

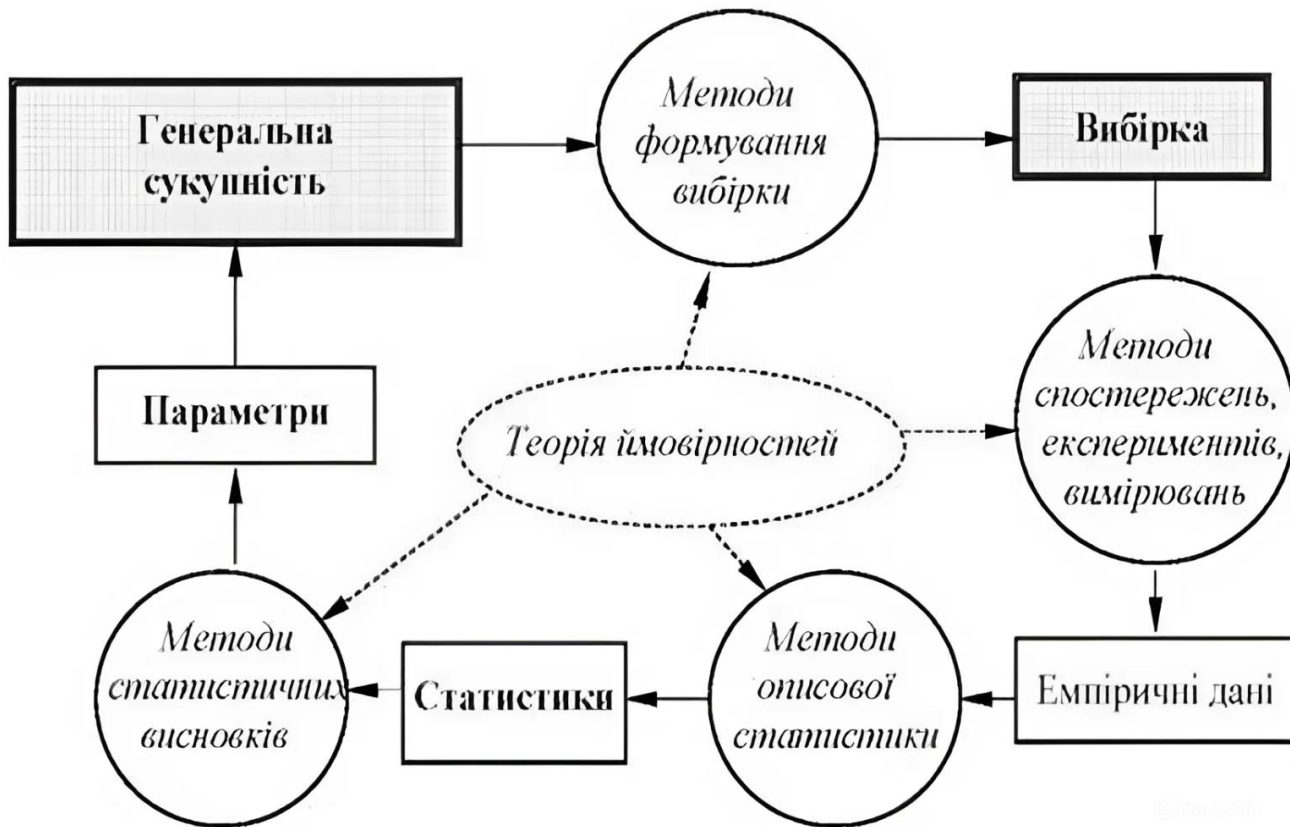
ЗАСТОСУВАННЯ СТАТИСТИЧНИХ МЕТОДІВ В АГРОНОМІЧНИХ ДОСЛІДЖЕННЯХ

ОСНОВНІ ПОНЯТТЯ ТА ЗАВДАННЯ МАТЕМАТИЧНОЇ СТАТИСТИКИ

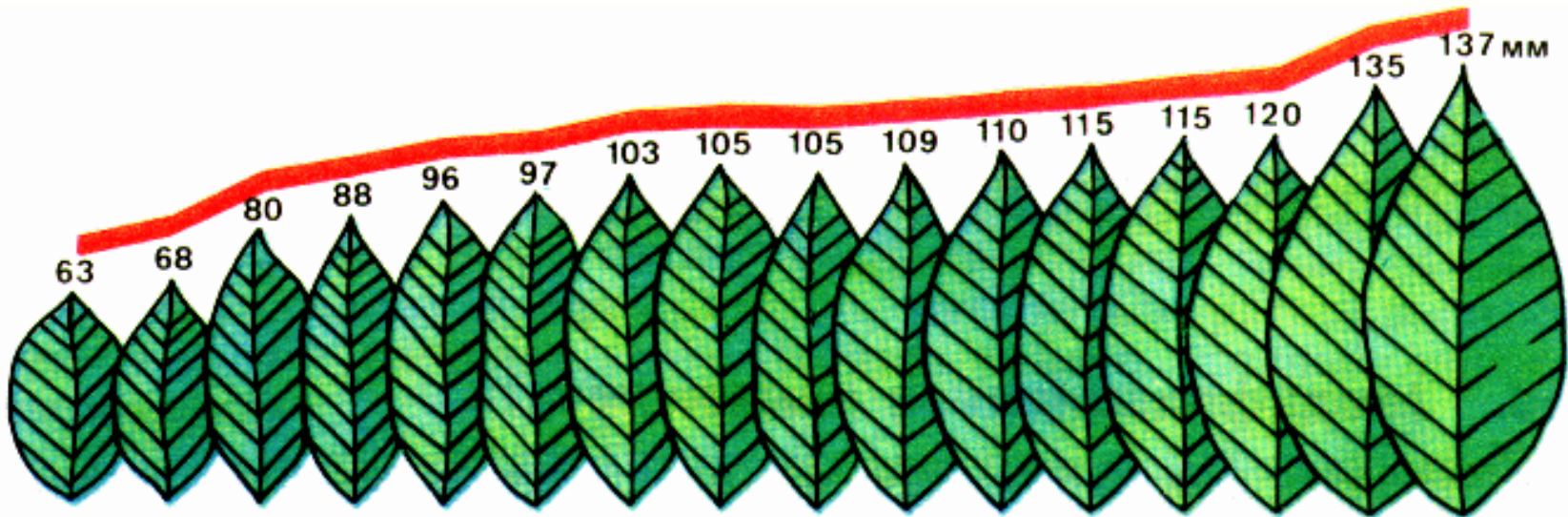
1. Аналіз варіаційних рядів кількісної і якісної мінливості
2. Обчислення статистичних характеристик малої та великої вибірки
3. Підготовка даних до статистичного аналізу
4. Бракування і відновлення даних
5. Дисперсійний аналіз
6. Кореляційний і регресійний аналіз



АНАЛІЗ ВАРІАЦІЙНИХ РЯДІВ КІЛЬКІСНОЇ І ЯКІСНОЇ МІНЛИВОСТІ



ВАРІАЦІЙНИЙ РЯД



Обчислення статистичних характеристик малої та великої вибірки

$$n = t^2 \frac{V^2}{\Delta\%},$$

$$s = \frac{X \text{ max} - X \text{ min}}{6},$$

$$V = \frac{s}{\bar{x}} 100,$$

$$\bar{x} = \frac{X \text{ max} + X \text{ min}}{2},$$

$$\bar{x} = \frac{\sum X}{n}$$

$$\bar{x} = \frac{\sum fX}{n}$$

$$s^2 = \frac{\sum X^2 - (\sum X)^2 : n}{n - 1}$$

$$s = \sqrt{s^2}$$

$$s_x = \frac{s}{\sqrt{n}} = \sqrt{\frac{s^2}{n}}$$

$$s_{\bar{x}\%} = \frac{s_{\bar{x}}}{\bar{x}} 100$$

Показники якісної мінливості

$$s = \sqrt{pq}$$

$$s = \sqrt[k]{p_1 \times p_2 \times \dots \times p_k}$$

$$\lg s = \frac{\lg p_1 + \lg p_2 + \dots + \lg p_k}{k}$$

$$V = \frac{s}{s_{\max}} 100$$

$$sp = \frac{s}{\sqrt{N}} = \sqrt{\frac{pq}{N}}$$

Закономірності нормального розподілу

$$\bar{x} \pm s$$

68,26 %

$$\bar{x} \pm 2s$$

95,46 %

$$\bar{x} \pm 3s$$

99,73 %

Підготовка даних до статистичного аналізу

База даних урожайності озимої пшениці залежно від кількості опадів та зрошувальної норми

Рік	Забезпеченість року	Опади, м ³ /га		Зрошувальна норма, м ³ /га		Сумарне водоспоживання шару 0-200 см, м ³ /га	Сума опадів і зрош. норми, м ³ /га	Сума весняно-літніх опадів і зрош. норми, м ³ /га	Врожайність, ц/га
		осінь-весна	весна-літо	всього	весна-літо				
1988	волог.	1458	2435	1700	1050	3914	5593	3485	73,9
1989	серед.	1317	993	2300	1450	4016	4610	2443	89,7
1990	серед.	878	1957	2050	1450	4608	4885	3407	82,4
1991	волог.	2136	1778	2050	1500	4279	5964	3278	64,4
1992	серед.	1258	1138	2050	1550	4347	4446	2688	43
1993	волог.	1713	1622	1800	1050	4016	5135	2672	73,6
1994	волог.	1065	1412	2300	1550	4506	4102	2287	47,6
1996	сухий	2179	669	1950	1600	4929	4798	2269	35,6
1997	волог.	1579	1804	1600	1100	5189	4983	2904	33,7
1998	волог.	2627	2485	800	500	6341	5912	2985	41,2
1999	серед.	1723	1520	1350	1000	5340	4593	2520	58,6
2000	сухий	2206	1264	1400	1100	5126	4870	2364	49,4
2001	волог.	998	1925	400	350	3971	3323	2275	55,9
2002	сухий	1246	1463	700	1500	5351	5009	1463	59,8
2003	сухий	2230	1456	1000	1000	5062	4686	2456	20,2
2004	волог.	1893	2731	1200	800	6138	5824	3531	76,4
2005	серед.	2626	1960	1200	800	6261	5786	2760	57,9
2006	середесух.	2337	1243	-	1000	5122	4580	1243	42,4
2007	сухий	1262	385	1500	2200	4401	3847	1885	52,5
2008	середвол.	1104	2914	1000	500	5347	5018	3914	84,0

Бракування і відновлення даних

- 1) Агрономічний аналіз (відповідність методики, критичний огляд даних, робота над помилками і уточненнями)
- 2) Первинна цифрова обробка матеріалів (перерахунок урожаю з ділянки на га, приведення до стандартної вологості й засміченості, складання таблиці врожаїв)
- 3) Статистична оцінка результатів досліджень



Дисперсійний аналіз

$$F = \frac{s_1^2}{s_2^2}$$

Суть дисперсійного аналізу полягає в розкладанні загальної суми квадратів відхилень і загального числа ступенів свободи на компоненти, що відповідають структурі дослідження і оцінці значущості дії і взаємодії факторів, що вивчаються.

Теорія планування досліджень і статистичний аналіз базуються на принципах рендомізації. В цьому випадку дисперсійний аналіз дає правильну похибку експерименту.

Якщо дослід не рендомізований, то можна отримати зміщену похибку дослідження і критерії значущості втрачають законну силу. Тому вони не можуть використовуватися як аргументи статистичного доведення ефектів .

Урожайність озимої пшениці в досліді

Варіанти	Урожайність, ц/га				Суми V	Середнє
1 st	41,5	47,0	44,0	43,0	175,5	43,88
2	49,6	54,2	49,0	51,0	203,8	50,95
3	58,6	57,2	60,0	61,3	237,1	59,28
4	42,1	41,2	47,6	41,3	172,2	43,05
5	56,5	57,0	58,0	51,2	222,7	55,68
Суми P	248,3	256,6	258,6	247,8	1011,3	50,57

Порядок обчислень

- Загальне число спостережень $N = \sum n = 4 \cdot 5 = 20$
- Корируючий фактор $C = (\sum X)^2 : N = 1011,3^2 : 20 = 51136,385$
- $C_Y = \sum X^2 - C = (41,5^2 + 44^2 + \dots + 51,2^2) - 51136,385 = 911,15$
- $C_V = \sum V^2 : n - C = (175,5^2 + 203,8^2 + 237,1^2 + 172,2^2 + 222,7^2) : 4 - 51136,385 = 813,42$
- $C_p = \sum P^2 : l - C = (248,3^2 + 256,6^2 + 258,6^2 + 247,8^2) : 5 - 51136,385 = 18,67$
- $C_z = C_Y - C_V - C_p = 911,15 - 813,42 - 18,67 = 79,06$

Варіант	Урожайність	Різниця з стандартом		Група
		ц	%	
1	43,88	–	–	
2	50,95	7,07	16,11	I
3	59,28	15,4	35,1	I
4	43,05	–0,83	–1,89	II
5	55,68	11,8	26,89	I

Кореляційний і регресійний аналіз

- Кореляція може бути лінійною і криволінійною за формою, прямою і зворотною за напрямом, простою, коли досліджується залежність між двома ознаками, і множинною, коли досліджуються три і більше ознак. Для оцінки тісноти кореляції використовують її коефіцієнт.
- Величина коефіцієнта кореляції лежить в межах від -1 до $+1$, є безрозмірною величиною, яка розраховується за формулою:

$$r = \frac{\sum (X - \bar{x})(Y - \bar{y})}{\sqrt{\sum (X - \bar{x})^2 \sum (Y - \bar{y})^2}} = \frac{\sum XY - (\sum X \sum Y) : n}{\sqrt{(\sum X^2 - (\sum X)^2 : n)(\sum Y^2 - (\sum Y)^2 : n)}}$$

Вважається, що при $r < 0,3$ кореляція між ознаками слабка, $r = 0,3 \dots 0,7$ – середня і при $r > 0,7$ – сильна. Але ступінь взаємозв'язку ознак найповніше характеризує коефіцієнт детермінації $d = r^2$, який показує долю (%) змін, які залежать від фактора, що вивчається.

Для оцінки надійності коефіцієнта кореляції розраховують його помилку і критерії істотності:

$$sr = \sqrt{\frac{1-r^2}{n-2}}$$

$$t_r = \frac{r}{s_r}$$

РЕГРЕСІЯ

$$b_{yx} = \frac{\sum (X - \bar{x})(Y - \bar{y})}{\sum (X - \bar{x})^2};$$

$$b_{xy} = \frac{\sum (X - \bar{x})(Y - \bar{y})}{\sum (Y - \bar{y})^2};$$

$$b_{yx} \cdot b_{xy} = r^2 = d$$

$$s_{byx} = s_r \sqrt{\frac{\sum (Y - \bar{y})^2}{\sum (X - \bar{x})^2}}; s_{bxy} = s_r \sqrt{\frac{\sum (X - \bar{x})^2}{\sum (Y - \bar{y})^2}}$$

$$t_b = b / s_b.$$

Приклад: Провести кореляційний і регресійний аналіз дослід, в якому досліджується взаємозв'язки маси 1000 зерен і урожайності озимої пшениці

Номер пари	Значення ознак		X ²	Y ²	XY
	X, г	Y, ц/га			
1	38,1	48	1451,61	2304	1828,8
2	43,3	58,2	1874,89	3387,24	2520,06
3	47,3	57,4	2237,29	3294,76	2715,02
4	42,1	50,7	1772,41	2570,49	2134,47
5	45,2	58,2	2043,04	3387,24	2630,64
6	47,4	56,3	2246,76	3169,69	2668,62
7	45,3	50	2052,09	2500	2265
8	44,3	51,6	1962,49	2662,56	2285,88
Сума	$\Sigma X = 353$	$\Sigma Y = 430,4$	$\Sigma X^2 = 15640,5$ 8	$\Sigma Y^2 = 23275,98$	$\Sigma XY = 19048,49$

Визначення допоміжних величин

$$\bar{x} = \sum X : n = 44,13 \qquad \bar{y} = \sum Y : n = 53,8$$

$$\sum (X - \bar{x})^2 = \sum X^2 - (\sum X)^2 : n = 15640,38 - 15576,125 = 64,46$$

$$\sum (Y - \bar{y})^2 = \sum Y^2 - (\sum Y)^2 : n = 23275,98 - 23155,52 = 120,46$$

$$\sum (X - \bar{x})(Y - \bar{y}) = \sum XY - (\sum X \sum Y) : n = 19048,49 - 18991,4 = 57,09$$

$$r = \frac{\sum (X - \bar{x})(Y - \bar{y})}{\sqrt{\sum (X - \bar{x})^2 \sum (Y - \bar{y})^2}} = \frac{57,09}{\sqrt{64,46 \times 120,46}} = 0,648$$

$$b_{yx} = \frac{\sum (X - \bar{x})(Y - \bar{y})}{\sum (X - \bar{x})^2} = \frac{57,09}{64,46} = 0,89 \text{ц / га}$$

Рівняння регресії:

$$Y = \bar{y} + b_{yx}(X - \bar{x}) = 53,8 + 0,89(X - 44,13) = 0,89X + 14,52$$

$$sr = \sqrt{\frac{1-r^2}{n-2}} = 0,097$$

$$s_{byx} = s_r \sqrt{\frac{\sum (Y - \bar{y})^2}{\sum (X - \bar{x})^2}} = 0,097 \sqrt{\frac{120,46}{64,46}} = 0,133 \text{ ц / га}$$

$$s_{yx} = s_r \sqrt{\sum (Y - \bar{y})^2} = 0,097 \sqrt{120,46} = 1,06 \text{ ц / га}$$

Критерії значущості $t_r = r/s_r = 6,68$; $v = n - 2 = 8 - 2 = 6$;
 $t_{05} = 2,45$

Довірчі інтервали $r \pm t_{05} sr = 0,65 \pm 2,45 \cdot 0,097 = 0,65 \pm 0,24$ (0,45 ÷ 0,89)
 $b_{yx} \pm t_{05} s_{byx} = 0,89 \pm 2,45 \cdot 0,0133 = 0,89 \pm 0,133$ (0,56 ÷ 1,22)

ГРАФІК РЕГРЕСІЇ

