А і В на рейтинг лівих партій в відсотках: фактор А - регіони (A_1 - західний, A_2 - південний, A_3 - східний); фактор В - віковий рівень опитаних (B_1 - до 30 років, B_2 - 30-55 років і B_3 -55-70 років). Результати досліджень наведені в таблиці:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	9,5	4,2	3,24
B_2	7,8	13,6	9,7
B_3	12,1	22,5	15,8

Варіант 16.

Досліджувався вплив факторів А і В на вміст протеїну в курячих яйцях, %: фактор А - три типи зернових; фактор В - два типи комбікормів.
Результати досліджень наведені в таблиці:

Фактор В	Фактор А		
	А ₁	А ₂	А ₃
В ₁	18,2	13,9	16,1
В ₂	14,9	15,6	19,3

Варіант 17.

Досліджувався вплив факторів А і В на приріст врожаю соняшнику, ц / га: фактор А - три види добрив; фактор В - глибина зрошення (дві глибини). Результатом досліджень наведені в таблиці:

Фактор В	Фактор А		
	А ₁	А ₂	А ₃
В ₁	6,3	7,1	9,4
В ₂	8,3	7,4	6,3

Варіант 18.

Досліджувався вплив факторів А і В на рейтинг правих партій (у відсотках): фактор А - регіони (А₁ - західний, А₂ - центральний, А₃ - східний); фактор В - вік опитаних (В₁ - 20-35 років, В₂ - 35-50 років і В₃ - 50-70 років). Результати досліджень наведені в таблиці::

Фактор В	Фактор А		
	А ₁	А ₂	А ₃
В ₁	25,2	10,6	6,4
В ₂	4,3	12,4	7,9
В ₃	14,3	7,9	14,8

Варіант 19.

Досліджувався вплив факторів А і В на жирність молока, %: фактор А - три види соковитих кормів; фактор В - два види біододішок. Результати досліджень наведені в таблиці:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	3,25	4,2	3,11
B_2	3,41	4,39	4,41

Варіант 20.

Досліджувався вплив факторів А і В на річні надої молока в літрах: фактор А - три види соковитих кормів; фактор В - два види грубих кормів з різним рівнем частки, %, спеціальних вітамінів. Результати досліджень наведені в таблиці:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	5050	3900	12500
B_2	6900	8900	3200

Варіант 21.

Досліджується вплив на зносостійкість деталей наступних факторів: А - матеріал для виготовлення деталей (застосовували три види сталі) і В - технологія виготовлення деталей (дві технології) Результати експерименту наведені в таблиці:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	9	12	12
B_2	16	11	8

Варіант 22.

Досліджується вплив на врожайність ячменю наступних факторів: А: посів здійснюється після чорного пару - А₁; після коренеплодів - А₂; після колосових культур - А₃; В - сортність ячменю (три сорти). Результати досліджень наведені в таблиці:

Фактор В	Фактор А		
	А ₁	А ₂	А ₃
В ₁	36,4	38,6	42,3
В ₂	40,2	40,5	51,2
В ₃	34,9	28,8	30,3

Варіант 23.

Досліджується вплив на міцність чавуну двох факторів: А - вміст кремнію в чавуні, а саме А₁ - 0,24%; А₂ - 0,42%; А₃ - 0,52%; В - температурний режим плавлення (два режими). Результати досліджень наведені в таблиці:

Фактор В	Фактор А		
	А ₁	А ₂	А ₃
В ₁	44,9	40,8	54,8
В ₂	30,4	36,9	32,6

Варіант 24.

Досліджується вплив на врожайність кукурудзи двох чинників: А - внесення добрив в ґрунт (три добрива); В - глибина поливу землі (три глибини поливу). Результати досліджень наведені в таблиці:

Фактор В	Фактор А		
	А ₁	А ₂	А ₃
В ₁	31,6	29,2	39,9
В ₂	42,2	38,4	49,8
В ₃	37,5	41,1	45,8

Варіант 25.

Досліджується вплив факторів А і В на число виготовлених втулок зі ста взятих, які відповідають нормам стандарту: А - використовували дві технології виготовлення; В - заготовки надходили з трьох заводів. Результати досліджень наведені в таблиці:

Фактор В	Фактор А	
	A_1	A_2
B_1	96	99
B_2	82	100
B_3	76	89

Варіант 26 .

Досліджується вплив факторів А і В на продуктивність праці підприємства певної галузі промисловості: А - фондозабезпеченість (три рівня); В - рівень оплати праці працівникам (два рівня) Результатом дослідження наведені в таблиці:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	9,7	4,6	5,24
B_2	7,85	11,6	9,78
B_3	10,1	24,5	16,8

Варіант 27.

Експериментально досліджувався вплив на зносостійкість деталей факторів А і В: фактор А - тип сталі (три типи); фактор В - технологія виготовлення (дві технології). Результати експерименту наведені в таблиці:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	13	9	16
B_2	6	12	14

Варіант 28.

Експериментально досліджувався вплив факторів А і В на опріснення морської води: фактор А - тип опріснювача (три типи); фактор В - три різних лабораторії. Результати експерименту (відсоток солі в опрісненій воді) наведені в таблиці:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	3,64	4,92	6,25
B_2	5,27	4,8	3,6
B_3	4,46	3,82	4,14

Варіант 29.

Експериментально досліджувався вплив факторів А і В на врожайність цукрових буряків в ц / га: фактор А - добрива (три види); фактор В - сума температур за період вегетації (три рівня). Результати досліджень наведені в таблиці:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	350	422	435
B_2	340	417	465
B_3	397	367	383

Варіант 30.

Досліджувався вплив факторів А і В на пружність стали в умовних одиницях: фактор А - % нікелю (три рівня); фактор В - % марганцю (два рівня). Результати досліджень наведені в таблиці:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	36,4	41,2	36,5
B_2	39,2	39,7	40,8

Варіант 31.

Досліджується вплив на зносостійкість деталей наступних факторів: А - матеріал для виготовлення деталей (застосовували три види сталі) і В - технологія виготовлення деталей (дві технології). Результати експерименту наведені в таблиці:

Фактор В	Фактор А		
	А ₁	А ₂	А ₃
В ₁	10	8	12
В ₂	12	11	14

Варіант 32.

Досліджується вплив на врожайність ячменю наступних факторів: А: посів здійснюється після чорного пару - А₁; після коренеплодів - А₂; після колосових культур - А₃; В - сортність ячменю (три сорти). Результати досліджень наведені в таблиці:

Фактор В	Фактор А		
	А ₁	А ₂	А ₃
В ₁	32,4	38,6	42,3
В ₂	38,9	40,5	50,2
В ₃	34,9	30,8	32,3

Варіант 33.

Досліджується вплив на міцність чавуну двох факторів: А - вміст кремнію в чавуні, а саме А₁ - 0,24%; А₂ - 0,42%; А₃ - 0,52%; В - температурний режим плавлення (два режими) Результатом досліджень наведені в таблиці:

Фактор В	Фактор А		
	А ₁	А ₂	А ₃
В ₁	44.5	41.8	51.8
В ₂	33.4	38.9	30.6

Варіант 34.

Досліджується вплив на врожайність кукурудзи двох чинників: А - внесення добрив в ґрунт (три добрива); В - глибина поливу землі (три глибини поливу). Результати досліджень наведені в таблиці:

Фактор В	Фактор А		
	А ₁	А ₂	А ₃
В ₁	32,6	28,2	38,9
В ₂	44,2	36,4	46,8
В ₃	38,5	40,1	47,8

Варіант 35.

Досліджується вплив факторів А і В на число виготовлених втулок зі ста взятих, які відповідають нормам стандарту: А - використовували дві технології виготовлення; В - заготовки надходили з трьох заводів. Результати досліджень наведені в таблиці:

Фактор В	Фактор А	
	А ₁	А ₂
В ₁	90	96
В ₂	83	100
В ₃	75	86

Варіант 36 .

Досліджується вплив факторів А і В на продуктивність праці підприємства певної галузі промисловості: А - фондозабезпеченість (три рівня); В - рівень оплати праці працівникам (два рівня). Результати дослідження наведені в таблиці:

Фактор В	Фактор А		
	А ₁	А ₂	А ₃
В ₁	9,5	4,2	3,24
В ₂	7,8	13,6	9,7
В ₃	12,1	22,5	15,8

Варіант 37.

Експериментально досліджувався вплив на зносостійкість деталей факторів А і В: фактор А - тип сталі (три типи); фактор В - технологія виготовлення (дві технології).
Результати експерименту наведені в таблиці:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	10	12	14
B_2	8	11	13

Варіант 38.

Експериментально досліджувався вплив факторів А і В на опріснення морської води: фактор А - тип опріснювача (три типи); фактор В - три різних лабораторії.
Результати експерименту (відсоток солі в опрісненій воді) наведені в таблиці:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	3,6	2,92	3,25
B_2	4,25	3,8	3,64
B_3	3,46	3,8	4,12

Варіант 39.

Експериментально досліджувався вплив факторів А і В на врожайність цукрових буряків в ц / га: фактор А - добрива (три види); фактор - сума температур за період вегетації (три рівня). Результати досліджень наведені в таблиці:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	360	402	433
B_2	340	412	455
B_3	395	399	375

Варіант 40.

Досліджувався вплив факторів А і В на пружність стали в умовних одиницях: фактор А -% нікелю (три рівня); фактор В -% марганцю (два рівня) Результатом досліджень наведені в таблиці:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	36,4	41,2	36,5
B_2	39,2	39,7	40,8

Варіант 41.

Досліджувався вплив факторів А і В на річні надої молока в літрах: фактор А - три види соковитих кормів; фактор В - два види грубих кормів з різним рівнем частки,%, спеціальних вітамінів. Результати досліджень наведені в таблиці:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	5050	3900	12500
B_2	6900	8900	3200

Варіант 42.

Досліджується вплив на зносостійкість деталей наступних факторів: А - матеріал для виготовлення деталей (застосовували три види сталі) і В - технологія виготовлення деталей (дві технології) Результати експерименту наведені в таблиці:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	9	12	12
B_2	16	11	8

Варіант 43.

Досліджується вплив на міцність чавуну двох факторів: А - вміст кремнію в чавуні, а саме А1 - 0,24%; А2 - 0,42%; А3 - 0,52%; В - температурний режим плавлення (два режими) Результати досліджень наведені в таблиці:

Фактор В	Фактор А		
	А ₁	А ₂	А ₃
В ₁	44,9	40,8	54,8
В ₂	30,4	36,9	32,6

Варіант 44.

Досліджується вплив на врожайність кукурудзи двох чинників: А - внесення добрив в ґрунт (три добрива); В - глибина поливу землі (Три глибини поливу). Результати досліджень наведені в таблиці:

Фактор В	Фактор А		
	А ₁	А ₂	А ₃
В ₁	31,6	29,2	39,9
В ₂	42,2	38,4	49,8
В ₃	37,5	41,1	45,8

Варіант 45.

Досліджується вплив факторів А і В на число виготовлених втулок зі ста взятих, які відповідають нормам стандарту: А - використовували дві технології виготовлення; В - заготовки надходили з трьох заводів. Результати досліджень наведені в таблиці:

Фактор В	Фактор А	
	А ₁	А ₂
В ₁	96	99
В ₂	82	100
В ₃	76	89

Варіант 46.

Досліджується вплив на врожайність ячменю наступних факторів: А: посів здійснюється після чорного пару - А₁; після коренеплодів - А₂; після колосових культур - А₃; В - сортність ячменю (три сорти). Результати досліджень наведені в таблиці:

Фактор В	Фактор А		
	А ₁	А ₂	А ₃
В ₁	36,4	38,6	42,3
В ₂	40,2	40,5	51,2
В ₃	34,9	28,8	30,3

Варіант 47 .

Досліджується вплив факторів А і В на продуктивність праці підприємства певної галузі промисловості: А - фондозабезпеченість (три рівня); В - рівень оплати праці працівникам (два рівня) Результати дослідження наведені в таблиці:

Фактор В	Фактор А		
	А ₁	А ₂	А ₃
В ₁	9,7	4,6	5,24
В ₂	7,85	11,6	9,78
В ₃	10,1	24,5	16,8

Варіант 48.

Експериментально досліджувався вплив на зносостійкість деталей факторів А і В: фактор А - тип сталі (три типи); фактор В - технологія виготовлення (дві технології). Результати експерименту наведені в таблиці:

Фактор В	Фактор А		
	А ₁	А ₂	А ₃
В ₁	13	9	16
В ₂	6	12	14

Варіант 49.

Експериментально досліджувався вплив факторів А і В на опріснення морської води: фактор А - тип опріснювача (три типи); фактор В - три різних лабораторії. Результати експерименту (відсоток солі в опрісненій воді) наведені в таблиці:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	3,64	4,92	6,25
B_2	5,27	4,8	3,6
B_3	4,46	3,82	4,14

Варіант 50.

Експериментально досліджувався вплив факторів А і В на врожайність цукрових буряків в ц / га: фактор А - добрива (три види); фактор В - сума температур за період вегетації (три рівня). Результати досліджень наведені в таблиці:

Фактор В	Фактор А		
	A_1	A_2	A_3
B_1	350	422	435
B_2	340	417	465
B_3	397	367	383

Рівень значущості $\alpha = 0,025$										
$f_1 \backslash f_2$	1	2	3	4	5	6	7	8	9	
1	647,79	799,50	864,16	899,58	921,85	937,11	948,22	956,66	963,28	
2	38,51	39,00	39,16	39,25	39,30	39,33	39,35	39,37	39,39	
3	17,44	16,04	15,44	15,10	14,88	14,74	14,62	14,54	14,47	
4	12,22	10,65	9,98	9,60	9,36	9,20	9,07	8,98	8,90	
5	10,01	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68	
6	8,81	7,26	6,60	6,23	5,99	5,82	5,69	5,60	5,52	
7	8,70	6,54	5,89	5,52	5,28	5,12	4,99	4,90	4,82	
8	7,57	6,06	5,42	5,05	4,82	4,65	4,63	4,43	4,36	
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03	
10	6,94	5,46	4,82	4,47	4,24	4,07	3,95	3,85	3,78	
11	6,72	5,26	4,63	4,27	4,04	3,88	3,76	3,66	3,59	
12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,44	
13	6,41	4,96	4,35	3,99	3,77	3,60	3,48	3,39	3,31	
14	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,28	3,21	
15	6,20	4,76	4,15	3,80	3,58	3,41	3,29	3,20	3,12	
16	6,11	4,69	4,08	3,73	3,50	3,34	3,22	3,12	3,05	
17	6,04	4,62	4,01	3,66	3,44	3,28	3,16	3,06	2,98	
18	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,00	2,93	
19	5,92	4,51	3,90	3,56	3,33	3,17	3,05	2,96	2,88	
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84	
21	5,83	4,42	3,82	3,47	3,25	3,09	2,97	2,87	2,80	
22	5,79	4,38	3,78	3,44	3,21	3,05	2,93	2,84	2,76	
23	5,75	4,35	3,75	3,41	3,18	3,02	2,90	2,81	2,73	
24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,70	
25	5,69	4,29	3,69	3,35	3,13	2,97	2,85	2,75	2,68	
26	5,66	4,26	3,67	3,33	3,10	2,94	2,82	2,73	2,64	
27	5,63	4,24	3,65	3,31	3,08	2,92	2,80	2,71	2,63	
28	5,61	4,22	3,63	3,29	3,6	2,90	2,78	2,69	2,61	
29	5,59	4,20	3,61	3,27	3,04	2,88	2,76	2,67	2,59	
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57	
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45	
60	5,29	3,92	3,34	3,01	2,79	2,63	2,51	2,41	2,33	
120	5,15	3,80	3,23	2,89	2,67	2,51	2,39	2,30	2,22	
∞	5,02	3,69	3,12	2,78	2,57	2,41	2,29	2,19	2,11	
Рівень значущості $\alpha = 0,025$										
$f_1 \backslash f_2$	10	12	15	20	24	30	40	60	120	∞
1	968,63	976,71	984,87	993,10	997,25	1001,4	1005,6	1009,8	1014,0	1018,3
2	39,40	39,41	39,43	39,45	39,46	39,46	39,47	39,48	39,49	39,50
3	14,42	14,34	14,25	14,17	14,12	14,08	14,04	13,99	13,95	13,90
4	8,84	8,75	8,66	8,56	8,51	8,46	8,41	8,36	8,31	8,26
5	6,62	6,52	6,42	6,38	6,28	6,23	6,17	6,12	6,07	6,01
6	5,46	5,36	5,27	5,17	5,12	5,06	5,01	4,96	4,90	4,85
7	4,76	4,66	4,57	4,47	4,41	4,36	4,31	4,25	4,20	4,14
8	4,29	4,20	4,10	4,00	3,95	3,89	3,84	3,78	3,73	3,67
9	3,96	3,86	3,77	3,67	3,61	3,56	3,50	3,45	3,39	3,33
10	3,72	3,62	3,52	3,42	3,36	3,31	3,25	3,20	3,14	3,08
11	3,52	3,43	3,33	3,23	3,17	3,12	3,06	3,00	2,94	2,88
12	3,37	3,26	3,18	3,07	3,02	2,96	2,91	2,85	2,79	2,72
13	3,25	3,15	3,05	2,95	2,89	2,84	2,78	2,72	2,66	2,60
14	3,15	3,05	2,95	2,84	2,79	2,73	2,67	2,61	2,55	2,49
15	3,06	2,96	2,86	2,76	2,70	2,64	2,58	2,52	2,46	2,32
16	2,99	2,89	2,79	2,68	2,63	2,57	2,51	2,45	2,38	2,32
17	2,92	2,82	2,72	2,62	2,56	2,50	2,44	2,38	2,32	2,25

18	2,87	2,77	2,67	2,56	2,50	2,44	2,38	2,32	2,26	2,19
19	2,82	2,72	2,62	2,51	2,45	2,39	2,33	2,27	2,20	2,13
20	2,77	2,68	2,57	2,46	2,41	2,35	2,29	2,22	2,16	2,09
21	2,73	2,64	2,53	2,42	2,37	2,31	2,25	2,18	2,11	2,04
22	2,70	2,60	2,50	2,39	2,33	2,27	2,21	2,14	2,08	2,00
23	2,67	2,57	2,47	2,36	2,30	2,24	2,18	2,11	2,04	1,97
24	2,64	2,54	2,44	2,33	2,27	2,21	2,15	2,08	2,01	1,94
25	2,61	2,51	2,41	2,30	2,24	2,18	2,12	2,05	1,98	1,91
26	2,59	2,49	2,39	2,28	2,22	2,16	2,09	2,03	1,95	1,88
27	2,57	2,47	2,36	2,25	2,19	2,13	2,07	2,00	1,93	1,85
28	2,55	2,45	2,34	2,23	2,17	2,11	2,05	1,98	1,91	1,83
29	2,53	2,43	2,32	2,21	2,15	2,09	2,03	1,96	1,89	1,81
30	2,51	2,41	2,31	2,20	2,14	2,07	2,01	1,94	1,87	1,79
40	2,39	2,29	2,18	2,07	2,01	1,94	1,88	1,80	1,72	1,64
60	2,27	2,17	2,06	1,94	1,85	1,82	1,74	1,67	1,59	1,48
120	2,16	2,05	1,95	1,82	1,76	1,69	1,61	1,53	1,43	1,31
∞	2,05	1,94	1,83	1,71	1,64	1,67	1,48	1,39	1,27	1,00
Рівень значущості $\alpha = 0,05$										
$f_1 \backslash f_2$	1	2	3	4	5	6	7	8	9	
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,39	
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,95	2,80	
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	
19	4,38	3,52	3,13	2,89	2,74	2,63	2,54	2,42		
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	
29	4,18	3,33	2,93	2,70	2,54	2,43	2,35	2,28	2,22	
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	

Задача № 8

Регресійний аналіз. Парна лінійна регресія. Етап структуризації математичної моделі рівняння регресії

Короткий теоретичний матеріал

Два типу взаємозв'язку між змінними x і y

1. У першому випадку можна вказати, яка з змінних є аргументом, а яка функцією. Тоді відзначають рівноправність цих змінних і вказують статистичний взаємозв'язок між ними кореляційного типу.

2. У другому випадку має місце ситуація, коли змінні нерівноправні, і при цьому одна з них вважається такою, яка пояснюватиме (незалежна, аргумент), а інша, яка пояснюється (залежна, функція).

Це той випадок, коли зміна однієї змінної тягне за собою зміну іншої. Наприклад, зниження процентної ставки веде до зростання інвестицій, підвищення валютного курсу веде до зменшення чистого експорту, зростання доходу веде до збільшення попиту.

При цьому найбільш типовим є випадок (для вивчення взаємовпливу економічних змінних) функції:

$$y = f(x).$$

Зазвичай такі залежності записуються у вигляді рівнянь - **рівняння регресії**.

Рівняння регресії - формула статистичного зв'язку між випадковими змінними.

Якщо формула лінійна, то говорять про лінійну регресію, якщо нелінійна - про нелінійну регресію. Формула статистичного взаємозв'язку між двома змінними називаються **парної регресією**.

Якщо ж формула статистичного взаємозв'язку виражається більш ніж двома змінними, то мова йде про **множинної регресії** (лінійної, нелінійної).

При статистичному дослідженні відповідних змінних на першому етапі завжди здійснюється вибір форми взаємозв'язку між змінними. Такий підхід дозволяє реалізувати етап структуризації (параметризації) рівнянь регресії. Регресійний аналіз вирішує дві основні задачі.

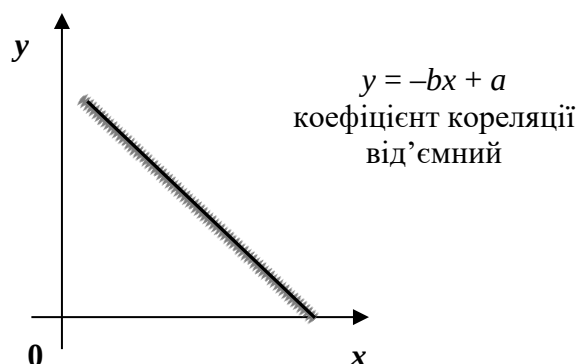
Задачі регресійного аналізу:

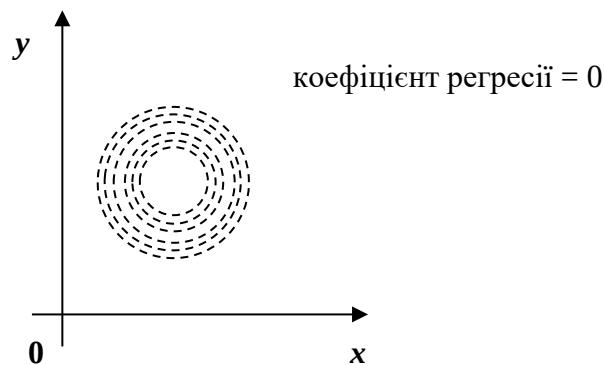
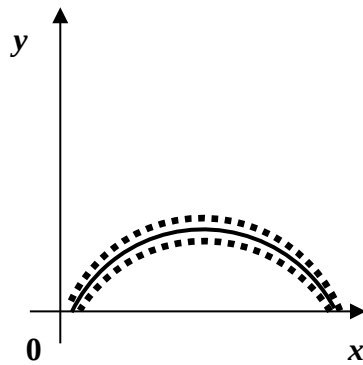
1. Структуризація (специфікація) математичної моделі, тобто, вибір і обґрунтування аналітичної залежності між змінними. На цьому етапі завжди отримують рівняння регресії. Слід зазначити, що поки рівняння регресії не буде відповідним чином оцінено, не буде проведено дослідження його адекватності, рівняння регресії буде являти собою всього лише гіпотезу.

2. Статистичний аналіз рівняння регресії. Проводиться з метою визначення адекватності рівняння регресії і встановлення значимого впливу незалежної змінної на залежну. Цей етап називається **етапом параметризації** рівняння регресії.

На першому етапі дослідження вирішують завдання специфікації рівняння регресії. Для парної регресії таке завдання вирішується на основі відповідного кореляційного поля.

Розглянемо варіанти кореляційного поля.



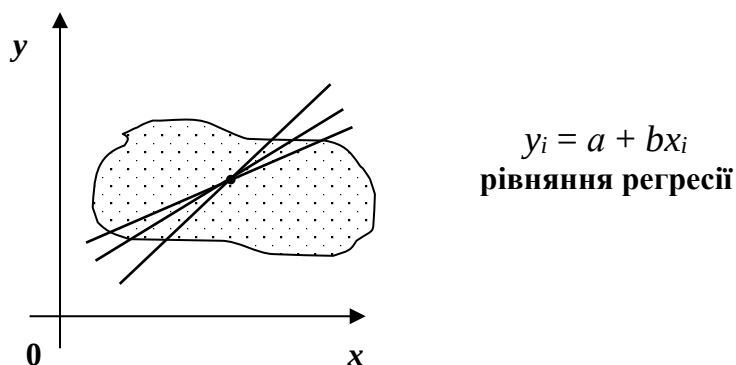


Метод найменших квадратів (МНК)

Вихідним пунктом аналізу між відповідними змінними зазвичай є оцінки лінійної залежності між змінними. Якщо є якась хмара точок, то через її можна провести таку пряму, яка найкращим чином описує функціональну залежність між змінними x і y .

У зв'язку з цим, вводять поняття щодо критерію близькості теоретичної лінії (рівняння регресії) і спостережуваними значеннями.

Як правило, такий критерій оцінює міру відстані від спостережуваних значень до лінії регресії. Зазвичай в якості такого критерію близькості використовують мінімум суми квадратів відхилень спостережуваних змінних y_i і теоретичних, обчислених за рівнянням регресії.



Таким чином, критерій може бути записаний так:

$$Q = \sum_{i=1}^m e_i^2 = \sum_{i=1}^m [y_i - (a + bx_i)]^2 \rightarrow \min$$

Тут x_i і y_i — відомі дані спостережень, а a і b — параметри рівняння регресії, що підлягають визначенню.

Зауважимо, що оскільки функція Q неперервна, опукла вниз і обмежена знизу нулями, то для неї буде існувати екстремум $\min$. Цілком очевидно, що для цих екстремальних точок значення параметрів a і b будуть найкращим чином описувати рівняння регресії.

У зв'язку з цим, завдання подальших досліджень полягає в тому, щоб виявити найбільш зручні та практичні формули для обчислення значень a і b .

Метод оцінювання параметрів лінійної регресії, що мінімізує суму квадратів відхилень спостережуваних значень від шуканої лінії називають методом найменших квадратів. За даним методом найкраща пряма завжди існує.

Оцінка коефіцієнтів парної лінійної регресії

Припускають, що зв'язок між x і y лінійна, тобто $y = \alpha + \beta x$. Тут мається на увазі те, що існує зв'язок для змінних генеральної сукупності.

Наявність існуючих збурень, врахованих і неврахованих факторів, а також з огляду на змінність статистичних спостережень параметрів на практиці, маємо:

$$y_i = \alpha + \beta x_i + e_i$$

У зв'язку з цим задача може бути сформульована таким чином: за наявними спостереженнями факторів $\{x_i\}$ і $\{y_i\}$ оцінити параметри α і β , забезпечуючи при цьому $\min$ величини критерію Q .

У той же час, якщо a є оцінкою параметра α , а b — параметра β , то оцінюване рівняння регресії буде перетворено до вигляду:

$$y_i = a + bx_i + e_i$$

$$m(e_i) = 0$$

При цьому розглянемо процедуру оцінювання a й b парної лінійної регресією. Цілком очевидно, що відповідно до МНК критерій близькості буде виражений:

$$Q = \sum_i e_i^2 = \sum_i [y_i - (a + bx_i)]^2 \rightarrow \min$$

Для даної функції двох незалежних змінних необхідна умова існування екстремуму, тобто:

$$\frac{\partial Q}{\partial a} = -2 \sum_i [y_i - (a + bx_i)] = 0$$

$$\frac{\partial Q}{\partial b} = -2 \sum_i [y_i - (a + bx_i)] \cdot x_i = 0$$

Тоді рівняння регресії матиме вигляд $y = ax + b$, а у дискретному вигляді $y_i = a + bx_i$.

У точці стаціонарності частинні похідні повинні дорівнювати нулю. Отриману систему представимо до більш зручній для аналізу формі:

$$\sum_{i=1}^n y_i - n \cdot a - b \sum_{i=1}^n x_i = 0 \tag{8.1}$$

$$\sum_{i=1}^n x_i y_i - a \sum_{i=1}^n x_i - b \sum_{i=1}^n x_i^2 = 0 \tag{8.2}$$

Аналіз (8.1) показує, що його можна представити у більш зручному вигляді:

$$\bar{y} = a + b\bar{x}.$$

Тут

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Таким чином, лінія регресії проходить через точку з середніми значеннями x і y ($\bar{x}$, $\bar{y}$).

Цілком очевидно, що з (8.1) мають місце співвідношення:

$$a = \bar{y} - b\bar{x} \quad (8.3)$$

— оцінка вільного члена лінії регресії

Для знаходження значення параметра b необхідно a зі співвідношення (8.3) підставити у вираз (8.2):

$$\sum_{i=1}^n x_i y_i - \sum_{i=1}^n (\bar{y} - b\bar{x}) x_i - b \sum_{i=1}^n x_i^2 = 0$$

Записуючи цей вираз у більш зручному для аналізу вигляді маємо:

$$\sum_{i=1}^n x_i y_i = n \cdot \bar{x} (\bar{y} - b\bar{x}) + b \sum_{i=1}^n x_i^2$$

Визначаючи з останнього співвідношення b , маємо:

$$b = \frac{\sum_{i=1}^n x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\text{cov}(x, y)}{\bar{D}(x)}$$

$$b = \frac{\overline{\text{cov}}(x, y)}{\bar{\sigma}_x \cdot \bar{\sigma}_x \cdot \bar{\sigma}_y} = \bar{r}_{xy} \cdot \frac{\bar{\sigma}_y}{\bar{\sigma}_x}$$

$$b = \bar{r}_{xy} \cdot \frac{\bar{\sigma}_y}{\bar{\sigma}_x} \quad (8.4)$$

Співвідношення (8.3) і (8.4) дозволяють отримати параметри a і b лінії регресії. При цьому, буде розв'язана перша задача дослідження, а саме задача специфікації рівняння регресії.

Аналітичні співвідношення завдання специфікації рівняння регресії:

1. Точкові характеристики вибірки

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$\bar{\sigma}_x^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \quad \bar{\sigma}_y^2 = \frac{1}{n} \sum_{i=1}^n y_i^2 - \bar{y}^2$$

2. Характеристики статистичного зв'язку між змінними

$$\overline{\text{cov}}(x, y) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \cdot \bar{y}$$

$$\bar{r}_{x,y} = \frac{\overline{\text{cov}}(x, y)}{\bar{\sigma}_x \cdot \bar{\sigma}_y}$$

3. Визначення параметрів рівняння регресії

$$b = \bar{r}_{x,y} \cdot \frac{\bar{\sigma}_y}{\bar{\sigma}_x}$$

$$a = \bar{y} + b\bar{x}$$

ЗАВДАННЯ:

1. За вибірковими даними побудувати кореляційне поле.
2. За даними кореляційного поля запропонувати гіпотезу про вигляд математичної моделі лінії регресії.
3. Знайти математичну модель лінійної регресії. Для чого:
 - 3.1. Обчислити оцінки математичного очікування випадкових величин або по залежностям

$$\bar{x} = \frac{1}{n} \sum_i x_i, \quad \bar{y} = \frac{1}{n} \sum_i y_i,$$

або використовуючи стандартні функції середовища MathCad.

- 3.2. Обчислити оцінки дисперсій випадкових величин або по залежностям:

$$\sigma_x^2 = \frac{1}{n} \sum_i x_i^2 - \bar{x}^2, \quad \sigma_y^2 = \frac{1}{n} \sum_i y_i^2 - \bar{y}^2,$$

або використовуючи стандартні функції середовища MathCad.

- 3.3. Обчислити кореляційний момент по залежності

$$\text{cov}(x, y) = \frac{1}{n} \sum_i x_i \cdot y_i - \bar{x} \cdot \bar{y}.$$

або використовуючи стандартні функції середовища MathCad.

- 3.4. Обчислити коефіцієнт кореляції по залежності

$$r_{x,y} = \frac{\text{cov}(x, y)}{\sigma_x \cdot \sigma_y},$$

або використовуючи стандартні функції середовища MathCad.

3.5. Обчислити коефіцієнт b лінійної регресії, використовуючи формулу

$$b = r_{x,y} \frac{\sigma_y}{\sigma_x}.$$

3.6. Обчислити коефіцієнт a лінійної регресії, використовуючи формулу

$$a = \bar{y} - b \cdot \bar{x}.$$

3.7. Лінію регресії отриманого рівняння $y = a + bx$ сумістити з кореляційним полем.



1. За вибірковими даними побудувати кореляційне поле.

➤ У текстовий редактор «Блокнот» ввести два файли, один – дані по змінній X , інший – дані по змінній Y .

➤ У середовищі MathCad ввести інформацію, необхідну до експорту даних з блокнота

Як це буде виглядати у середовищі MathCad

ORIGIN := 1

N := 15

i := 1.. N

$x_i := \text{READ}("f:\text{dan}\backslash x1.txt")$

$y_i := \text{READ}("f:\text{dan}\backslash y5.txt")$

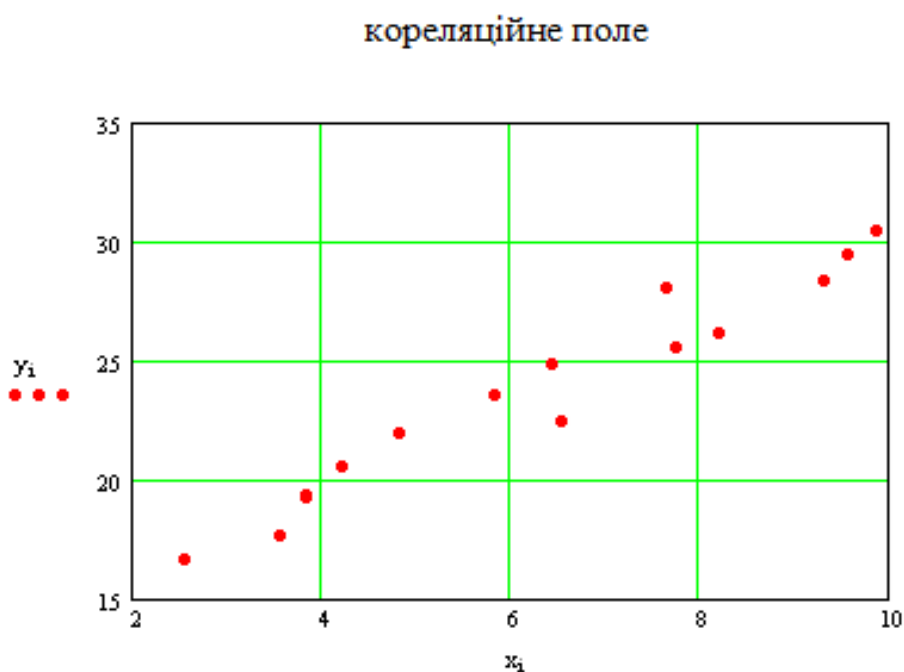
Як це зробити у середовищі MathCad

➤ Увійти у графічний редактор середовища.

➤ По осі абсцис вказати - x_i .

➤ По осі ординат вказати - y_i

Як це буде виглядати у середовищі MathCad



2. За даними кореляційного поля запропонувати гіпотезу про вигляд математичної моделі лінії регресії. При необхідності сформулювати наступний **висновок**: по виду кореляційного поля висувається гіпотеза про лінійний зв'язок між спостережуваними змінними.

3.1 . Обчислити оцінки математичного очікування випадкових величин:

по залежностям $\bar{x} = \frac{1}{n} \sum_i x_i$, $\bar{y} = \frac{1}{n} \sum_i y_i$, або використовуючи стандартні функції середовища MathCad.

Як це зробити у середовищі MathCad

Оцінка математичного очікування спостережуваних змінних

$$xsr := \frac{1}{N} \cdot \sum_{i=1}^N x_i \quad xsr = 6.257$$

$$ysr := \frac{1}{N} \cdot \sum_{i=1}^N y_i \quad ysr = 23.609$$

3.2. Обчислити оцінки дисперсій випадкових величин по залежностям

$$\sigma_x^2 = \frac{1}{n} \sum_i x_i^2 - \overline{x^2}, \quad \sigma_y^2 = \frac{1}{n} \sum_i y_i^2 - \overline{y^2},$$

або використовуючи стандартні функції середовища MathCad .

Як це зробити у середовищі MathCad

Оцінка дисперсії спостережуваних змінних

$$kvX := \frac{1}{N} \cdot \sum_{i=1}^N (x_i)^2 \quad kvY := \frac{1}{N} \cdot \sum_{i=1}^N (y_i)^2$$

$$kvX = 44.524 \quad kvY = 575.41$$

$$dispX := kvX - xsr^2 \quad dispY := kvY - ysr^2$$

$$dispX = 5.37 \quad dispY = 18.01$$

3.3. Обчислити кореляційний момент по залежності $cov(x, y) = \frac{1}{n} \sum_i x_i \cdot y_i - \overline{x \cdot y}$.

або використовуючи стандартні функції середовища MathCad.

Як це зробити у середовищі MathCad

Оцінка кореляційного моменту і коефіцієнта кореляції

$$\text{prxy} := \left(\sum_{i=1}^N x_i \cdot y_i \right) \quad \text{prxy} = 2.36 \times 10^3$$

$$\text{covxy} := \frac{1}{N} \cdot \text{prxy} - \text{xsr} \cdot \text{ysr}$$

$$\text{covxy} = 9.605$$

3.4. Обчислити коефіцієнт кореляції по залежності $r_{x,y} = \frac{\text{cov}(x, y)}{\sigma_x \cdot \sigma_y}$,
або використовуючи стандартні функції середовища MathCad.

Як це зробити у середовищі MathCad

$$\text{kkor} := \frac{\text{covxy}}{\sqrt{\text{dispX}} \cdot \sqrt{\text{dispY}}}$$

$$\text{kkor} = 0.977$$

3.5. Обчислити коефіцієнт b лінійної регресії, використовуючи формулу $b = r_{x,y} \frac{\sigma_y}{\sigma_x}$.

Як це зробити у середовищі MathCad

Обчислення оцінок коефіцієнтів рівняння регресії

$$b := \text{kkor} \cdot \frac{\sqrt{\text{dispY}}}{\sqrt{\text{dispX}}} \quad b = 1.788$$

3.6. Обчислити коефіцієнт a лінійної регресії, використовуючи формулу $a = \bar{y} - b \cdot \bar{x}$.

Як це зробити у середовищі MathCad

$$a := y_{sr} - b \cdot x_{sr} \qquad a = 12.418$$

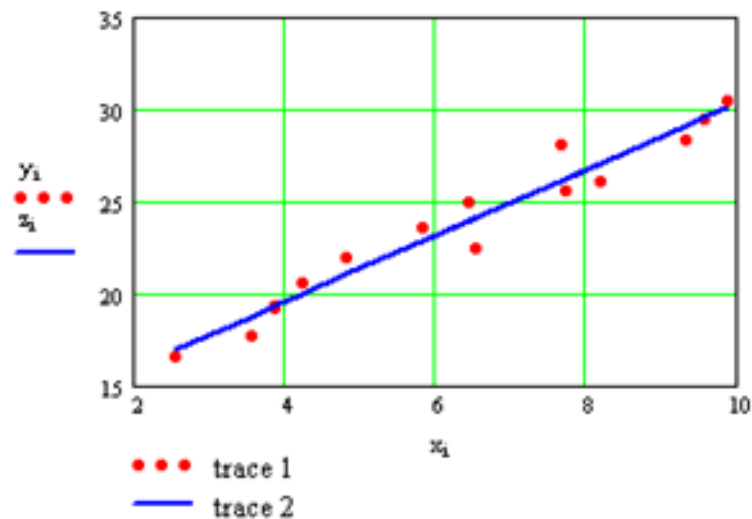
3.8. Лінію регресії отриманого рівняння $y = a + bx$ сумістити з кореляційним полем.

Як це зробити у середовищі MathCad

Лінія регресії

$$y1(x) := a + b \cdot x$$

$$z_1 := a + b \cdot x_1$$



Висновок : при зміні аргументу на одиницю значення функції збільшиться на 1.788.

Варіанти завдань

Вибір завдання

- Якщо Ваш варіант – число 17, то вибираються стовпці X_1 и Y_7
- Якщо Ваш варіант – число 20, то вибираються стовпці X_2 и Y_0
- Якщо Ваш варіант – число 5 (або 05), то вибираються стовпці X_0 и Y_5

ЗНАЧЕННЯ ФАКТОРА X

X ₀	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉
2,06	2,53	2,17	3,65	3,22	2,16	4,57	2,25	6,15	1,86
2,58	3,54	2,90	3,82	3,87	2,65	5,42	2,98	5,66	1,91
3,14	3,84	3,29	3,76	4,95	3,49	5,29	2,15	7,50	2,14
3,54	3,84	4,13	5,24	5,10	3,16	6,33	2,71	6,90	3,39
4,18	4,22	5,25	5,03	5,98	3,85	7,63	3,70	8,31	3,95
4,78	4,81	4,92	5,52	7,28	4,58	7,53	4,59	8,25	4,30
5,11	6,53	5,79	5,62	6,90	5,33	7,73	4,77	9,39	5,10
5,67	5,82	5,87	6,98	7,54	5,89	8,44	5,34	9,73	5,47
6,02	6,43	6,99	6,91	7,91	6,20	9,49	5,45	9,33	5,97
6,65	7,73	7,04	7,95	8,40	6,39	9,18	6,00	10,50	6,16
7,05	8,19	8,14	7,24	8,14	6,95	10,14	6,25	11,10	6,46
7,52	7,65	8,06	9,27	8,76	7,25	9,94	6,79	11,51	6,07
8,03	9,31	8,57	8,46	9,67	7,80	10,92	8,24	12,42	6,71
8,56	9,26	9,45	10,30	10,28	8,47	11,89	8,51	12,40	7,16
9,03	9,86	9,06	10,72	10,59	9,22	11,14	9,15	13,14	8,81
<i>Значення X для прогнозу (X_p)</i>									
9,52	9,69	10,30	10,05	11,58	9,32	11,73	9,78	12,56	8,07

ЗНАЧЕННЯ ПОКАЗНИКА Y

Y ₀	Y ₁	Y ₂	Y ₃	Y ₄	Y ₅	Y ₆	Y ₇	Y ₈	Y ₉
7,24	10,89	16,21	12,11	15,21	16,62	10,22	12,50	19,66	14,87
8,02	11,92	17,75	12,30	15,42	17,63	10,58	13,88	20,53	15,78
9,28	12,45	16,39	13,82	16,44	19,22	12,01	15,16	21,31	16,79
10,12	13,27	18,87	14,84	17,93	19,36	12,84	16,06	22,59	18,03
11,12	14,12	19,60	15,86	18,52	20,52	13,28	16,66	23,27	18,29
12,19	15,23	21,21	16,41	19,80	21,95	15,13	17,65	24,44	19,93
13,01	16,07	21,84	17,80	20,76	22,45	15,84	18,46	25,85	20,32
14,12	17,40	23,00	18,61	21,30	23,56	17,08	19,54	26,74	21,18
15,21	18,68	24,44	19,57	22,25	24,90	17,99	20,58	27,36	22,47
16,29	19,46	25,36	21,26	24,14	25,53	18,32	21,77	28,37	23,47
17,01	20,52	25,54	21,08	24,17	26,11	19,49	22,15	29,22	24,07
18,03	21,32	27,14	22,99	25,66	28,02	20,59	23,80	30,50	25,57
19,19	22,58	27,95	23,43	26,50	28,37	21,35	24,79	31,21	27,07
20,21	23,73	28,99	24,63	27,46	29,48	23,20	25,57	32,56	27,62
21,22	25,02	30,80	25,41	29,02	30,42	24,21	27,18	33,66	28,42

Задача № 9

Регресійний аналіз. Парна лінійна регресія. Етап параметризації математичної моделі рівняння регресії

Короткий теоретичний матеріал

Оцінка адекватності математичної моделі

Необхідно ще раз підкреслити, що поки кількісно НЕ будуть оцінені параметри рівняння регресії і не буде виконана процедура перевірки зроблених оцінок, форма лінії регресії залишається **гіпотезою**.

Оцінка значень параметрів обраної форми рівняння регресії називається **процесом параметризації рівняння регресії**.

Однак слід мати на увазі, що на першому етапі виконують аналіз отриманого рівняння регресії. Розглянемо приклад.

Приклад 1.

Досліджуючи функцію попиту на телятину в одній з європейських країн, отримано наступне рівняння регресії: $y = 1,2 - 1,3 \cdot x$.

Отримане рівняння попиту на телятину як функція ціни телятини показує, якщо припустити, що за інших рівних умов ціна на телятину зростає на одиницю, то попит на неї впаде на 1,3 одиниць.

Дослідження коефіцієнта кореляції

Коефіцієнт кореляції використовується у якості заходів для ступеня лінійного зв'язку між двома змінними. Близька до 0 величина коефіцієнта кореляції говорить про відсутність *лінійного зв'язку* між змінними. Зауважимо, що тут мова йде не взагалі про відсутність зв'язку між змінними, а лише про відсутність лінійного зв'язку. Тоді, чим ближче *значення коефіцієнта кореляції до одиниці*, тим *тісніше буде лінійний зв'язок між змінними x і y* .

В аналізі коефіцієнта кореляції виникає наступне питання: якщо він дорівнює нулю для генеральної сукупності, то це не означає, що він буде дорівнювати нулю і для вибіркової сукупності (і навпаки). Тоді, чим більше відхилення такого параметра, тим менше воно ймовірно для відповідної вибірки. **Зауважимо, що вибірковий коефіцієнт кореляції є величиною випадковою**, тому необхідно використовувати відповідний критерій для аналізу таких оцінок. Для аналізу вибіркового коефіцієнта кореляції (статистичний аналіз) використовують ***t*-статистику Стьюдента**.

Основні ідеї *t*-статистики Стьюдента

Для аналізу вибіркового коефіцієнта кореляції на **I етапі** розраховують *t*-статистику Стьюдента за такою формулою:

$$t_{\text{набл}} = \bar{r}_{x,y} \frac{\sqrt{n-2}}{1 - \bar{r}_{x,y}^2}.$$

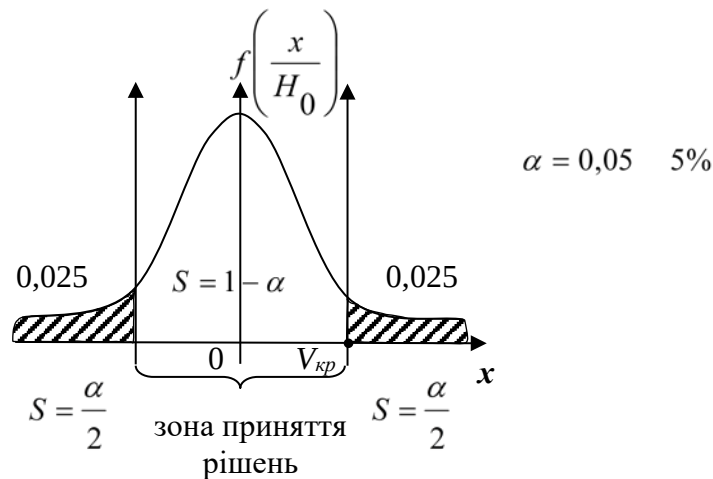
Тут *t*-статистика має розподіл Стьюдента з $(n - 2)$ ступенями свободи.

Для коефіцієнта кореляції (r) висувається нульова гіпотеза H_0 , яка полягає в тому, що передбачає рівність нулю коефіцієнта кореляції. Ця гіпотеза відхиляється, якщо вибіркове значення коефіцієнта кореляції r далеко від нульового значення, тоді приймається гіпотеза H_1 .

Суть гіпотези H_1 полягає в тому, що у генеральній сукупності $r \neq 0$ і статистично значимий. Це означає, що між змінними x і y існує лінійна залежність. Отже, за *t*-статистикою Стьюдента визначається критичне значення критерію ($V_{кр.}$), що знаходиться у залежності від прийнятого рівня значущості.

Рівень значущості α — ймовірність, з якою перевіряється статистична гіпотеза. При цьому
1% — 0,99 5% — 0,95

Розглянемо відповідну геометричну інтерпретацію.



 — зона віхилення нульової гіпотези.

Якщо обчислене значення критерію (що спостерігається) потрапляє у зону прийняття критерію, то роблять висновок про рівність коефіцієнта кореляції нулю ($r = 0$) у генеральній сукупності.

Чим далі обчислене значення критерію (що спостерігається) ухиляється від $V_{кр}$, тем тим більше підстави відкидати H_0 (нульову гіпотезу), а приймати гіпотезу H_1 щодо статистичної значущості коефіцієнта кореляції r .

Розглянемо процедуру статистичної значущості r на прикладі.

Приклад 2. Взято десять спостережень показників інфляції та безробіття у США у 1931-1940 рр. Для них розраховані вибіркові значення r ($r = -0,227$).

Очевидно, що такий зв'язок від'ємна, що відповідає економічній теорії (**крива Філіпса**). Однак чи є цей параметр (r) статистично значимий?

На I етапі висувається гіпотеза H_0 про рівність $r = 0$ і обчислюють t -статистику Стюдента з $(n - 2)$ ступенями свободи.

$$t_{набл} = \bar{r} \frac{\sqrt{n-2}}{\sqrt{1-\bar{r}^2}} = -0,227 \cdot \frac{\sqrt{10-2}}{\sqrt{1-(-0,227)^2}} \approx -0,66$$

Далі задають рівень значущості $\alpha = 5\%$ ($\alpha = 0,05$) і з ймовірністю 0,95 перевіряють статистичну гіпотезу. Цілком очевидно, що відповідний критерій буде складатися з двох хвостів, площа кожного дорівнює 0,025.

Тоді, за таблицею критичних точок розподілу Стюдента відшукують $t_{кр}$ (критичне).

$$t_{кр} = 2,314; \quad t_{набл.} = -0,66; \quad |t_{набл.}| = 0,66.$$

Отже, для виконаного дослідження отримано: $|t_{набл.}| < t_{кр}$.

У зв'язку з цим роблять висновок - гіпотеза H_0 приймається (у генеральній сукупності $r = 0$) - це означає, що між інфляцією та безробіттям немає лінійного зв'язку. Але це не означає, що між цими параметрами зв'язок взагалі відсутній. Тут, швидше за все, необхідно говорити щодо нелінійного зв'язку з одного боку, або щодо більш складних зв'язках рівня інфляції і безробіття з іншими параметрами.

Перевірка загальної якості рівняння регресії. Коефіцієнт детермінації

Для аналізу загальної якості рівняння регресії зазвичай використовують коефіцієнт детермінації R^2 .

R^2 — коефіцієнт детермінації (квадрат коефіцієнта множинної кореляції). Для окремого випадку парної регресії R^2 — це квадрат коефіцієнта вибіркової кореляції.

Суть R^2 :

- R^2 характеризує частку варіації (розкиду) залежної змінної щодо значення аргументу.
- іншими словами, R^2 показує: яка частина коливання у обумовлено колеблемостью x , (а яка частина коливання у пояснюється іншими функціями).

Очевидно, чим ближче R^2 до одиниці, тим більш адекватна математична модель (коли $R^2 = 1$, то між x і y суворий лінійний зв'язок).

Приклад 3. При деяких дослідженнях отримано значення коефіцієнта кореляції ($r = 0,9$). Проаналізувати коефіцієнт детермінації.

Тоді $R^2 = 0,81$. Це означає, що 81% загального розсіювання змінної у пояснюється лінійним зв'язком з x , і тільки 19% розсіювання у пояснюються іншими чинниками, що діють на вибірку вибірково.

Висновок: R^2 є мірою визначно лінійної регресії і тоді, чим більше R^2 , тим менше спостережувані значення у ухиляються від обчисленої лінії регресії.

Як для коефіцієнта кореляції (r), так і для коефіцієнта детермінації (R^2) необхідно проводити дослідження його статистичної значущості. При цьому тут використовують **F-статистика Фішера**, яка розраховується за співвідношенням:

$$F = \frac{R^2}{1 - R^2} \cdot (n - 2)$$

Суть F-статистики Фішера:

H_0 говорить про рівність нулю всіх коефіцієнтів регресії, за винятком вільного члена ($r = 0$). Це означає, що для генеральної сукупності висувають значення $y = \bar{y}$, тобто при цьому R^2 і F -статистика Фішера обертаються до 0 ($R^2 = 0$, $F = 0$).

Альтернативна гіпотеза H_1 говорить про те, що не всі коефіцієнти рівняння регресії обертаються до нуля. Отже, $R^2 \neq 0$, і він пояснює міру розсіювання змінної у відносно змінної x .

Приклад 4. Нехай за спостережуваним значенням x і y обчислене значення $R^2 = 0,7$. При цьому у розрахунку використовують 15 спостережень ($n = 15$). Виконати дослідження статистичної значимості коефіцієнта детермінації.

$$F_{\text{набл.}} = \frac{0,7}{1 - 0,7} (15 - 2) \cong 30,3$$

По таблиці розподілу Фішера з (1,13) ступенями свободи і при рівні значущості 5% знаходять критичні значення критерію Фішера: $F_{\text{кр.}} = 4,67$

Тоді, $F_{\text{набл.}} > F_{\text{кр.}}$, і H_0 відкидається. Це означає, що значення R^2 статистично значимо і з ймовірністю 0,95 робиться висновок про адекватність прийнятої математичної моделі.

Побудова довірчих інтервалів для коефіцієнтів регресії

I. Стандартне відхилення кутового коефіцієнта і його довірчий інтервал

Рівняння регресії має вигляд: $y = a + bx$. Тоді стандартне (середньоквадратичне) відхилення параметра b може бути обчислено, беручи до уваги оцінку дисперсії. Цілком очевидно, що незміщеною і обґрунтованою статистичною оцінкою $\bar{\sigma}^2$ буде співвідношення:

$$S^2 = \frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{n - 2}$$

Якщо припустити, що розкид спостережуваних значень щодо теоретичної лінії розподілений за нормальним законом, то довірчий інтервал для параметра b буде визначатись за формулою: $b \pm \Delta b$, де

$$\Delta b = t_{\alpha,k} \cdot \frac{S}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

$t_{\alpha,k}$ — значення функції Стюдента, яке визначається для заданого рівня значущості (α) і числом ступенів свободи (k).

II. Стандартне відхилення для вільного члена і його довірчий інтервал

Вільний член рівняння регресії визначається за співвідношенням:

$$a = \bar{y} - b\bar{x}.$$

Беручи до уваги ту обставину, що відхилення спостережуваних значень від лінії регресії розподілено за нормальним законом, то можна отримати:

$$\Delta a = t_{\alpha,k} \cdot S \cdot \frac{\sqrt{\sum_{i=1}^n x_i^2}}{\sqrt{n \cdot \sum_{i=1}^n (x_i - \bar{x})^2}} = \Delta b \cdot \frac{\sqrt{\sum_{i=1}^n x_i^2}}{n}.$$

Довірчий інтервал для параметра a будемо визначати за формулою: $a \pm \Delta a$.

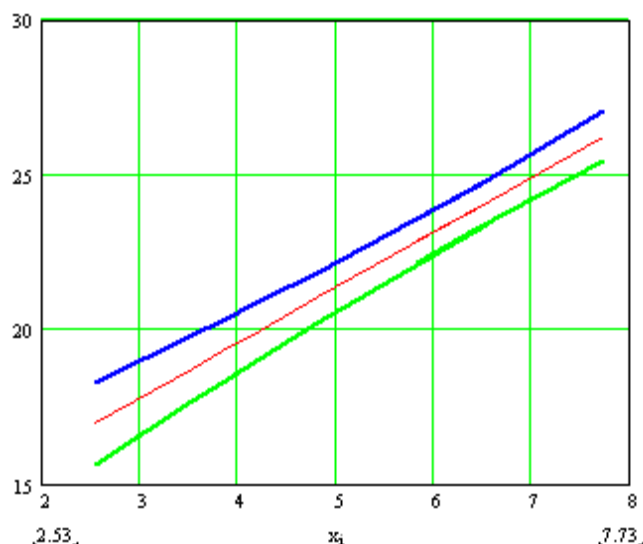
Довірча зона регресії

Коефіцієнти рівняння регресії визначаються на підставі аналізу вибірових даних. Отже, вони носитиме випадковий характер, і від вибірки до вибірки матимуть різні значення. Тоді виникає питання: яким чином у генеральній сукупності буде розташовуватися лінія регресії? У цьому випадку визначають довірчу зону регресії, що визначає місце положення лінії регресії у генеральній сукупності.

Аналогічно попереднім міркуванням, мають співвідношення виду:

$$\Delta y_i = t_{\alpha,k} \cdot S[\bar{y}_i] = t_{\alpha,k} \cdot \frac{S}{\sqrt{n}} \cdot \sqrt{1 + \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\bar{D}_b(x)}}$$

В такому випадку, поєднуючи безперервною лінією всі значення $y_i + \Delta y_i$ і $y_i - \Delta y_i$, отримують довірчу зону регресії для базисних значень спостереження.



Прогноз і його довірчий інтервал

Прогноз - наукове передбачення ймовірнісних шляхів розвитку деяких процесів у більш-менш віддаленому майбутньому.

Період випередження - проміжок часу від моменту, для якого є останні статистичні дані до моменту, якому належить прогноз.

Прогнозоване значення змінної визначається на підставі співвідношення: $y_p = a + bx_p$

Для прогнозованого значення довірчий інтервал визначається наступним чином:

$$y_p \pm \Delta y_p, \text{ где}$$

$$\Delta y_p = t_{\alpha, k} \cdot S \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n \cdot \bar{D}_b(x)}}$$

ЗАВДАННЯ:

1. Дослідити коефіцієнт кореляції, використовуючи t-статистику Стюдента.

При цьому

- обчислити $t = r \cdot \frac{\sqrt{n-2}}{\sqrt{1-r^2}},$

- визначити за таблицями розподілення критичних точок Стюдента з ймовірністю 0.95, $t_{кр}$.
- порівнюючи спостережуване і критичне значення t - статистики Стюдента, сформулювати висновок про статистичну значущість коефіцієнта кореляції.

2. Використовуючи критерій Фішера, оцінити адекватність прийнятої економетричної моделі. При цьому

- обчислити $F = \frac{R^2}{1-R^2} \cdot (n-2),$

- визначити за таблицями розподілення критичних точок Фішера $F_{кр}$ з ймовірністю 0.95,
- прийняти $R^2 = r_{xy}^2$.
- порівнюючи спостережуване і критичне значення f - статистики Фішера, сформулювати висновок про статистичну значущість коефіцієнта детермінації.

3. Провести аналіз R^2 .

4. Побудувати довірчий інтервал для коефіцієнта при невідомій рівняння регресії. Для цього

$$\text{провести обчислення } b \pm \Delta b, \quad \Delta b = t_{\alpha,k} \cdot \frac{s}{\sqrt{n} \cdot \sigma_x}.$$

Коефіцієнт S обчислити за формулою $S^2 = \frac{\sum_i (y_i - \bar{y})^2}{n-2}$, склавши відповідну для обчислень таблицю.

5. Визначити довірчу зону регресії і за отриманими даними побудувати відповідні залежності,

$$\text{при цьому } y_i \pm \Delta y_i, \quad \Delta y_i = t_{\alpha,k} \cdot \frac{S}{\sqrt{n}} \cdot \sqrt{1 + \frac{(x_i - \bar{x})^2}{D_x}}.$$

Для отримання сукупності значень Δy_i скласти відповідні таблиці.

6. Визначити прогноз відповідного фактора. Знайти довірчий інтервал для прогнозу

$$y_p = a + bx_p, \quad \Delta y_{p_i} = t_{\alpha,k} \cdot S \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n \cdot D_x}}.$$

7. Визначити значення коефіцієнта еластичності для прогнозованого значення.

Як це зробити у середовищі MathCad

1. Дослідити коефіцієнт кореляції, використовуючи t- статистику Стюдента.

При цьому

$$\text{- обчислити } t = r \cdot \frac{\sqrt{n-2}}{\sqrt{1-r^2}},$$

- визначити за таблицями розподілення критичних точок Стюдента з ймовірністю 0.95, $t_{кр}$.

- порівнюючи спостережуване і критичне значення t - статистики Стюдента, сформулювати висновок про статистичну значущість коефіцієнта кореляції.

Якщо $|t_{\text{набл}}| > t_{\text{кр}}$, то коефіцієнт кореляції r статистично значимий і між спостережуваними змінними X і Y існує лінійний зв'язок.

Як це зробити у середовищу MathCad

1. Дослідження коефіцієнта кореляції

$$tr := kkor \cdot \frac{\sqrt{N-2}}{\sqrt{1-kkor^2}}$$

$$tr = 12.849$$

З таблиці розподілу Стюдента з рівнем значущості $\alpha = 0.05$ і числом ступенів свободи $N - 2$ знаходять критичне значення критерію

$$tkr := 2.31$$

Оскільки $|tr| > tkr$, то коефіцієнт кореляції статистично значущий.

2. Використовуючи критерій Фішера, оцінити адекватність прийнятої економічної моделі. При цьому

- обчислити $F = \frac{R^2}{1-R^2} \cdot (n-2)$,
- визначити по таблиці розподілення критичних точок Фішера з ймовірністю 0.95, $F_{кр}$.
- прийняти $R^2 = r_{xy}^2$.
- порівняти спостережуване і критичне значення F -статистики Фішера, зробити висновок про статистичну значущість коефіцієнта детермінації і про адекватність математичної моделі рівняння регресії у цілому (*чим більше $F_{\text{набл}}$ у порівнянні з $F_{кр}$, тим більш адекватною є математична модель*).

Як це зробити у середовищі MathCad

Дослідження коефіцієнта детермінації

$$Rkv := kkor^2$$

$$Rkv = 0.954$$

Аналіз коефіцієнта детермінації показує, що 95.4% колеблiмостi у пояснюється колеблiмостю x і тільки 4.6% колеблiмостi у пояснюється іншими факторами, діючими вибірково.

3. Провести аналіз R^2 .

Як це зробити у середовищі MathCad

Аналіз статистичної значущості коефіцієнта детермінації

$$k_{det} := R_{kv}$$

виконується по F-статистикою Фішера

$$Fr := \frac{k_{det}}{1 - k_{det}^2} \cdot (N - 2)$$

$$Fr = 84.504$$

З таблиці розподілу Стюдента з рівнем значущості $\alpha = 0.05$ і $(1, N - 2)$ числом ступенів свободи знаходять критичне значення критерію:

$$F_{kr} := 7.67$$

Оскільки $Fr > F_{kr}$, то коефіцієнт детермінації статистично значущий. Математична модель адекватна.

4. Побудувати довірчий інтервал для коефіцієнта при невідомому рівнянні регресії. Для цього

- провести обчислення $b \pm \Delta b$, $\Delta b = t_{\alpha, k} \cdot \frac{s}{\sqrt{n} \cdot \sigma_x}$.

- коефіцієнт S обчислити за формулою $S^2 = \frac{\sum_i (y_i - \hat{y}_i)^2}{n - 2}$, склавши відповідну для обчислень таблицю.

Як це зробити у середовищі MathCad

Будуємо довірчий інтервал для коефіцієнта при невідомому рівнянні регресії.

$$b = 1.788$$

$$a = 12.418$$

$$S_{kv} := \frac{1}{N - 2} \cdot \sum_{i=1}^{10} (y_i - a - b \cdot x_i)^2$$

$$ysr = 23.609$$

$$xsr = 6.257$$

$$S_{kv} = 0.908$$

$$\text{disp}X = 5.37$$

$$Dx := \text{disp}X$$

$$\text{det}B := 2.31 \cdot \frac{\sqrt{Skv}}{\sqrt{N} \cdot \sqrt{Dx}}$$

$$\text{det}B = 0.457$$

$$b1 := b + \text{det}B$$

$$b1 = 2.246$$

$$b2 := b - \text{det}B$$

$$b2 = 1.331$$

5. Визначити довірчу зону регресії і за отриманими даними побудувати відповідні залежності,

$$\text{при цьому } y_i \pm \Delta y_i, \quad \Delta y_i = t_{\alpha,k} \cdot \frac{S}{\sqrt{n}} \cdot \sqrt{1 + \frac{(x_i - \bar{x})^2}{D_x}}.$$

Для отримання сукупності значень Δy_i скласти відповідні таблиці.

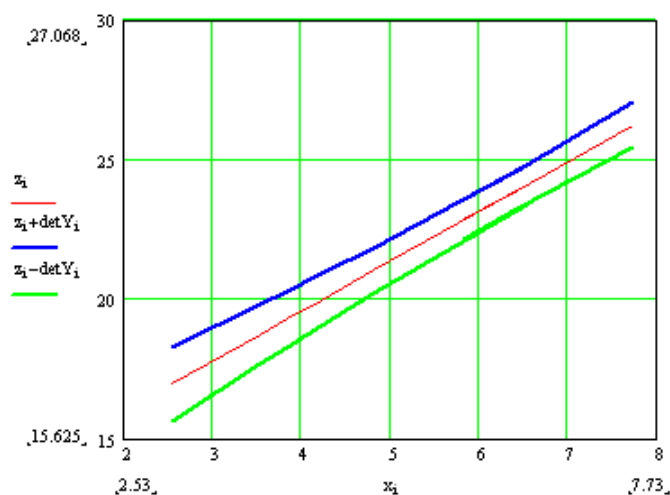
Знайдені коефіцієнти рівняння регресії a , b залежать від параметрів вибірки, і від вибірки до вибірки будуть змінюватися. Тому рівняння регресії носять наближений характер.

У зв'язку з цим необхідно побудувати довірчу зону регресії, у якій з довірчою ймовірністю α буде перебувати справжня лінія регресії (для будь-якої вибірки з генеральної сукупності).

Як це зробити у середовищі MathCad

Довірча зона регресії

$$\text{det}Y_i := 2.31 \cdot \sqrt{\frac{Skv}{N}} \cdot \sqrt{1 + \frac{(x_i - x_{sr})^2}{D_x}}$$



6. Визначити прогноз відповідного фактора. Знайти довірчий інтервал для прогнозу

$$y_p = a + bx_p, \quad \Delta y_{p_i} = t_{\alpha, k} \cdot S \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n \cdot D_x}}.$$

Отримане рівняння регресії застосовується для обчислення прогнозного значення фактора Y .
Задаючи значення фактора X_p , по рівнянню регресії визначається Y_p .

Так як коефіцієнти рівняння регресії носять наближений характер, Y_p також буде величиною наближеною. У зв'язку з цим необхідно побудувати довірчий інтервал, якому з довірчою ймовірністю буде належати істинне значення прогнозованої величини Y_p .

Розмах довірчого інтервалу визначається за наведеною формулою.

Як це зробити у середовищі MathCad

Визначення прогнозу

$$X_p := 9.69$$

$$Y_p := a + b \cdot X_p \quad Y_p = 29.749$$

СТАТИСТИЧНІ ТАБЛИЦІ

КРИТИЧНІ ТОЧКИ РОЗПОДІЛУ СТЬЮДЕНТА Таблиця 9.1

Число ступенів свободи k	Рівень значущості α (двостороння критична область)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,31	12,7	31,82	63,7	318,3	637,0
2	2,92	4,30	6,97	9,92	22,33	31,6
3	2,35	3,18	4,54	5,84	10,22	12,9
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,86
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,40
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,90	3,65	3,96
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,73	2,09	2,53	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,51	3,79
23	1,71	2,07	2,50	2,81	3,49	3,77
24	1,71	2,06	2,49	2,80	3,47	3,74
25	1,71	2,06	2,49	2,79	3,45	3,72
26	1,71	2,06	2,48	2,78	3,44	3,71
27	1,71	2,05	2,47	2,77	3,42	3,69
28	1,70	2,05	2,46	2,76	3,40	3,66
29	1,70	2,05	2,46	2,76	3,40	3,66
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,70	3,31	3,55
60	1,67	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,64	1,96	2,33	2,58	3,09	3,29

КРИТИЧНІ ТОЧКИ РОЗПОДІЛУ ФІШЕРА

 k_1, k_2 - числа ступенів свободи)

Таблиця 9.2

Рівень значущості $\alpha = 0,01$										
k_2	k_1									
	1	2	3	4	5	6	7	8	9	10
1	4052	4999	5404	5624	5764	5859	5928	5981	6022	6056
2	98,5	99,0	99,2	99,3	99,3	99,3	99,4	99,4	99,4	99,4
3	34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,5	27,3	27,2
4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5
5	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1
6	13,8	11,0	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87
7	12,3	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62
8	11,3	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81
9	10,6	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26
10	10,1	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69
17	8,40	6,11	5,19	4,67	4,34	4,10	3,93	3,79	3,68	3,59
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37
k_2	k_1									
	11	12	13	14	15	16	17	18	19	20
1	6083	6107	6126	6143	6157	6170	6181	6191	6201	6209
2	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4
3	27,1	27,1	27,0	26,9	26,9	26,8	26,8	26,8	26,7	26,7
4	14,5	14,4	14,3	14,2	14,2	14,2	14,1	14,1	14,0	14,0
5	9,96	9,89	9,82	9,77	9,72	9,68	9,64	9,61	9,58	9,55
6	7,79	7,72	7,66	7,60	7,56	7,52	7,48	7,45	7,42	7,40
7	6,54	6,47	6,41	6,36	6,31	6,28	6,24	6,21	6,18	6,16
8	5,73	5,67	5,61	5,56	5,52	5,48	5,44	5,41	5,38	5,36
9	5,18	5,11	5,05	5,01	4,96	4,92	4,89	4,86	4,83	4,81
10	4,77	4,71	4,65	4,60	4,56	4,52	4,49	4,46	4,43	4,41

11	4,46	4,40	4,34	4,29	4,25	4,21	4,18	4,15	4,12	4,10
12	4,22	4,16	4,10	4,05	4,01	3,97	3,94	3,91	3,88	3,86
13	4,02	3,96	3,91	3,86	3,82	3,78	3,75	3,72	3,69	3,66
14	3,86	3,80	3,75	3,70	3,66	3,62	3,59	3,56	3,53	3,51
15	3,73	3,67	3,61	3,56	3,52	3,49	3,45	3,42	3,40	3,37
16	3,62	3,55	3,50	3,45	3,41	3,37	3,34	3,31	3,28	3,26
17	3,52	3,46	3,40	3,35	3,31	3,27	3,24	3,21	3,19	3,16
18	3,43	3,37	3,32	3,27	3,23	3,19	3,16	3,13	3,10	3,08
19	3,36	3,30	3,24	3,19	3,15	3,12	3,08	3,05	3,03	3,00
20	3,29	3,23	3,18	3,13	3,09	3,05	3,02	2,99	2,96	2,94
<i>Рівень значущості $\alpha = 0,05$</i>										
k₂	k₁									
	1	2	3	4	5	6	7	8	9	10
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35

k_2	k_1									
	11	12	13	14	15	16	17	18	19	20
1	243,0	243,9	244,7	245,4	245,9	246,5	246,9	247,3	247,7	248,0
2	19,40	19,41	19,42	19,42	19,43	19,43	19,44	19,44	19,44	19,45
3	8,76	8,74	8,73	8,71	8,70	8,69	8,68	8,67	8,67	8,66
4	5,94	5,91	5,89	5,87	5,86	5,84	5,83	5,82	5,81	5,80
5	4,70	4,68	4,66	4,64	4,62	4,60	4,59	4,58	4,57	4,56
6	4,03	4,00	3,98	3,96	3,94	3,92	3,91	3,90	3,88	3,87
7	3,60	3,57	3,55	3,53	3,51	3,49	3,48	3,47	3,46	3,44
8	3,31	3,28	3,26	3,24	3,22	3,20	3,19	3,17	3,16	3,15
9	3,10	3,07	3,05	3,03	3,01	2,99	2,97	2,96	2,95	2,94
10	2,94	2,91	2,89	2,86	2,85	2,83	2,81	2,80	2,79	2,77
11	2,82	2,79	2,76	2,74	2,72	2,70	2,69	2,67	2,66	2,65
12	2,72	2,69	2,66	2,64	2,62	2,60	2,58	2,57	2,56	2,54
13	2,63	2,60	2,58	2,55	2,53	2,51	2,50	2,48	2,47	2,46
14	2,57	2,53	2,51	2,48	2,46	2,44	2,43	2,41	2,40	2,39
15	2,51	2,48	2,45	2,42	2,40	2,38	2,37	2,35	2,34	2,33
16	2,46	2,42	2,40	2,37	2,35	2,33	2,32	2,30	2,29	2,28
17	2,41	2,38	2,35	2,33	2,31	2,29	2,27	2,26	2,24	2,23
18	2,37	2,34	2,31	2,29	2,27	2,25	2,23	2,22	2,20	2,19
19	2,34	2,31	2,28	2,26	2,23	2,21	2,20	2,18	2,17	2,16
20	2,31	2,28	2,25	2,22	2,20	2,18	2,17	2,15	2,14	2,12

КРИТИЧНІ ТОЧКИ РОЗПОДІЛУ СТЬЮДЕНТА

Таблиця 9.3

Число ступенів свободи k	Рівень значущості α (двостороння критична область)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,31	12,7	31,82	63,7	318,3	637,0
2	2,92	4,30	6,97	9,92	22,33	31,6
3	2,35	3,18	4,54	5,84	10,22	12,9
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,86
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,40
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59

11	1,80	2,20	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,90	3,65	3,96
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,73	2,09	2,53	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,51	3,79
23	1,71	2,07	2,50	2,81	3,49	3,77
24	1,71	2,06	2,49	2,80	3,47	3,74
25	1,71	2,06	2,49	2,79	3,45	3,72
26	1,71	2,06	2,48	2,78	3,44	3,71
27	1,71	2,05	2,47	2,77	3,42	3,69
28	1,70	2,05	2,46	2,76	3,40	3,66
29	1,70	2,05	2,46	2,76	3,40	3,66
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,70	3,31	3,55
60	1,67	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,64	1,96	2,33	2,58	3,09	3,29

КРИТИЧНІ ТОЧКИ РОЗПОДІЛУ ФІШЕРА

(k_1 , k_2 - числа ступенів свободи)

Таблиця 9.3

Рівень значущості $\alpha = 0,01$										
k_2	k_1									
	1	2	3	4	5	6	7	8	9	10
1	4052	4999	5404	5624	5764	5859	5928	5981	6022	6056
2	98,5	99,0	99,2	99,3	99,3	99,3	99,4	99,4	99,4	99,4
3	34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,5	27,3	27,2
4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5
5	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1
6	13,8	11,0	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87
7	12,3	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62
8	11,3	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81
9	10,6	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26
10	10,1	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85

11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69
17	8,40	6,11	5,19	4,67	4,34	4,10	3,93	3,79	3,68	3,59
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37
<i>Рівень значущості $\alpha = 0,01$</i>										
k₂	k₁									
	11	12	13	14	15	16	17	18	19	20
1	6083	6107	6126	6143	6157	6170	6181	6191	6201	6209
2	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4
3	27,1	27,1	27,0	26,9	26,9	26,8	26,8	26,8	26,7	26,7
4	14,5	14,4	14,3	14,2	14,2	14,2	14,1	14,1	14,0	14,0
5	9,96	9,89	9,82	9,77	9,72	9,68	9,64	9,61	9,58	9,55
6	7,79	7,72	7,66	7,60	7,56	7,52	7,48	7,45	7,42	7,40
7	6,54	6,47	6,41	6,36	6,31	6,28	6,24	6,21	6,18	6,16
8	5,73	5,67	5,61	5,56	5,52	5,48	5,44	5,41	5,38	5,36
9	5,18	5,11	5,05	5,01	4,96	4,92	4,89	4,86	4,83	4,81
10	4,77	4,71	4,65	4,60	4,56	4,52	4,49	4,46	4,43	4,41
11	4,46	4,40	4,34	4,29	4,25	4,21	4,18	4,15	4,12	4,10
12	4,22	4,16	4,10	4,05	4,01	3,97	3,94	3,91	3,88	3,86
13	4,02	3,96	3,91	3,86	3,82	3,78	3,75	3,72	3,69	3,66
14	3,86	3,80	3,75	3,70	3,66	3,62	3,59	3,56	3,53	3,51
15	3,73	3,67	3,61	3,56	3,52	3,49	3,45	3,42	3,40	3,37
16	3,62	3,55	3,50	3,45	3,41	3,37	3,34	3,31	3,28	3,26
17	3,52	3,46	3,40	3,35	3,31	3,27	3,24	3,21	3,19	3,16
18	3,43	3,37	3,32	3,27	3,23	3,19	3,16	3,13	3,10	3,08
19	3,36	3,30	3,24	3,19	3,15	3,12	3,08	3,05	3,03	3,00
20	3,29	3,23	3,18	3,13	3,09	3,05	3,02	2,99	2,96	2,94
<i>Рівень значущості $\alpha = 0,05$</i>										
k₂	k₁									
	1	2	3	4	5	6	7	8	9	10

1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
Уровень значимости $\alpha = 0,05$										
k₂	k₁									
	11	12	13	14	15	16	17	18	19	20
1	243,0	243,9	244,7	245,4	245,9	246,5	246,9	247,3	247,7	248,0
2	19,40	19,41	19,42	19,42	19,43	19,43	19,44	19,44	19,44	19,45
3	8,76	8,74	8,73	8,71	8,70	8,69	8,68	8,67	8,67	8,66
4	5,94	5,91	5,89	5,87	5,86	5,84	5,83	5,82	5,81	5,80
5	4,70	4,68	4,66	4,64	4,62	4,60	4,59	4,58	4,57	4,56
6	4,03	4,00	3,98	3,96	3,94	3,92	3,91	3,90	3,88	3,87
7	3,60	3,57	3,55	3,53	3,51	3,49	3,48	3,47	3,46	3,44
8	3,31	3,28	3,26	3,24	3,22	3,20	3,19	3,17	3,16	3,15
9	3,10	3,07	3,05	3,03	3,01	2,99	2,97	2,96	2,95	2,94
10	2,94	2,91	2,89	2,86	2,85	2,83	2,81	2,80	2,79	2,77
11	2,82	2,79	2,76	2,74	2,72	2,70	2,69	2,67	2,66	2,65
12	2,72	2,69	2,66	2,64	2,62	2,60	2,58	2,57	2,56	2,54
13	2,63	2,60	2,58	2,55	2,53	2,51	2,50	2,48	2,47	2,46
14	2,57	2,53	2,51	2,48	2,46	2,44	2,43	2,41	2,40	2,39
15	2,51	2,48	2,45	2,42	2,40	2,38	2,37	2,35	2,34	2,33

16	2,46	2,42	2,40	2,37	2,35	2,33	2,32	2,30	2,29	2,28
17	2,41	2,38	2,35	2,33	2,31	2,29	2,27	2,26	2,24	2,23
18	2,37	2,34	2,31	2,29	2,27	2,25	2,23	2,22	2,20	2,19
19	2,34	2,31	2,28	2,26	2,23	2,21	2,20	2,18	2,17	2,16
20	2,31	2,28	2,25	2,22	2,20	2,18	2,17	2,15	2,14	2,12

ВКАЗІВКИ ЩОДО ВИКОНАННЯ КОНТРОЛЬНОЇ РОБОТИ

Навчальним планом з дисципліни "Методи прикладного статистичного аналізу" передбачено виконання контрольної роботи. Кількість задач контрольної роботи визначається викладачем. Перед розв'язуванням задач необхідно вивчити відповідний розділ теоретичного матеріалу.

При виконанні контрольної роботи студент повинний дотримувати таких правил:

1. Титульна сторінка роботи оформлюється за зразком, наведеним нижче.
2. Контрольна робота виконується в зошиті, в якому необхідно залишити поля для зауважень рецензента і кілька чистих аркушів для доповнень і відповідей на ці зауваження.
3. Розв'язування кожної задачі треба починати з наведення її повної умови.
4. Рішення задач необхідно супроводжувати поясненнями, графіками та посиланнями на відповідні теоретичні поняття та формули.
5. Якщо контрольна робота після перевірки не зарахована, треба виправити помилки згідно з зауваженнями викладача. Це необхідно робити у кінці роботи (або в окремому зошиті), написавши спочатку титул "Робота над помилками". Вносити зміни до тексту вже перевіреної роботи забороняється. Доопрацьована контрольна робота надсилається для повторної перевірки разом з першим варіантом.
6. Студент, що не виконав контрольну роботу, до заліку не допускається.

Вибір варіантів контрольних робіт

Номер варіанту контрольної роботи обирається згідно з двома останніми цифрами залікової книжки, або студентського квитка. При чому, якщо цей номер перевищує число 50, то номер контрольної роботи визначається наступним чином: від числа 100 віднімається число, що відповідає двом останнім цифрам залікової книжки, або студентського квитка. Наприклад, якщо дві останні цифри залікової книжки - 48, то студент виконує 48 варіант. В разі, коли дві останні цифри залікової книжки 85, то варіант контрольної роботи обирається наступним чином: $100-85=15$ і студент виконує 15 варіант контрольної роботи.

Приклад оформлення титульного аркуша

 КАФЕДРА ПРИКЛАДНОЇ МАТЕМАТИКИ ТА ОБЧИСЛЮВАЛЬНОЇ ТЕХНІКИ	
КОНТРОЛЬНА РОБОТА з дисципліни "Методи прикладного статистичного аналізу"	
група _____	спеціальність _____
_____ <i>(прізвище, ім'я та по батькові)</i>	
Варіант № _____	
№ залікової книжки _____	
Дата здачі роботи до деканату _____	
Викладач _____	
Дніпропетровськ 201	

Рекомендована література

1. Вентцель Е.С. Теория вероятностей. – М.: Физматгиз, 1972.
2. Гнеденко Б.В. Курс теории вероятностей. – М.: Физматгиз, 1961.
3. Гмурман В.Е. Теория вероятностей и математическая статистика. – М.: Высшая школа, 1978.
4. Гурский Е.М. Теория вероятностей с элементами математической статистики. – М.: Высшая школа, 1971.
5. Дудин-Барковский И.В., Смирнов Н.В. Краткий курс математической статистики для технических приложений. – Физматгиз, 1965.
6. Кендалл М, Стьюард А. Основы математической статистики. Том 1,2,3. –Наука, 1973,1976.
7. С. Уилкс Математическая статистика. Наука, 1967.
8. Ю.Нейман. Вводный курс теории вероятностей и математической статистики. Наука, 1968.
9. А. Хальд. Математическая статистика с математическими приложениями. Мир, 1965.
10. Е. Леман. Проверка статистических гипотез. Наука, 1964.

Навчальне видання

Швачич Геннадій Григорович
Коноваленков Володимир Степанович
Соболенко Олександр Вікторович
Заборова Тамара Михайлівна
Христян Володимир Іванович
Єгорцева Євгенія Євгенівна

Навчальний посібник щодо вивчення дисципліни
"Методи прикладного статистичного аналізу"

Підписано до друку _____ Формат 60x84/16. Ум. друк. арк.
Оперативна поліграфія. Зам. № _____. Тираж _____ прим.

НМетАУ

49000, м. Дніпропетровськ, пр. Гагаріна, 4.