*СТАТИСТИКА-

ЦЕ НАУКА , ЯКА ВИВЧАЄ, ОБРОБЛЯЄ Й АНАЛІЗУЄ КІЛЬКІСНІ ДАНІ ПРО НАЙРІЗНОМАНІТНІШІ МАСОВІ ЯВИЩА В ЖИТТІ.

* **МАТЕМАТИЧНА СТАТИСТИКА-**

ЦЕ РОЗДІЛ МАТЕМАТИКИ, ЯКИЙ ВИВЧАЄ МАТЕМАТИЧНІ МЕТОДИ ОБРОБКИ Й ВИКОРИСТАННЯ СТАТИСТИЧНИХ ДАНИХ ДЛЯ НАУКОВИХ І ПРАКТИЧНИХ ВИСНОВКІВ.

Місце математичної статистики в системі статистичних дисциплін

- Відомий статистик-математик професор Є.Є. Слуцький ще у 1912 році писав про значимість методів математичної статистики: "Ми приходимо таким чином до кардинальної вимоги, яку життя ставить діячам статистики: статистик повинен бути математиком, бо його наука є наука математична".

Місце математичної статистики в системі статистичних дисциплін

- В умовах широкого застосування методів сучасної математики в усіх галузях наукових досліджень, фундаментальних і прикладних, а також у вирішенні ряду практичних проблем суспільного життя увага надається математичній статистиці.

Місце математичної статистики в системі статистичних дисциплін

- У біології статистичні значення допомагають під час вивчення генетики, фізіології, екології. Нині жодна серйозна експериментальна робота з біології, медицини не обходиться без статистично обґрунтованого обсягу виконаних експериментів і довірчої оцінки отриманих результатів.

Висновки

- У даний час математична статистика знаходит широке застосування в економіці різних галузе народного господарства, біології, фізиці, хім медицині та ін. На основі її методів можна вирішуват і багато аналітичних задач в галузі економік Зокрема, кількісні характеристики, одержані результаті математико-статистичного аналіз дозволяють мати більш глибоке уявлення пр характер причинно-наслідкових зв'язків явищ, також одержати стійкі надійні параметри дл здійснення економічних розрахунків і особливо метою прогнозування

* Задачі математичної статистики

1. Визначення закону розподілу випадкової величини за статистичними даними;

2. Перевірка правдоподібності гіпотез;

3. Знаходження невідомих параметрів розподілення.

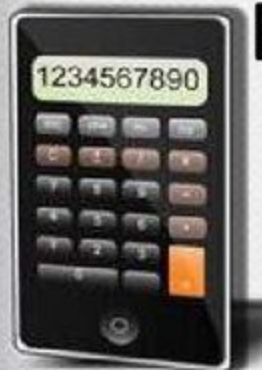
* Основні терміни та поняття статистики:

Генеральна сукупність	**Множина всіх можливих результатів спостереження (вимірювання)**
Статистична вибірка, ряд	**Множина результатів, які реально одержані в даному спостереженні (вимірюванні)**
Варіанта	**Одне із значень елементів вибірки**
Варіаційний ряд	**Упорядкована множина всіх варіант**

Основні поняття математичної статистики з прикладами розв'язування задач

Приклад 1. За результатами контрольної роботи з теорії ймовірностей група за списком одержала такі оцінки:

8; 1; 3; 4; 9; 9; 10; 6; 8; 11; 7; 7; 8; 10; 6.



Варіаційний ряд має такий вигляд: 1, 3, 4, 6, 6, 7, 7, 8, 8, 8, 9, 9, 10, 10, 11.

* ЧИСЛОВІ ХАРАКТЕРИСТИКИ РЯДІВ ДАНИХ (вибірки)

РОЗМАХ ВИБІРКИ - різниця між найбільшим і найменшим значеннями величини у вибірці.

МОДА ВИБІРКИ - значення елемента вибірки, яке зустрічається частіше за інші.

МЕДІАНА ВИБІРКИ - це серединне значення впорядкованого ряду значень випадкової величини.

СЕРЕДНЄ ЗНАЧЕННЯ ВИБІРКИ - середнє арифметичне всіх чисел ряду даної вибірки.

* РОЗГЛЯНЕМО ЗАДАЧУ:

У відділі жіночого взуття протягом трьох днів було проведено обстеження для вивчення попиту на певні розміри взуття. За ці дні було продано 14 пар взуття 37, 38, 39, 40 і 41 розмірів.

1. Запишемо № розміру взуття і кількість відповідних йому проданих пар в таблицю:

* **Генеральна сукупність - кількість всіх пар взуття в магазині.**
* **Вибірка - 14 пар взуття.**
* **37, 37, 37, 38, 38, 38, 38, 38, 39, 39, 39, 40, 40, 41- статистичний, ранжований ряд.**

* **37, 37 , 37, 38, 38, 38, 38, 38, 39, 39, 39, 40, 40, 41 – вибірка**

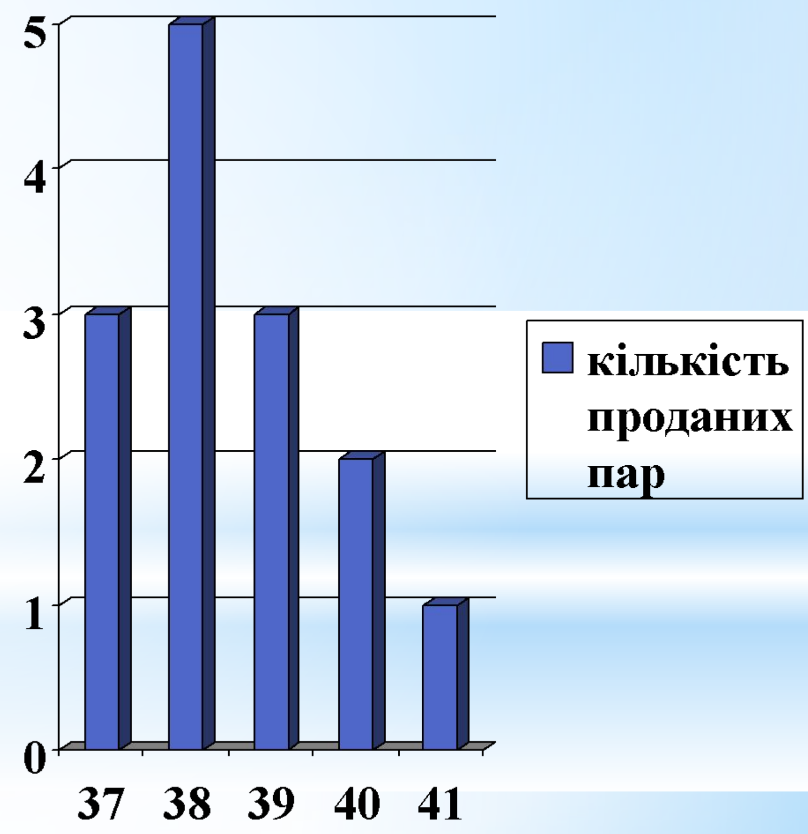
Розмір (варіанта)	37	38	39	40	41
Кількість проданих пар взуття (частота варіанти)	3	5	3	2	1
Відношення кількості проданих пар даного розміру до кількості всіх проданих пар (відносна частота варіанти)	3/14	5/14	3/14	2/14	1/14

- Об'єм вибірки: **14.**
- Розмах вибірки: **41-37=4.**
- Мода вибірки: **38.**
- Медіана вибірки: **38.**
- Середнє значення :

(37+37+37+38+38+38+38+38+ +39+39+39+40+40+41):14=38,5.

Розмір взуття	37	38	39	40	41	варіанти
Кількість проданих пар	3	5	3	2	1	частота варіант

- Для наочного зображення даних спостереження можна використати стовпчикову діаграму. Також можна використовувати зображення у вигляді ламаної, яка називається полігоном частот.
- Графічні зображення дозволяють візуально охопити всю сукупність даних і скласти картину дослідження в цілому.



Задача. Випадкова величина β- кут ковзання літака в момент скидання бомби. Проведено 20 бомбометань, в кожному зареєстровано кут ковзання β в тисячних долях радіана.

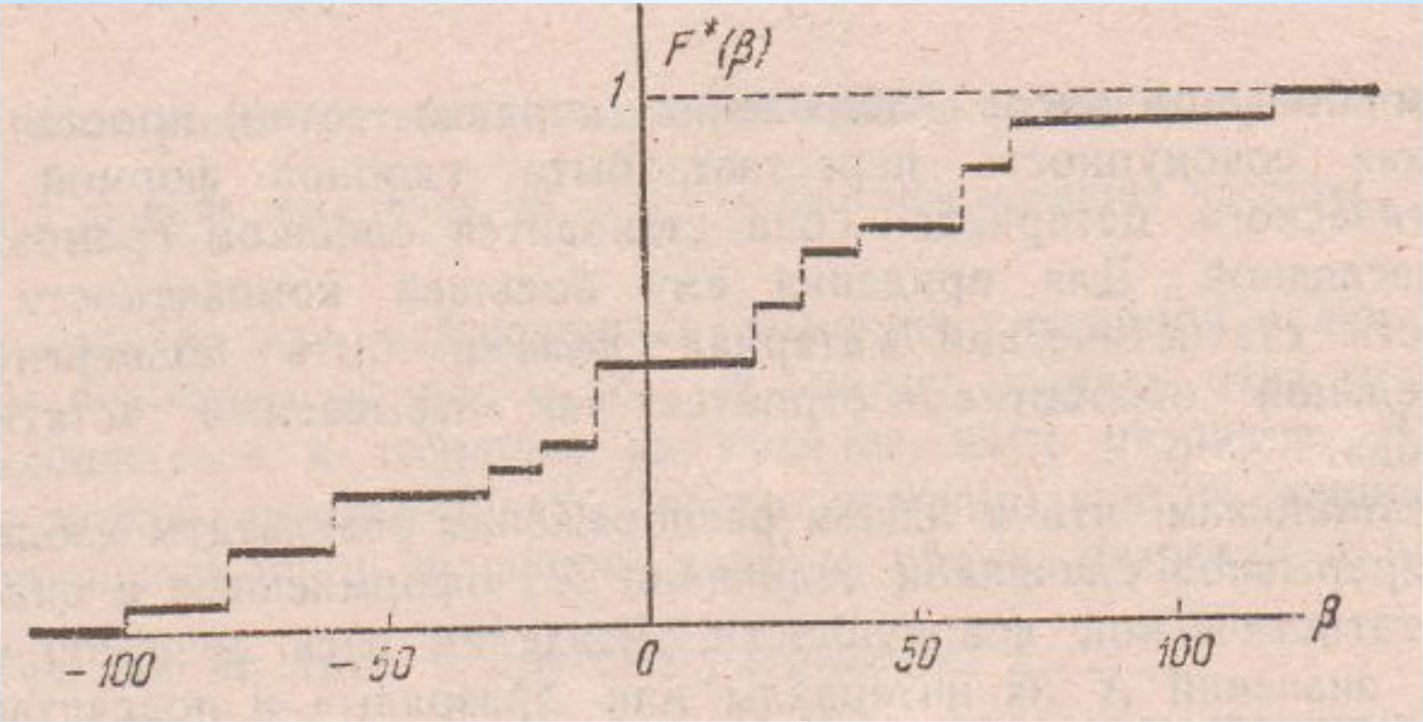
Результати зведені в простий статистичний ряд. Побудувати функцію розподілу.

i	β	i	β	i	β
1	-20	8	-30	15	-10
2	-60	9	120	16	20
3	-10	10	-100	17	30
4	30	11	-80	18	-80
5	60	12	20	19	60
6	70	13	40	20	70
7	-10	14	-60		

в тисячних долях радіана. Результати зведено в простий статистичний ряд. Побудувати функцію розподілу

$$F^*\left(x\right) = P^*\left(x \langle X\right)$$

* Статистична функція розподілу



* Інтервальний статистичний ряд

Задача 15.85. Дані вимірювання частоти серцевих скорочень старшокласників подано у такому вигляді:

Частота скорочень x_i	65	69	70	71	72	73	74	75	76	77
Кількість учнів f_i	5	10	15	16	20	14	5	5	5	5

Потрібно:

а) побудувати графічне зображення даних таблиці;

б) згрупувати частоту серцевих скорочень у чотири групи і побудувати її інтервальний ряд розподілу; виконати відповідне графічне зображення.

Розв'язання: а) наведена таблиця є дискретним рядом розподілу частоти серцевих скорочень. Його графічним зображенням є полігон (рис. 15.12);

б) визначимо величину інтервалу: $x_{max} = 77$, $x_{min} = 65$, $l = \frac{77-65}{4} = \frac{12}{4} = 3$.

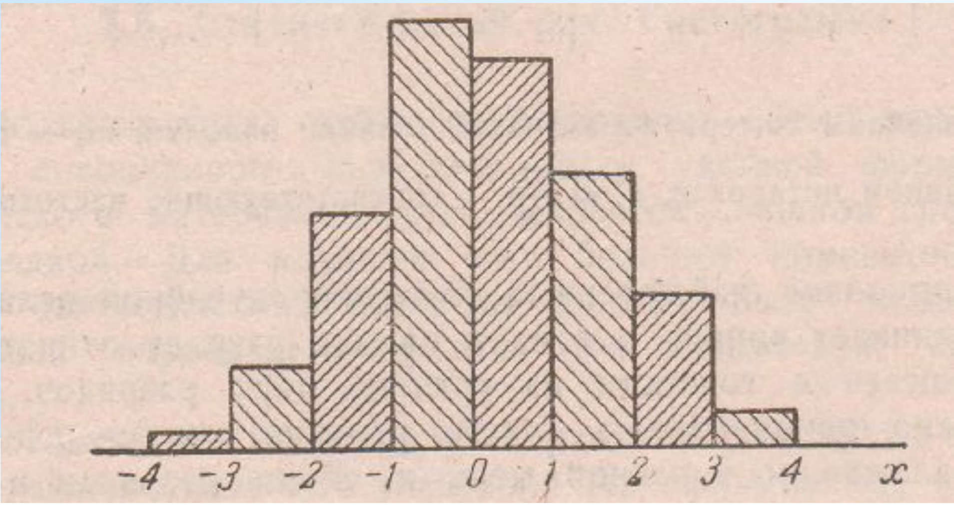
Будуємо інтервальний ряд розподілу:

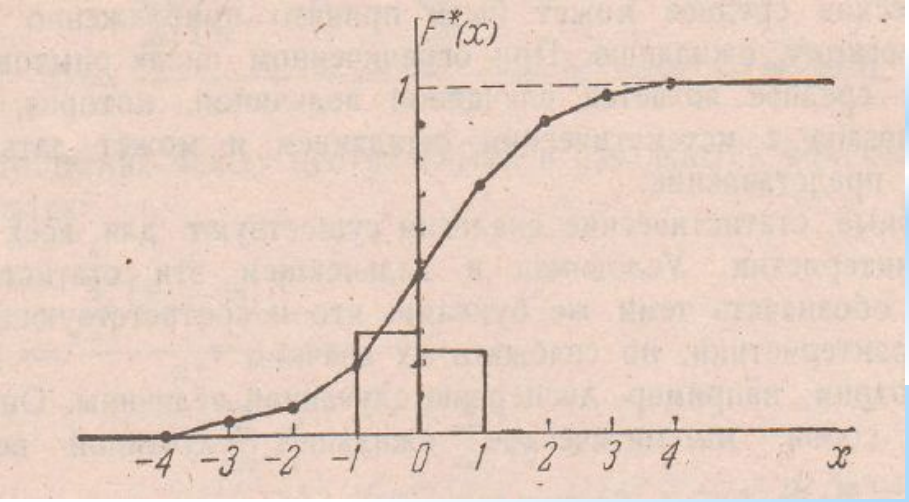
Частота скорочень I_i	65 – 68	68 – 71	71 – 74	74 – 77
Кількість учнів f_i	5	25	50	20

Задача 2. Проведено 500 вимірювань бокової похибки наводки при стрільбі з літака по наземній цілі. Результати вимірювань (в тисячних долях радіана) зведено в статистичний ряд. Побудувати гістограму, наближену статистичну функцію розподілу за даними статистичного ряду.

I_i	$-4; -3$	$-3; -2$	$-2; -1$	$-1; 0$	$0, 1$	$1; 2$	$2; 3$	$3; 4$
m_i	6	25	72	133	120	88	46	10
p_i^*	0,012	0,050	0,144	0,266	0,240	0,176	0,092	0,020

* Гістограма і статистична функція розподілення, побудовані за даними статистичного ряду





Приклад 3. Тридцятьма учасниками зовнішнього незалежно

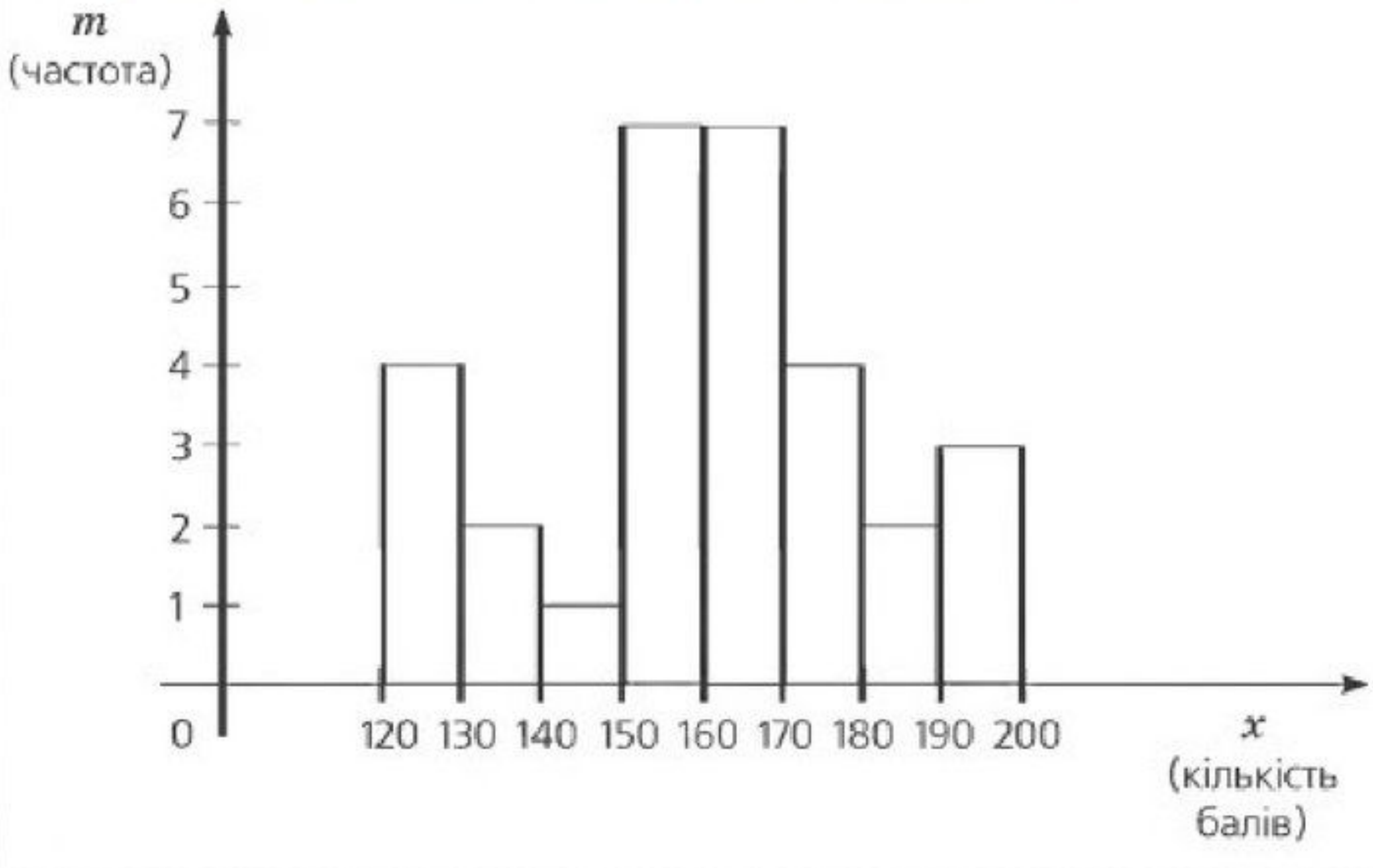
оцінювання з математики були отримані такі бали:

120; 124; 125; 128; 135; 137; 142; 150; 151; 151; 151; 153; 155; 158; 16

160; 160; 165; 165; 166; 168; 170; 172; 175; 175; 180; 186; 190; 194; 200.

Гістограма частот:

Для приклади 3 гістограма частот має такий вигляд:



* Вирівнювання статистичного розподілу задачі 2 за допомогою нормального закону.

I_i	−4; −3	−3; −2	−2; −1	−1; 0	0; 1	1; 2	2; 3	3; 4
m_i	6	25	72	133	120	88	46	10
p_i^*	0,012	0,050	0,144	0,266	0,240	0,176	0,092	0,020

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

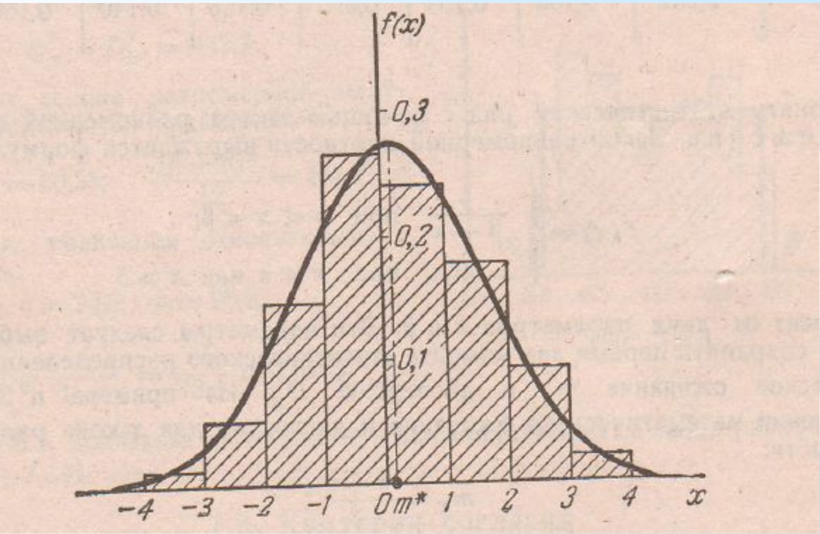
$$m_x^* = -3,5\cdot 0,012 - 2,5\cdot 0,050 - 1,5\cdot 0,144 - 0,5\cdot 0,266 + 0,5\cdot 0,240 + 1,5\cdot 0,176 + 2,5\cdot 0,092 + 3,5\cdot 0,020 = 0,168.$$

$$\alpha_2^* = \sum_{i=1}^{\infty} \tilde{x}_i^2 p_i^* = 2,126.$$

$$D_x^* = \alpha_2^* - (m_x^*)^2 = 2,126 - 0,028 = 2,098.$$

$$m = 0,168; \quad \sigma = 1,448.$$

$$f(x) = \frac{1}{1,448\sqrt{2\pi}} e^{-\frac{(x-0,168)^2}{2\cdot 1,448^2}}$$



x	−4	−3	−2	−1	0	1	2	3	4
$f(x)$	0,004	0,025	0,090	0,199	0,274	0,234	0,124	0,041	0,008

* **Задача. З метою дослідження закону розподілу похибки вимірювання дальності за допомогою радіодальноміра зроблено 400 вимірювань. Результати досліду представлені у вигляді статистичного ряду. Вирівняти ряд за допомогою закону рівномірної густини**

I_i (м)	20; 30	30; 40	40; 50	50; 60	60; 70	70; 80	80; 90	90; 100
m_i	21	72	66	38	51	56	64	32
p_i^*	0,052	0,180	0,165	0,095	0,128	0,140	0,160	0,080

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \text{при} \quad \alpha < x < \beta; \\ 0 & \text{при} \quad x < \alpha \text{ или } x > \beta \end{cases}$$

$$m_x = \frac{\alpha + \beta}{2};$$

$$D_x = \frac{(\beta - \alpha)^2}{12}.$$

$\tilde{x}'_i$	−35	−25	−15	−5	5	15	25	35
p_i^*	0,052	0,180	0,165	0,095	0,128	0,140	0,160	0,080

$$m^*_{x'} = \sum_{i=1}^{k} \tilde{x}'_i p^*_i = 0{,}26.$$

$$\alpha^*_2 = \sum_{i=1}^{k} (\tilde{x}'_i)^2 p^*_i = 447{,}8,$$

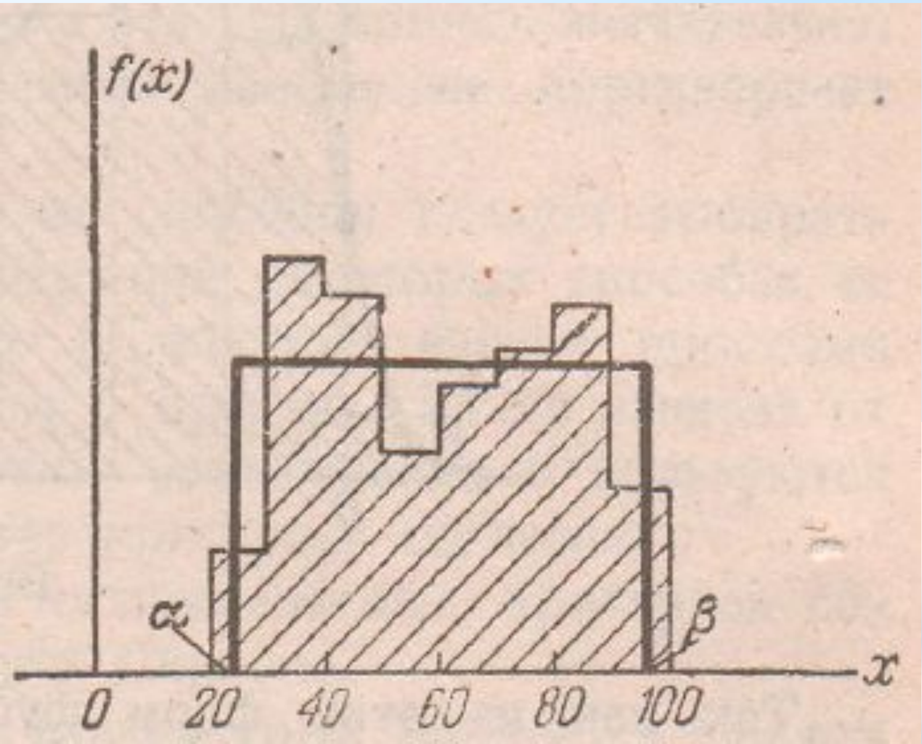
$$D^*_{x'} = \alpha^*_2 - (m^*_{x'})^2 = 447{,}7.$$

$$m^*_x = m^*_{x'} + 60 = 60{,}26$$

$$\frac{\alpha+\beta}{2} = 60{,}26; \qquad \frac{(\beta-\alpha)^2}{12} = 447{,}7.$$

$$\alpha \approx 23{,}6; \quad \beta \approx 96{,}9,$$

$$f(x) = \frac{1}{\beta-\alpha} = \frac{1}{73{,}3} \approx 0{,}0136.$$



Критерії погодження (перевірки)
1. Критерій Пірсона;

$$k_r(u) = \begin{cases} \dfrac{1}{2^{\frac{r}{2}}\Gamma\left(\frac{r}{2}\right)} u^{\frac{r}{2}-1} e^{-\frac{u}{2}} & \text{при} \quad u > 0, \\ 0 & \text{при} \quad u < 0, \end{cases}$$

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt$$

Таблица 4

Значения χ^2 в зависимости от r и p

r \ p	0,99	0,98	0,95	0,90	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,000	0,001	0,004	0,016	0,064	0,148	0,455	1,074	1,642	2,71	3,84	5,41	6,64	10,83
2	0,020	0,040	0,103	0,211	0,446	0,713	1,386	2,41	3,22	4,60	5,99	7,82	9,21	13,82
3	0,115	0,185	0,352	0,584	1,005	1,424	2,37	3,66	4,64	6,25	7,82	9,84	11,34	16,27
4	0,297	0,429	0,711	1,064	1,649	2,20	3,36	4,88	5,99	7,78	9,49	11,67	13,28	18,46
5	0,554	0,752	1,145	1,610	2,34	3,00	4,35	6,06	7,29	9,24	11,07	13,39	15,09	20,5
6	0,872	1,134	1,635	2,20	3,07	3,83	5,35	7,23	8,56	10,64	12,59	15,03	16,81	22,5
7	1,239	1,564	2,17	2,83	3,82	4,67	6,35	8,38	9,80	12,02	14,07	16,62	18,48	24,3
8	1,646	2,03	2,73	3,49	4,59	5,53	7,34	9,52	11,03	13,36	15,51	18,17	20,1	26,1
9	2,09	2,53	3,32	4,17	5,38	6,39	8,34	10,66	12,24	14,68	16,92	19,68	21,7	27,9
10	2,56	3,06	3,94	4,86	6,18	7,27	9,34	11,78	13,44	15,99	18,31	21,2	23,2	29,6
11	3,05	3,61	4,58	5,58	6,99	8,15	10,34	12,90	14,63	17,28	19,68	22,6	24,7	31,3
12	3,57	4,18	5,23	6,30	7,81	9,03	11,34	14,01	15,81	18,55	21,0	24,1	26,2	32,9
13	4,11	4,76	5,89	7,04	8,63	9,93	12,34	15,12	16,98	19,81	22,4	25,5	27,7	34,6
14	4,66	5,37	6,57	7,79	9,47	10,82	13,34	16,22	18,15	21,1	23,7	26,9	29,1	36,1
15	5,23	5,98	7,26	8,55	10,31	11,72	14,34	17,32	19,31	22,3	25,0	28,3	30,6	37,7
16	5,81	6,61	7,96	9,31	11,15	12,62	15,34	18,42	20,5	23,5	26,3	29,6	32,0	39,3
17	6,41	7,26	8,67	10,08	12,00	13,53	16,34	19,51	21,6	24,8	27,6	31,0	33,4	40,8
18	7,02	7,91	9,39	10,86	12,86	14,44	17,34	20,6	22,8	26,0	28,9	32,3	34,8	42,3
19	7,63	8,57	10,11	11,65	13,72	15,35	18,34	21,7	23,9	27,2	30,1	33,7	36,2	43,8
20	8,26	9,24	10,85	12,44	14,58	16,27	19,34	22,8	25,0	28,4	31,4	35,0	37,6	45,3
21	8,90	9,92	11,59	13,24	15,44	17,18	20,3	23,9	26,2	29,6	32,7	36,3	38,9	46,8
22	9,54	10,60	12,34	14,04	16,31	18,10	21,3	24,9	27,3	30,8	33,9	37,7	40,3	48,3
23	10,20	11,29	13,09	14,85	17,19	19,02	22,3	26,0	28,4	32,0	35,2	39,0	41,6	49,7
24	10,86	11,99	13,85	15,66	18,06	19,94	23,3	27,1	29,6	33,2	36,4	40,3	43,0	51,2
25	11,52	12,70	14,61	16,47	18,94	20,9	24,3	28,2	30,7	34,4	37,7	41,7	44,3	52,6
26	12,20	13,41	15,38	17,29	19,82	21,8	25,3	29,2	31,8	35,6	38,9	42,9	45,6	54,1
27	12,88	14,12	16,15	18,11	20,7	22,7	26,3	30,3	32,9	36,7	40,1	44,1	47,0	55,5
28	13,56	14,85	16,93	18,94	21,6	23,6	27,3	31,4	34,0	37,9	41,3	45,4	48,3	56,9
29	14,26	15,57	17,71	19,77	22,5	24,6	28,3	32,5	35,1	39,1	42,6	46,7	49,6	58,3
30	14,95	16,31	18,49	20,6	23,4	25,5	29,3	33,5	36,2	40,3	43,8	48,0	50,9	59,7

* **Для задачі №2 перевірити узгодженість теоретичного і статистичного розподілень.**

$m = 0{,}168, \qquad \sigma = 1{,}448,$

$$p_i = \Phi^*\left(\frac{x_{i+1} - m}{\sigma}\right) - \Phi^*\left(\frac{x_i - m}{\sigma}\right),$$

I_i	—4; —3	—3; —2	—2; —1	—1; 0	0; 1	1; 2	2; 3	3; 4
m_i	6	25	72	133	120	88	46	10
np_i	6,2	26,2	71,2	122,2	131,8	90,5	38,2	10,5

$$\chi^2 = \sum_{i=1}^{8} \frac{(m_i - np_i)^2}{np_i} = 3{,}94.$$

$r = 8 - 3 = 5.$

при $\chi^2 = 3{,}00 \quad p = 0{,}70;$
при $\chi^2 = 4{,}35 \quad p = 0{,}50.$

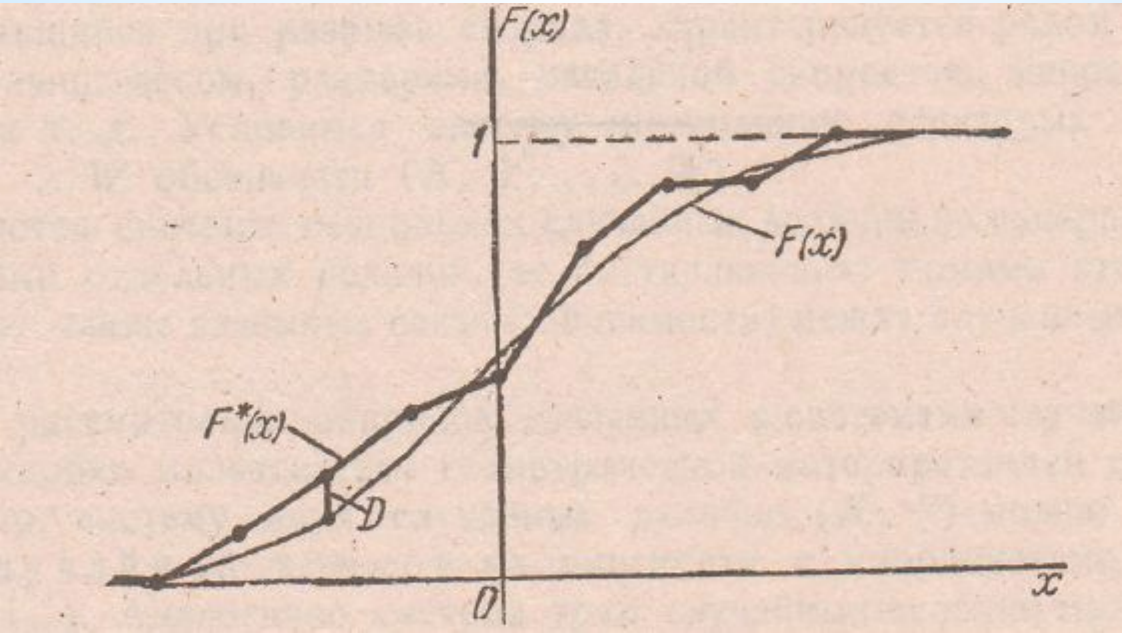
* Критерій Колмогорова

$$D = \max |F^*(x) - F(x)|.$$

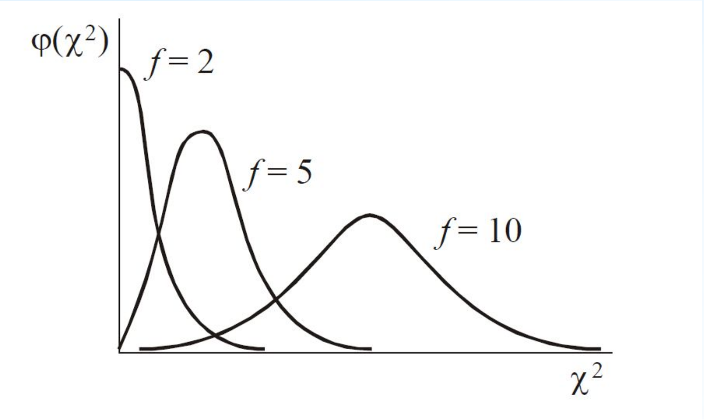
$$D\sqrt{n} \geqslant \lambda$$

$$P(\lambda) = 1 - \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2\lambda^2}.$$

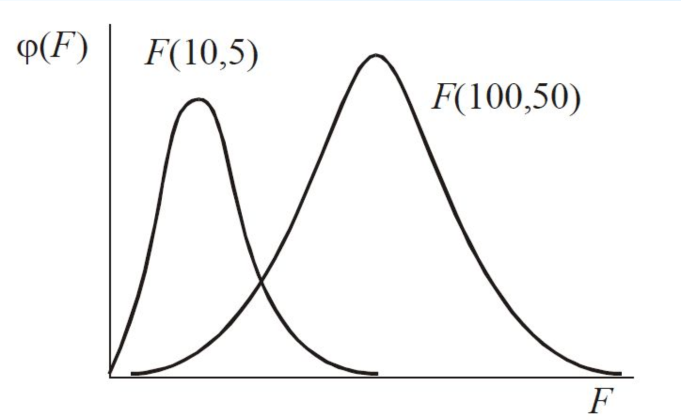
λ	P(λ)	λ	P(λ)	λ	P(λ)
0,0	1,000	0,7	0,711	1,4	0,040
0,1	1,000	0,8	0,544	1,5	0,022
0,2	1,000	0,9	0,393	1,6	0,012
0,3	1,000	1,0	0,270	1,7	0,006
0,4	0,997	1,1	0,178	1,8	0,003
0,5	0,964	1,2	0,112	1,9	0,002
0,6	0,864	1,3	0,068	2,0	0,001



* Критерії погодження



χ^2-розподілення з різними значеннями *f*

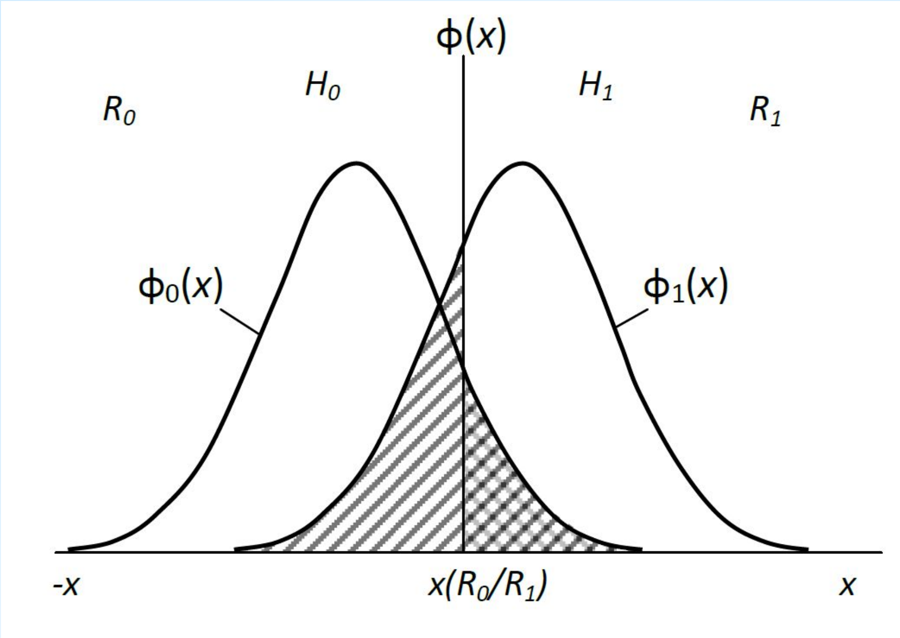


Розподілення Фішера з різними значеннями *f*

$$\varphi(F) = \frac{\Gamma\left(f_1 + \frac{f_2}{2}\right)}{\Gamma\left(\frac{f_1}{2}\right)\cdot\Gamma\left(\frac{f_2}{2}\right)} f_1^{f_1/2} \cdot f_2^{f_2/2} \frac{F^{f_1/2} - 1}{(f_2 + f_1 F)^{f_1 + f_2/2}}$$

,

$$0 \le F < \infty$$



Густина ймовірності розподілу величини *x*

* Статистичні гіпотези і їх перевірка

В якості прикладу проведемо статистичну обробку результатів визначення залишкового вмісту (с, %) стиролу в двох зразках полімеру, отриманого методом безперервної полімеризації у масі

Концентрація стиролу в зразках полімеру

№	Концентрація стиролу, %	
	Зразок I	Зразок II
1	0,56	0,47
2	0,58	0,43
3	0,57	0,42
4	0,54	0,45
5	0,54	

$$\bar{x}_{\mathrm{I}} = \frac{2{,}79}{5} = 0{,}5580;\ \bar{x}_{\mathrm{II}} = \frac{1{,}77}{4} = 0{,}4425;$$

$$\widetilde{D}_{x(I)} = \frac{0{,}001280}{4} = 0{,}0003200;\ \widetilde{\sigma}_{x(I)} = \sqrt{0{,}0003200} = 0{,}01789;$$

$$\widetilde{D}_{x(II)} = \frac{0{,}001475}{3} = 0{,}0004917;\ \widetilde{\sigma}_{x(II)} = \sqrt{0{,}0004917} = 0{,}02217.$$

$$\tau_{\mathrm{I}} = \frac{|0{,}58 - 0{,}558|}{0{,}01789} = 1{,}230 \qquad \tau_{\mathrm{I}} < \tau(0{,}05, n):\ 1{,}230 < 1{,}87$$

$\tau_{\mathrm{II}} = \dfrac{|0{,}47 - 0{,}4425|}{0{,}02217} = 1{,}240$ При $n = 4$ $\tau(\alpha, n)$ дорівнює 1,69. $\tau_{\mathrm{II}} < \tau(0{,}05, n)$

$$F = \frac{\widetilde{D}_{x1}}{\widetilde{D}_{x2}} = \frac{\widetilde{D}_{x(II)}}{\widetilde{D}_{x(I)}} = \frac{0{,}0004917}{0{,}0003200} = 1{,}537.$$

Значення F-розподілу Фішера $F(\alpha, f_1, f_2)$ дорівнює 6,591.

Виконується нерівність $F < F(0{,}05, f_1, f_2)$.

$$\widetilde{s}_A^2 = \frac{4 \cdot 0{,}00032 + 3 \cdot 0{,}0004917}{4+3} = 0{,}0003936;$$

$$f_A = 4 + 3 = 7$$

$$t_{\bar{x}} = \frac{|0{,}5580 - 0{,}4425|}{\sqrt{0{,}0003936}} \sqrt{\frac{5 \cdot 4}{5+4}} = 8{,}679$$

Оцінюємо спочатку можливість прийняття гіпотези. Для цього знаходимо $t(0{,}025, f_A) = 2{,}365$. Так як нерівність $t \le t(0{,}025, f_A)$ не виконується (8,679 > 2,365), то з 5 %-вим рівнем значущості гіпотеза, що перевіряється, не може бути прийнята.

Тепер оцінимо можливість відкидання гіпотези. Знаходимо $t(0{,}005, f_A) = 3{,}499$. У цьому випадку нерівність $t > t(0{,}005, f_A)$ виконується (8,679 > 3,499) і гіпотеза, що оцінюється, H_0 повинна бути відкинута.