МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
МИКОЛАЇВСЬКИЙ НАЦІОНАЛЬНИЙ АГРАРНИЙ УНІВЕРСИТЕТ
Факультет ТВППТСБ

Кафедра генетики, годівлі тварин та біотехнології

СТАТИСТИЧНІ МЕТОДИ У БІОТЕХНОЛОГІЇ

Методичні рекомендації
для виконання лабораторно-практичних робіт та вивчення дисципліни для здобувачів вищої освіти СВО «Магістр» освітньої спеціальності 162 – «Біотехнології та біоінженерія» денної форми навчання

Миколаїв – 2020

УДК 573.6.086.83
С-78

Друкується за рішенням науково-методичної комісії факультету ТВППТСБ Миколаївського національного аграрного університету від 22 травня 2020 р., протокол № 10.

Укладач:

С С. Крамаренко – д-р біол. наук, професор, професор кафедри генетики, годівлі тварин та біотехнології Миколаївського національного аграрного університету.

Рецензенти:

О.В. Жуков – д-р біол. наук, доцент, професор кафедри зоології та екології Дніпропетровського національного університету імені Олеся Гончара;
Є.В. Баркарь – кандидат с.-г. наук, доцент, доцент кафедри генетики, годівлі тварин та біотехнології Миколаївського НАУ.

ЗМІСТ

ВСТУП

Методичні рекомендації розроблені для здобувачів вищої освіти СВО «Магістр» зі спеціальності 162 - «Біотехнології та біоінженерія». Їх основна мета – допомогти викладачу при проведенні лабораторно-практичних занять із здобувачами вищої освіти денної форм навчання із дисципліни "Статистичні методи у біотехнології" й надати їм сучасної теоретичної підготовки з біометрії та математичної статистики перед опануванням комп'ютерних прикладних програм (MS Excel, STATISTICA, SPSS, Statgraphics і т.п.). Результати цієї роботи повинні логічно перетворитися у випускну роботу.

Дисципліна «Статистичні методи у біотехнології» спрямована на статистичний опис результатів експериментів і спостережень, побудову математичних моделей масових процесів і явищ та планування і прогнозування результатів подальших досліджень. Необхідність використання математико-статистичних методів у біотехнології обумовлена в першу чергу тим, що властивості біологічних об'єктів зазвичай істотно варіюються в межах певної вибірки та залежать від багатьох чинників, а біотехнологічні та виробничі процеси зазнають певні флуктуації у часі.

ПРАКТИЧНА РОБОТА № 1

Тема: Аналіз вибіркових даних
Мета: Набути навичок розрахунку вибіркових показників на підставі варіаційного ряду

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ

Варіаційний ряд розподілу – це подвійний ряд, в якому для ранжованих значень варіант вибірки наведено частоти їх зустрічаємості. Форму запису сукупності у вигляді варіаційного ряду часто-густо використовують для статистичного аналізу великих вибірок. Розглянемо на одному прикладі особливості побудови варіаційного ряду та розрахунку вибіркових показників за його допомогою. Проаналізуємо дані щодо вмісту жиру у молоці матерів.

1. Для побудови інтервального варіаційного ряду по-перше необхідно визначити кількість інтервалів та його ширину. Існує велика кількість правил щодо обрання необхідної кількості інтервалів, найвідоміше з них – це правило Старджеса:

$$i = 1 + 3{,}32 \cdot \lg \quad n, \qquad (1.1)$$

де n – обсяг вибірки.

Але доцільніше для кількості інтервалів варіаційного ряду рекомендується обирати непарне число, найближче до значення, що дає правило Старджеса. В нашому прикладі значення, що дає правило Старджеса дорівнює: $i = 1 + 3{,}32 \cdot \lg \ \ 113 = 7{,}82$. Доцільно обрати найближче непарне число, тобто, 7.

Ширина інтервалу визначається за формулою:

$$h = \frac{x_{\max} - x_{\min}}{i}. \qquad (1.2)$$

Але для полегшення подальших розрахунків ширина інтервалу повинна бути кратною 10 (наприклад, 0,1 , 20 або 250 та т.п.). Для цього мінімальне значення вибірки ($x_{\min}$) у формулі 1.2 можна трохи зменшити, а максимальне ($x_{\max}$) – трохи збільшити. (Але ніяк не павпаки!)

В нашому прикладі $x_{\max}$ = 4,49 %, $x_{\min}$ = 3,25 %. Різниця між ними складає 1,24.Найближче число, яке дробове до 7 та 10 – це 1,40. Тому ми трохи зменшимо мінімальне значення (до 3,20) й підвищимо максимальне (до 4,60). Тоді ширина інтервалу складатиме:

$$h = \frac{4{,}60 - 3{,}20}{7} = 0{,}20.$$

Межами нашого варіаційного ряду будуть значення:

3,20 – 3,40 – 3,60 – 3,80 – 4,00 – 4,20 – 4,40 – 4,60.

Але для того, щоб значення вибірки, що попадають на межу між класами однозначно трактувалося, підвищимо нижні межі класових інтервалів (починаючи із другого) на одну одиницю мінімального розряду (у нашому випадку, на 0,01). Таким чином вихідна матриця для нашого варіаційного ряду матиме наступний вигляд – табл. 1.1 (перші два стовпчики).

Таблиця 1.1

i	*Інтервал*					
1	3,20-3,40					
2	3,41-3,60					
3	3,61-3,80					
4	3,81-4,00					
5	4,01-4,20					
6	4,21-4,40					
7	4,41-4,60					
	Всього					

2. Після того як класові інтервали визначені підраховується абсолютні частоти, з якими варіанти вибірки попадають у той чи інший класовий інтервал (стовпчик n_i в табл. 1.1). Крім того необхідно розрахувати накопичені частоти (стовпчик s_i в табл. 1.1). Накопичена частота класового інтервалу – це кількість значень у вибірки, що менше чи дорівнюють верхній межі класового інтервалу.

Для першого інтервалу: $s_1 = n_1$; для другого: $s_2 = s_1 + n_2$; для третього: $s_3 = s_2 + n_3$ і т.п. Накопичені частоти потрібні нам для розрахунку медіани вибірки та для перевірки відповідності розподілу вибірки нормальному закону (див. Практичну роботу № 2).

i	*Інтервал*	n_i	s_i			
1	3,20-3,40	4	4			
2	3,41-3,60	21	25			
3	3,61-3,80	40	65			
4	3,81-4,00	30	95			
5	4,01-4,20	15	110			
6	4,21-4,40	0	110			
7	4,41-4,60	3	113			
	Всього	113	-			

3. Необхідно визначити середини класових інтервалів (ось для чого ширина інтервалу була обрана дробовою 10). Для першого інтервалу середина складає:

$$x_1 = \frac{3{,}20 + 3{,}40}{2} = 3{,}30,$$

для другого:

$$x_2 = \frac{3{,}40 + 3{,}60}{2} = 3{,}50,$$

і т.п. (Одиницю, на яку ми збільшили нижню межу кожного інтервалу в рахунок не беремо.)

4. Розраховуємо вибіркове середнє арифметичне значення за формулою:

$$\bar{x} = \frac{\sum n_i x_i}{n}. \quad (1.3)$$

Для цього необхідно у межах кожного інтервалу помножити його середину на абсолютну частоту. Тобто, для першого інтервалу:

$$n_1 \cdot x_2 = 4 \cdot 3{,}30 = 13{,}20 \text{ і т.п.}$$

Всі необхідні для цього значення розраховуємо у табл.1.1 (стовпчик $n_i \cdot x_i$) та знаходимо суму значень цього стовпчика:

i	*Інтервал*	n_i	s_i	x_i	$n_i \cdot x_i$	
1	3,20-3,40	4	4	3,30	13,20	
2	3,41-3,60	21	25	3,50	73,50	
3	3,61-3,80	40	65	3,70	148,00	
4	3,81-4,00	30	95	3,90	117,00	
5	4,01-4,20	15	110	4,10	61,50	
6	4,21-4,40	0	110	4,30	0,00	
7	4,41-4,60	3	113	4,50	13,50	
Всього		113	-	-	426,70	

Таким чином, вибіркове середнє арифметичне значення складатиме:

$$\bar{x} = \frac{426{,}70}{113} = 3{,}776 \approx 3{,}78.$$

5. Значення середнього квадратичного відхилення (с.к.в.) розраховується за формулою:

$$\sigma = \sqrt{\frac{\sum n_i (x_i - \bar{x})^2}{n-1}}. \quad (1.4)$$

Для цього необхідно у межах кожного інтервалу помножити його абсолютну частоту на квадрат різниці між серединою цього інтервалу та вибірковим середнім арифметичним значенням. Для першого інтервалу ця величина складатиме:

$$n_1 \cdot (x_1 - \bar{x})^2 = 4 \cdot (3{,}30 - 3{,}78)^2 = 0{,}9216 \text{ і т.п.}$$

Всі необхідні для розрахунку с.к.в. значення занести в табл. 1.1 та знайти їх суму:

i	*Інтервал*	n_i	s_i	x_i	$n_i \cdot x_i$	$n_i \cdot (x_i - \bar{x})^2$
1	3,20-3,40	4	4	3,30	13,20	0,9216
2	3,41-3,60	21	25	3,50	73,50	1,6464
3	3,61-3,80	40	65	3,70	148,00	0,2560
4	3,81-4,00	30	95	3,90	117,00	0,4320
5	4,01-4,20	15	110	4,10	61,50	1,5360
6	4,21-4,40	0	110	4,30	0,00	0,0000
7	4,41-4,60	3	113	4,50	13,50	1,5552
Всього		113	-	-	426,70	6,3472

Таким чином, вибіркове значення с.к.в. складатиме:

$$\sigma = \sqrt{\frac{6{,}3472}{113-1}} = 0{,}238.$$

Але, оскільки розрахунок с.к.в. проводився не на підставі окремих значень, а на підставі варіаційного ряду, необхідно врахувати поправку Шепарда. Для цього від розрахованого значення необхідно відняти величину (h/12). Таким чином, остаточне значення с.к.в. складатиме:

$$\sigma = 0{,}238 - \frac{0{,}20}{12} = 0{,}221 \approx 0{,}22.$$

6. Коефіцієнт варіації розраховується за формулою:

$$CV = \frac{\sigma}{\bar{x}} \cdot 100\%, \tag{1.5}$$

а його статистична помилка:

$$S_{CV} = \frac{CV}{\sqrt{2n}}. \tag{1.6}$$

Для даних із нашого прикладу значення коефіцієнта варіації складатиме: $CV = \frac{0{,}22}{3{,}78} \cdot 100\% = 5{,}82\%$ із статистичною помилкою: $S_{CV} = \frac{5{,}82}{\sqrt{2 \cdot 113}} = 0{,}39\%$.

7. Альтернативним показником, який також як і вибіркове середнє арифметичне використовується для оцінки центру розподілу є вибіркова медіана. Для інтервального варіаційного ряду значення медіани розраховується за формулою:

$$Me = x_0 + h \cdot \left(\frac{\frac{n}{2} - s_{Me-1}}{n_{Me}} \right), \tag{1.7}$$

де x_0 – нижня межа медіанного інтервалу; $s_{\text{Me-1}}$ – накопичені частоти інтервалу, який передує медіанному; n_{Me} – частота медіанного інтервалу.

Медіанним є інтервал 3,61-3,80, оскільки на цей інтервал приходиться перша накопичена частота, що перевищує половину всього обсягу вибірки (65 перевищує 113/2 = 56,5). Таким чином значення медіани складає:

$$Me = 3{,}61 + 0{,}2 \cdot \left(\frac{\frac{113}{2} - 25}{40} \right) = 3{,}77 \ .$$

►**<u>Завдання</u>**: Розрахувати вибіркові характеристики для показнику, що вказаний викладачем.◄

Література. **1**: с.7-30; **2**: с.23-27; **3**: с.7-14; **4**: с.66-75, с.98-135, с.131-135.

Питання для самоперевірки:

1. Що таке варіаційний ряд? Які є типи варіаційного ряду?
2. Що таке середнє арифметичне значення? Які властивості має середнє арифметичне?
3. Що таке середнє квадратичне відхилення? Які властивості притаманні с.к.в.?
4. Що таке медіана? Які переваги структурних середніх перед ступеневими?

ПРАКТИЧНА РОБОТА № 2

Тема: Нормальний розподіл та його властивості
Мета: Набути навичок перевірки відповідності розподілу вибірки нормальному закону

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ

Нормальний розподіл (або *розподіл Гауса-Лапласа*) має сукупність, розподіл варіант якої відповідають формулі:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} \ . \qquad (2.1)$$

Для перевірки відповідності нормальному закону розподілу вибірки, що задана у вигляді варіаційного ряду необхідно виконати наступні розрахунки. Розглянемо цю методику на підставі варіаційного ряду, що було нами отримано у Практичній роботі № 1.

1. Необхідно побудувати нову таблицю, що повинна мати наступні дані щодо варіаційного ряду (табл. 2.1).

2. Для середини кожного інтервалу необхідно розрахувати його нормоване відхилення за формулою:

$$t_i = \frac{x_i - \overline{x}}{\sigma}. \tag{2.2}$$

Таблиця 2.1

i	*Інтервал*	n_i	s_i	x_i					
1	3,20-3,40	4	4	3,30					
2	3,41-3,60	21	25	3,50					
3	3,61-3,80	40	65	3,70					
4	3,81-4,00	30	95	3,90					
5	4,01-4,20	15	110	4,10					
6	4,21-4,40	0	110	4,30					
7	4,41-4,60	3	113	4,50					
Всього		113	-	-					

Наприклад, для середини першого інтервалу нормоване відхилення має значення: $t_1 = \frac{3{,}30 - 3{,}78}{0{,}22} = -2{,}182$.

Заповнюємо таблицю 2.1 проміжними даними:

i	*Інтервал*	n_i	s_i	x_i	t_i				
1	3,20-3,40	4	4	3,30	-2,182				
2	3,41-3,60	21	25	3,50	-1,273				
3	3,61-3,80	40	65	3,70	-0,364				
4	3,81-4,00	30	95	3,90	0,545				
5	4,01-4,20	15	110	4,10	1,455				
6	4,21-4,40	0	110	4,30	2,364				
7	4,41-4,60	3	113	4,50	3,273				
Всього		113	-	-	-				

3. Далі для значень нормованого відхилення необхідно визначити ординати нормальної кривої (теоретичні). Це можна зробити двома способами. По-перше, знайти відповідні значення у таблиці ординат

нормальної нормованої кривої, що наведена у Додатку А. (Тут необхідно враховувати, що *f(-t) = f(t)*, тобто знак, що знаходиться перед значенням нормованого відхилення можна не враховувати.)

По-друге, ординати нормальної кривої можна розрахувати на підставі формули:

$$f(t) \approx 0{,}4 \cdot e^{-\frac{t^2}{2}} . \tag{2.3}$$

Всі отримані значення заносимо у відповідний стовпчик табл. 2.1.

i	*Інтервал*	n_i	s_i	x_i	t_i	*f(t)*			
1	3,20-3,40	4	4	3,30	-2,182	0,0370			
2	3,41-3,60	21	25	3,50	-1,273	0,1780			
3	3,61-3,80	40	65	3,70	-0,364	0,3744			
4	3,81-4,00	30	95	3,90	0,545	0,3447			
5	4,01-4,20	15	110	4,10	1,455	0,1389			
6	4,21-4,40	0	110	4,30	2,364	0,0245			
7	4,41-4,60	3	113	4,50	3,273	0,0019			
Всього		113	-	-	-	1,0993			

Правильність розрахунків можна перевірити наступним чином. Сума значень ординат повинна дорівнювати:

$$\sum f(t) = \frac{\sigma}{h}, \tag{2.4}$$

де h – ширина класового інтервалу варіаційного ряду (див. Практичну роботу № 1). В нашому випадку: $\frac{\sigma}{h} = \frac{0{,}22}{0{,}2} = 1{,}10$, що з точністю до третього знаку після коми відповідає сумі значень ординат нормальної кривої (1,0993).

4. Розраховуються теоретичні значення частот кожного інтервалу за формулою:

$$n^{teo} = f(t) \cdot \frac{nh}{\sigma} \tag{2.5}$$

Наприклад, теоретична частота для першого інтервалу складатиме: $n^{teo} = 0{,}0370 \cdot \frac{113 \cdot 0{,}20}{0{,}22} = 3{,}80$.

Після розрахунку всіх теоретичних частот табл. 2.1 прийме вигляд:

i	*Інтервал*	n_i	s_i	x_i	t_i	*f(t)*	n^{teo}	s^{teo}	
1	3,20-3,40	4	4	3,30	-2,182	0,0370	3,80	3,80	
2	3,41-3,60	21	25	3,50	-1,273	0,1780	18,28	22,08	
3	3,61-3,80	40	65	3,70	-0,364	0,3744	38,46	60,55	
4	3,81-4,00	30	95	3,90	0,545	0,3447	35,41	95,96	
5	4,01-4,20	15	110	4,10	1,455	0,1389	17,27	110,22	
6	4,21-4,40	0	110	4,30	2,364	0,0245	2,52	112,74	
7	4,41-4,60	3	113	4,50	3,273	0,0019	0,19	112,93	
Всього		113	-	-	-	1,0993	112,93	-	

В ідеалі сума фактичних частот повинна дорівнювати сумі теоретичних, тобто $\Sigma n_i = \Sigma n^{teo}$. (В нашому випадку різниця у другому знаку після коми, що викликане округленням значень при проведенні проміжних розрахунків.)

Наступний стовпчик – це накопичені теоретичні значення (див. Практичну роботу № 1).

5. Графічним зображенням відповідності розподілу вибірки нормальному закону є гістограма (рис. 2.1).

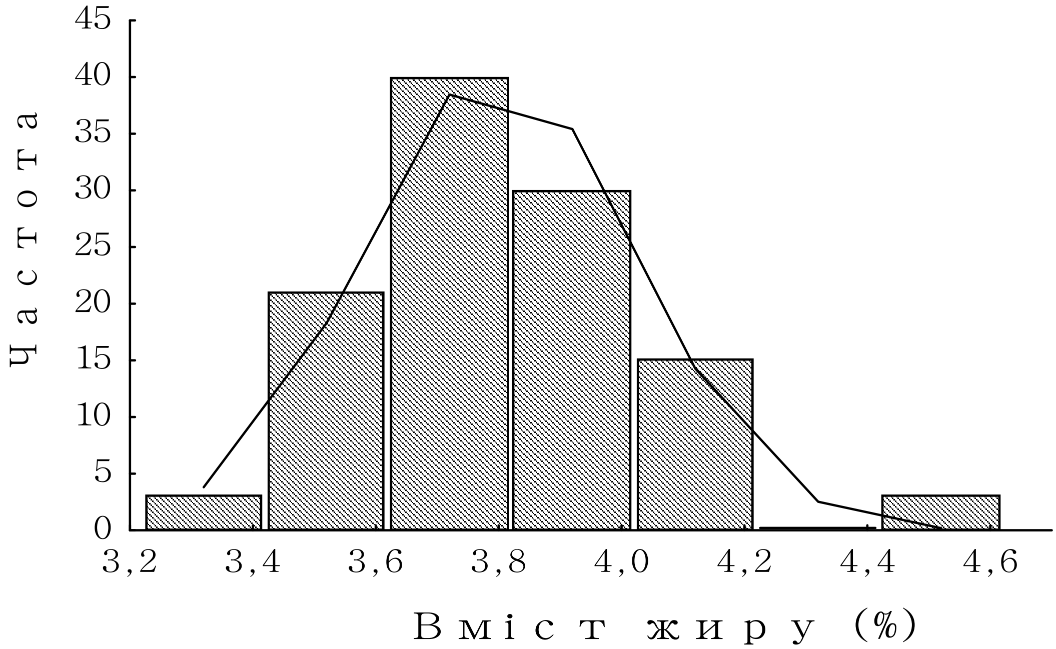


Рис. 2.1. Гістограма розподілу за вмістом жиру матерів корів дослідного стада. Наведена теоретична лінія нормального розподілу

Крім того, відповідність вибіркового розподілу нормальному закону можна перевірити на підставі *кумуляти* (рис. 2.2).

6. Для перевірки відповідності вибіркового розподілу нормальному закону можна використовувати критерій *Колмогорова-Смирнова*, який представляє собою модуль максимальної різниці між фактичними та теоретичними накопиченими частотами:

$$D = \max | s_i - s^{teo} |. \quad (2.6)$$

Для розрахунку критерію Колмогорова-Смирнова необхідно в межах кожного інтервалу розрахувати модуль різниці між фактичними та теоретичними накопиченими частотами. Наприклад, для першого інтервалу це значення складатиме: $D = \max | 4 – 3,80 | = 0,20$. Тоді табл. 2.1 буде мати вигляд:

i	*Інтервал*	n_i	s_i	x_i	t_i	$f(t)$	n^{teo}	s^{teo}	D
1	3,20-3,40	4	4	3,30	-2,182	0,0370	3,80	3,80	0,20
2	3,41-3,60	21	25	3,50	-1,273	0,1780	18,28	22,08	2,92
3	3,61-3,80	40	65	3,70	-0,364	0,3744	38,46	60,55	**4,45**
4	3,81-4,00	30	95	3,90	0,545	0,3447	35,41	95,96	0,96
5	4,01-4,20	15	110	4,10	1,455	0,1389	17,27	110,22	0,22
6	4,21-4,40	0	110	4,30	2,364	0,0245	2,52	112,74	2,74
7	4,41-4,60	3	113	4,50	3,273	0,0019	0,19	112,93	0,07
Всього		113	-	-	-	1,0993	112,93	-	-

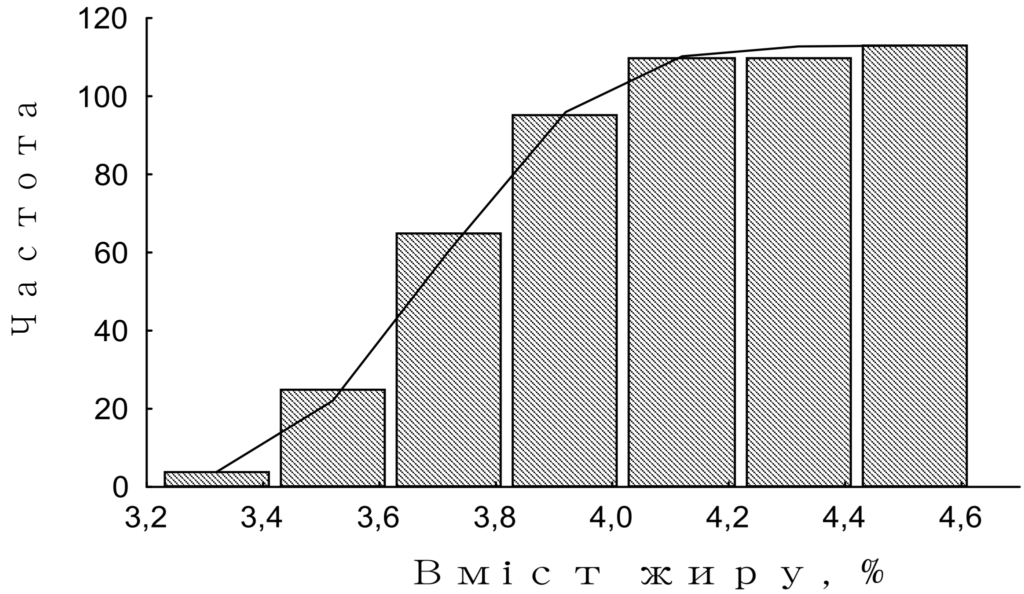


Рис. 2.2. Кумулята розподілу за вмістом жиру матерів корів дослідного стада. Наведена теоретична лінія нормального розподілу

Знаходимо максимальне значення – в нашому випадку воно дорівнює 4,45 (за модулем) для третього інтервалу.

Оцінка вірогідності отриманого значення критерію Колмогорова-Смирнова перевіряється на підставі значення λ:

$$\lambda = \frac{D}{\sqrt{n}}. \qquad (2.7)$$

Якщо розраховане значення λ не перевищує 1,36, вибірковий розподіл відповідає нормальному закону; якщо λ більше, ніж 1,36 – вибірковий розподіл вірогідно відхиляється від нормального.

Для нашого приклада: $\lambda = \frac{4,45}{\sqrt{113}} = 0,42$; таким чином розподіл матерів корів за вмістом жиру відповідає нормальному закону.

7. Крім того, відповідність вибіркового розподілу нормальному закону можна перевірити, розрахувавши показники форми кривої розподілу, а саме – коефіцієнт асиметрії (*As*) та коефіцієнт ексцесу (*Ex*):

$$As = \frac{\sum (x_i - \bar{x})^3}{n\sigma^3}, \qquad (2.8)$$

$$Ex = \frac{\sum (x_i - \bar{x})^4}{n\sigma^4} - 3. \qquad (2.9)$$

Перший з них дає оцінку скошеності кривої розподілу відповідно його центру, а другий визначає наявність піку або плато у центрі розподілу.

Формули 2.8 та 2.9 використовуються для малих вибірок; якщо ми маємо справу із варіаційним рядом оцінки коефіцієнтів асиметрії та ексцесу розраховуються наступним чином.

Будуються нова таблиця, яка має такі стовпчики (табл. 2.2).

Таблиця 2.2

i	*Інтервал*	n_i	a	$n_i \cdot a$	$n_i \cdot a^2$	$n_i \cdot a^3$	$n_i \cdot a^4$
1	3,20-3,40	4	-2	-8	16	-32	64
2	3,41-3,60	21	-1	-21	21	-21	21
3	3,61-3,80	40	0	0	0	0	0
4	3,81-4,00	30	+1	30	30	30	30
5	4,01-4,20	15	+2	30	60	120	240
6	4,21-4,40	0	+3	0	0	0	0
7	4,41-4,60	3	+4	12	48	192	768
Суми		113	-	43	175	289	1123

Показник *a* беруть таким чином, що для інтервального класу с найбільшою частотою він дорівнює нулю, для значень, що менше модального класу – зменшуються на одиницю, а для значень, що більше модального класу – збільшуються на одиницю. (До речі, вибір значень *a* не грає суттєвої ролі.)

На підставі розрахованих сум чотирьох останніх стовпчиків розраховуються проміжні величини:

$$m_1 = \frac{\sum n_i a}{n}, \tag{2.10}$$

$$m_2 = \frac{\sum n_i a^2}{n}, \tag{2.11}$$

$$m_3 = \frac{\sum n_i a^3}{n}, \tag{2.12}$$

$$m_4 = \frac{\sum n_i a^4}{n}. \tag{2.13}$$

Тоді оцінки коефіцієнтів асиметрії та ексцесу можна отримати за формулами:

$$As = \frac{m_3 - 3m_1 m_2 + 2m_1^3}{\left(\sqrt{m_2 - m_1^2}\right)^3}, \tag{2.14}$$

та

$$Ex = \frac{m_4 - 4m_1 m_3 + 6m_1^2 m_2 - 3m_1^4}{\left(\sqrt{m_2 - m_1^2}\right)^4} - 3. \tag{2.15}$$

Для даних із нашого прикладу: $m_1 = \frac{43}{113} = 0{,}381$; $m_2 = \frac{175}{113} = 1{,}549$; $m_3 = \frac{289}{113} = 2{,}558$; $m_4 = \frac{1123}{113} = 9{,}938$. Тоді значення коефіцієнту асиметрії складатиме:

$$As = \frac{2{,}558 - 3 \cdot 0{,}381 \cdot 1{,}549 + 2 \cdot 0{,}381^3}{\left(\sqrt{1{,}549 - 0{,}381^2}\right)^3} = 0{,}541,$$

а коефіцієнта ексцесу:

$$Ex = \frac{9{,}938 - 4 \cdot 0{,}381 \cdot 2{,}558 + 6 \cdot 0{,}381^2 \cdot 1{,}549 - 3 \cdot 0{,}381^4}{\left(\sqrt{1{,}549 - 0{,}381^2}\right)^4} - 3 = 0{,}718.$$

Оцінка вірогідності отриманих значень проводиться на підставі значень їх статистичних помилок. Для великих вибірок статистичні помилки коефіцієнтів асиметрії та ексцесу розраховуються за формулами:

$$S_{As} = \sqrt{\frac{6}{n+3}}; \qquad (2.16)$$

$$S_{Ex} = \sqrt{\frac{24}{n+5}}. \qquad (2.17)$$

Для даних із нашого прикладу ці значення складають: S_{As} = 0,227; S_{Ex} = 0,451.

Якщо відношення коефіцієнту асиметрії чи ексцесу до своїх помилок перевищуватиме 3, то вважається, що вибірковий розподіл вірогідно відхиляється від нормального закону. В інших випадках – вибірковий розподіл відповідає нормальному закону.

В нашому випадку відношення коефіцієнта асиметрії до своєї помилки складає: $\frac{As}{S_{As}} = \frac{0{,}541}{0{,}227} = 2{,}38$, а відношення коефіцієнта ексцесу до своєї помилки: $\frac{Ex}{S_{Ex}} = \frac{0{,}718}{0{,}451} = 1{,}59$. В обох випадках ці значення не перевищують 3; таким чином, ми ще раз підтвердили, що розподіл матерів корів за вмістом жиру в молоці вірогідно не відхиляється від нормального.

►**<u>Завдання</u>**: 1. Перевірити відповідність розподілу нормальному закону для показнику, вказаному викладачем.

2. Розрахувати для цього ж показника оцінки коефіцієнтів асиметрії та ексцесу. ◄

Література. **1**: с.31-36, с.57-60; **3**: с.18-21, с.59-62, с.90-91; **4**: с.141-150, с.217-231.

Питання для самоперевірки:

1. Які властивості має нормальний розподіл?
2. Що таке коефіцієнт асиметрії та які він має властивості?
3. Що таке коефіцієнт ексцесу та які він має властивості?
4. Для чого використовується критерій Колмогорова-Смирнова?

ПРАКТИЧНА РОБОТА № 3

Тема: Інтервальні оцінки вибіркових показників

Мета: Набути навичок розрахунку довірчих інтервалів для вибіркових показників

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ

Інтервальною називають оцінку, яка характеризується двома числами – границями інтервалу, який охоплює оцінюваний параметр. Така оцінка являє собою деякий інтервал, в якому із заданою ймовірністю знаходиться шуканий параметр. За центр інтервалу беруть вибіркову точкову оцінку.

Довірча ймовірність – це достатньо висока ймовірність практично вірогідній події, яка гарантує отримання надійний статистичних висновків. Вона означається P, а ймовірність перевищити цей рівень – α. Відповідно, $\alpha = 1 - P$. Ймовірність α називають *рівнем значимості*, або *рівням істотності*, який характеризує відносну кількість помилкових висновків від загальної кількості усіх статистичних висновків.

Довірчим інтервалом для параметра θ називається такий інтервал, відносно якого можна із близькою до одиниці довірчою ймовірністю $P = 1 - \alpha$ стверджувати, що він містить шукане значення параметра θ.

Для більшості вибіркових показників інтервальне оцінювання безпосередньо пов'язано із розрахунками їх відповідних статистичних помилок. Проілюструємо методику розрахунку довірчих інтервалів для показника вмісту жиру в молоці матерів корів дослідного стада (див. Практичну роботу № 1).

1. Розрахунок довірчого інтервалу для вибіркового середнього арифметичного.

Статистична помилка вибіркового середнього арифметичного розраховується за формулою:

$$S_{\bar{x}} = \frac{\sigma}{\sqrt{n}}, \tag{3.1}$$

де σ – вибіркове середнє квадратичне відхилення (с.к.в.); n – обсяг вибірки.

Таким чином, для даних з нашого прикладу помилка середнього арифметичного складатиме: $S_{\bar{x}} = \frac{0{,}22}{\sqrt{113}} = 0{,}02$.

Тоді із 95 % довірчою ймовірністю значення вибіркового середнього арифметичного знаходиться у інтервалі:

$$\bar{x} - t \cdot S_{\bar{x}} \le \bar{x} \le \bar{x} + t \cdot S_{\bar{x}}, \tag{3.2}$$

де t – значення критерію Ст'юдента для числа ступенів свободи $df = n - 1$ (табл. 3.1).

Обсяг вибірки складає 113 голів. Тоді число ступенів свободи для критерію Ст'юдента складатиме $df = 113 - 1 = 112$. В табл. 3.1 немає точного значення для такого числа ступенів свободи. Але, оскільки, для 110 та 120 ступенів свободи табличне значення не змінюється (1,98), його можна прийняти за шукане. Таким чином, для нашого прикладу 95 % довірчий інтервал для вибіркового середнього арифметичного складатиме:

$$3{,}78 - 1{,}98 \cdot 0{,}02 \le \bar{x} \le 3{,}78 + 1{,}98 \cdot 0{,}02,$$

або

$$3{,}74 \le \bar{x} \le 3{,}82.$$

Таблиця 3.1

Критичні точки розподілу Ст'юдента
(для α = 0,05)

df	*t*	*df*	*t*	*df*	*t*
1	12,71	16	2,12	40	2,02
2	4,30	17	2,11	50	2,01
3	3,18	18	2,10	60	2,00
4	2,78	19	2,09	70	1,99
5	2,57	20	2,09	80	1,99
6	2,45	21	2,08	90	1,99
7	2,36	22	2,07	100	1,98
8	2,31	23	2,07	110	1,98
9	2,26	24	2,06	120	1,98
10	2,23	25	2,06	130	1,98
11	2,20	26	2,06	140	1,98
12	2,18	27	2,05	150	1,98
13	2,16	28	2,05	200	1,97
14	2,14	29	2,05	500	1,96
15	2,13	30	2,04	∞	1,96

В тих випадках, коли обсяг генеральної сукупності з якої відібрана вибірка має фіксоване значення (*N*), в формулу 3.1 необхідно вносити виправлення:

$$S_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{1 - \frac{n}{N}}\ . \tag{3.3}$$

2. <u>Розрахунок довірчого інтервалу для вибіркового с.к.в</u>.

Статистична помилка вибіркового с.к.в. розраховується за формулою:

$$S_{\sigma} = \frac{\sigma}{\sqrt{2n}}. \tag{3.4}$$

Таким чином, для даних з нашого прикладу помилка вибіркового с.к.в. складатиме: $S_{\sigma} = \dfrac{0{,}22}{\sqrt{2 \cdot 113}} = 0{,}015.$

Але при побудові довірчого інтервалу для вибіркового с.к.в. використовується не розподіл Ст'юдента, а розподіл Хі-квадрат Пірсона. З 95 % довірчою ймовірністю можна стверджувати, що істинне значення вибіркового с.к.в. лежить у межах:

$$\sigma_н = \sqrt{\frac{(n-1)\cdot\sigma^2}{\chi_1^2}}; \qquad (3.5)$$

$$\sigma_в = \sqrt{\frac{(n-1)\cdot\sigma^2}{\chi_2^2}}, \qquad (3.6)$$

де χ_1^2 - значення критерію Хі-квадрат Пірсона для числа ступенів свободи $df = n - 1$ та рівня значимості $\alpha = 0{,}025$; χ_2^2 - значення критерію Хі-квадрат Пірсона для числа ступенів свободи $df = n - 1$ та рівня значимості $\alpha = 0{,}975$.

Для зручності розрахунки довірчого інтервалу для вибіркового с.к.в. можна проводити за формулами:

$$\sigma_н = c_1 \cdot \sigma; \qquad (3.7)$$

$$\sigma_в = c_2 \cdot \sigma, \qquad (3.8)$$

де коефіцієнти c_1 та c_2 знаходяться в табл. 3.2.

В тих випадках, коли число ступенів свободи для вибірки відсутнє в табл. 3.2 шукані значення коефіцієнтів c_1 та c_2 можна отримати лінійною інтерполяцією.

Наприклад, в Практичній роботі № 1 ми отримали вибіркове значення с.к.в. 0,22. Як розрахувати 95 % довірчий інтервал цього показника?

Число ступенів свободи для вибірки складає $df = 113 - 1 = 112$. В табл. 3.2 найближчі значення – це 110 та 120. Тоді методом лінійної інтерполяції коефіцієнти для розрахунків нижньої та верхньої довірчих меж для вибіркового с.к.в. можна розрахувати за формулою:

$$c = c' + \frac{(c'' - c')\cdot(df - df')}{(df'' - df')},$$

де df' – табличне число ступенів свободи, що менше вибіркового, а c' – відповідне значення допоміжного коефіцієнту; df'' – табличне число ступенів свободи, що більше вибіркового, а c'' – відповідне значення допоміжного коефіцієнту.

Нижче вибіркового числа ступенів свободи в таблиці наведені значення для $df = 110$, а вище - $df = 120$. Тоді на підставі лінійної інтерполяції можна визначити значення допоміжних коефіцієнтів:

$$c_1 = 0,884 + \frac{(0,888 - 0,884) \cdot (112 - 110)}{120 - 110} = 0,885;$$

$$c_2 = 1,152 + \frac{(1,145 - 1,152) \cdot (112 - 110)}{120 - 110} = 1,151.$$

Таблиця 3.2

Коефіцієнти для розрахунку довірчого інтервалу для вибіркового с.к.в.

df	c_1	c_2	*df*	c_1	c_2
20	0,765	1,444	120	0,888	1,145
30	0,799	1,337	130	0,892	1,138
40	0,821	1,280	140	0,895	1,133
50	0,837	1,243	150	0,899	1,128
60	0,849	1,217	160	0,901	1,123
70	0,858	1,198	170	0,904	1,119
80	0,866	1,183	180	0,907	1,115
90	0,873	1,171	190	0,909	1,112
100	0,879	1,161	200	0,911	1,109
110	0,884	1,152	250	0,920	1,096

Відповідно нижня та верхня межі для вибіркового с.к.в. складатиме:

$$\sigma_н = 0,885 \cdot 0,22 = 0,195;$$

$$\sigma_в = 1,151 \cdot 0,22 = 0,253.$$

3. Розрахунок довірчого інтервалу для вибіркової медіани.

95% довірчий інтервал для вибіркової медіани можна також розрахувати за формулою 3.2, де замість вибіркового середнього арифметичного та його помилки підставляється вибіркова медіана та її помилка, відповідно:

$$Me - t \cdot S_{Me} \le Me \le Me + t \cdot S_{Me}, \qquad (3.9)$$

де статистична помилка вибіркової медіани розраховується за формулою:

$$S_{Me} = \frac{Q_3 - Q_1}{\sqrt{n}}. \qquad (3.10)$$

В цій формулі використовуються значення першого (Q_1) та третього квартилів (Q_3). *Квартилі* розподіляють весь вибірковий ранжований ряд на чотири рівні частини. Таким чином, 25 % вибіркових варіант мають значення нижчі, ніж перша квартиль, 50 % значень знаходяться між першою та третьою квартиллю, і ще 25 % значень переважають третю квартиль. (Медіана вважається другою квартиллю.)

Значення першої та третьої квартилів для даних, що згруповані у варіаційний ряд можна розрахувати за формулами:

$$Q_1 = x_{Q_1} + h \cdot \left(\frac{\frac{n}{4} - s_{Q_1 - 1}}{n_{Q_1}} \right) \tag{3.11}$$

та

$$Q_3 = x_{Q_3} + h \cdot \left(\frac{\frac{3 \cdot n}{4} - s_{Q_3 - 1}}{n_{Q_3}} \right), \tag{3.12}$$

де x_{Q_1} та x_{Q_3} - нижні межі інтервалів, що містять першу та треті квартилі, відповідно; $s_{Q_1 - 1}$ та $s_{Q_3 - 1}$ - накопичені частоти інтервалів, що передують інтервалам, що містять першу та третю квартилі, відповідно; n_{Q_1} та n_{Q_3} - частоти інтервалів, що містять першу та третю квартилі, відповідно.

Оскільки у вибірці 113 значень нижче першої квартилі знаходяться (113/4) = 28,25 варіант. Таким чином перша квартиль знаходиться у інтервалі, накопичена частота якої вперше перважає 28,25; це інтервал 3,61-3,80. Менше третьої квартилі знаходиться (3·113/4) = 84,75 варіант. Таким чином, третя квартиль знаходиться у інтервалі, накопичена частота якої вперше переважає 84,75; це інтервал 3,81-4,00. Таким чином значення першої і третьої квартилів будуть дорівнювати:

$$Q_1 = 3{,}61 + 0{,}2 \cdot \left(\frac{\frac{113}{4} - 25}{40} \right) = 3{,}63,$$

$$Q_3 = 3{,}81 + 0{,}2 \cdot \left(\frac{\frac{3 \cdot 113}{4} - 65}{30} \right) = 3{,}94.$$

Таким чином, помилка вибіркової медіани дорівнюватиме:

$$S_{Me} = \frac{3{,}94 - 3{,}63}{\sqrt{113}} = 0{,}03.$$

Тоді з 95 % ймовірністю можна стверджувати, що вибіркова медіана знаходиться у межах:

$$3{,}77 - 1{,}98 \cdot 0{,}03 \le Me \le 3{,}77 + 1{,}98 \cdot 0{,}03,$$

або

$$3{,}71 \le Me \le 3{,}83.$$

►**<u>Завдання</u>**: Розрахувати 95 % довірчі інтервали для середнього арифметичного, с.к.в. та медіани показника, вказаного викладачем. ◄

Література. **1**: с.42-44; **2**: с.27-29; **3**: с.14-16, с.57-59; **4**: 202-213.

Питання для самоперевірки:

1. Що таке довірчий інтервал оцінки?
2. Що визначає рівень значимості?
3. Які особливості має розподіл Ст'юдента?
4. Які особливості побудови довірчого інтервалу для с.к.в.?
5. Що таке перший та третій квартилі?

ПРАКТИЧНА РОБОТА № 4

Тема: Критерії Ст'юдента
Мета: Набути навичок перевірки статистичних гіпотез відносно центру розподілу

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ

Статистичною гіпотезою (H) називається припущення щодо параметрів чи форми розподілу генеральної сукупності (чи сукупностей), що перевіряється на основі даних вибіркового розподілу (чи сукупностей). З параметрів розподілу найчастіше виникає необхідність перевірити чи рівність середніх арифметичних у двох вибірках (перевірка у відношенні центральної тенденції), чи вибіркових варіанс (перевірка у відношенні рівня мінливості). Дійсна лабораторна робота буде присвячена задачам першого типу.

1. <u>Розрахунок критерію Ст'юдента у випадку залежних вибірок</u>

Перевірка нульової гіпотези в тому випадку, коли вибірка залежні проводиться, використовуючи наступну формулу критерію Ст'юдента:

$$t = \frac{\left|\bar{d}\right|}{\sqrt{\dfrac{\sum d_i^2 - n\bar{d}^2}{n(n-1)}}}, \tag{4.1}$$

де n – обсяг вибірок, а $d_i = x_i - y_i$. Табличні значення критерію Ст'юдента наведено в табл. 3.1.

Приклад 1. Було визначено зміст жиру в молоці в 10 корів та їхніх дочок. Чи можна сказати, що дочки більш жирномолочні, ніж їх матері?

Усі вихідні дані і проміжні величини приведені в табл. 4.1.

Таблиця 4.1

№ п/п	Зміст жиру в молоці матерів, % (x)	Зміст жиру в молоці дочок, % (y)	d_i	d_i^2
1	3,6	3,5	0,1	0,01
2	3,3	3,6	-0,3	0,09
3	3,7	3,6	0,1	0,01
4	3,5	3,8	-0,3	0,09
5	3,0	3,5	-0,5	0,25
6	3,1	3,4	-0,3	0,09
7	4,0	3,9	0,1	0,01
8	3,4	3,7	-0,3	0,09
9	3,8	4,0	-0,2	0,04
10	3,7	4,1	-0,4	0,16
Суми	35,1	37,1	-2,0	0,84

Середня різниця тоді буде дорівнює $\bar{d} = \frac{-2,0}{10} = -0,2$. Отже, оцінка критерію Ст'юдента буде складати:

$$t = \frac{0,2}{\sqrt{\frac{0,84 - 10 \cdot 0,2^2}{10 \cdot 9}}} = 2,86 \quad .$$

Число ступенів свободи дорівнює $df = 10 - 1 = 9$. Табличні значення для цього числа ступенів свободи складають $t = 2,26$. Оскільки розраховане значення критерію Ст'юдента переважає табличне, дочки виявляються вірогідно більш жирномолочними, чим матері.

2. <u>Розрахунок критерію Ст'юдента у випадку незалежних вибірок</u>

У цьому випадку вибір формули для критерію Ст'юдента залежить від рівності варіанс двох вибіркових сукупностей. Ця перевірка проводиться, використовуючи F-критерій Фішера-Снедекора :

$$F = \frac{\sigma_1^2}{\sigma_2^2} . \tag{4.2}$$

Значення критерію Фішера-Снедекора завжди перевищує одиницю, оскільки являє собою відношення більшої варінси до меншого. Отримане таким чином значення критерію порівнюється з табличним для відповідних чисел ступенів волі: $df_1 = n_1 - 1$ і $df_2 = n_2 - 1$. Відповідно, n_1 і n_2 – обсяги порівнюваних вибірок. Однак, якщо обсяги вибірок досить великі (більш 20-25 варіант), рівень значимості F-критерію можна оцінити, використовуючи u-апроксимацію:

$$u = \frac{(1-k_2)\cdot\sqrt[3]{F} - (1-k_1)}{\sqrt{k_2\sqrt[3]{F^2} + k_1}}, \tag{4.3}$$

де $k_1 = 2/(9\cdot df_1)$, а $k_2 = 2/(9\cdot df_2)$. Критичне значення u-критерію, при якому визнається достовірним розходження між двома порівнюваними вибірковими варіансами, складає 1,645.

При рівності варіанс, значення критерію Ст’юдента розраховується за формулою:

$$t = \frac{|\bar{x} - \bar{y}|\sqrt{\dfrac{n_x \cdot n_y}{n_x + n_y}}}{\sqrt{\dfrac{\sigma_x^2(n_x - 1) + \sigma_y^2(n_y - 1)}{n_x + n_y - 2}}}. \tag{4.4}$$

Це вираження значне спрощується, якщо обсяги вибірок рівні між собою:

$$t = \frac{|\bar{x} - \bar{y}|}{\sqrt{\dfrac{\sigma_x^2 + \sigma_y^2}{n}}}. \tag{4.5}$$

Розраховане значення критерію порівнюється з табличним (табл. 3.1) для числа ступенів свободи: $df = n_x + n_y - 2$. (У випадку рівності обсягів вибірок ця формула перетвориться у $df = 2\cdot(n - 1)$).

Приклад 2. В двох групах корів червоної степової породи була проведена оцінка жирномолочності. У першій групі, що містила 50 тварин, середній зміст жиру в молоці склав 3,0 % із середнім квадратичним відхиленням – 0,15 %. В другій групі, що містила 75 корів, середній зміст жиру в молоці складало 3,8 % із середнім квадратичним відхиленням – 0,18 %.

Чи можна стверджувати, що обидві групи вірогідно розрізняються за рівнем жирномолочності?

Спочатку, використовуючи критерій Фішера-Снедекора, з'ясуємо, вірогідно розрізняються варіанси порівнюваних вибірок чи ні.

Підставляємо вихідні значення у формулу 4.2, причому в чисельнику записуємо значення варіанси для другої вибірки (оскільки воно більше):

$$F = \frac{0{,}18^2}{0{,}15^2} = 1{,}44.$$

Оскільки обсяги вибірок великі, оцінку значимості цієї величини можна розрахувати, використовуючи апроксимацію по формулі 4.3:

$$u = \frac{(1-0{,}0045)\cdot\sqrt[3]{1{,}44} - (1-0{,}003)}{\sqrt{0{,}0045\cdot\sqrt[3]{1{,}44^2} + 0{,}003}} = 1{,}36 \;.$$

Ця величина не перевищує граничне (тобто 1,645), тому робиться висновок, що оцінки варіанс у порівнюваних вибірках вірогідно не відрізняються, отже, можна використовувати формулу 4.4 для розрахунку величини критерію Ст'юдента:

$$t = \frac{|3{,}0-3{,}8|\sqrt{\dfrac{50\cdot 75}{50+75}}}{\sqrt{\dfrac{0{,}15^2\cdot(50-1)+0{,}18^2\cdot(75-1)}{50+75-2}}} = 26{,}3.$$

Число ступенів свободи для цього приклада складає *df* = 50 + 75 – 2 = 123.

Безпосередня інтерпретація отриманого значення критерію Ст'юдента не представляється можливим, оскільки в табл.3.1 не наведено табличне значення для числа ступенів свободи рівному 123. Але як можна помітити, при збільшенні числа ступенів волі від 120 до 130 табличні значення критерію Ст'юдента не змінюються (1,98), тому можна вважати, що розраховане значення критерію Ст'юдента набагато перевищує табличне. Таким чином, можна затверджувати, що рівень жирномолочності корів другої групи вірогідно вище, ніж у корів з першої групи.

У тому випадку, якщо оцінки вибіркових варіанс вірогідно відрізняються між собою, оцінку вірогідності розходжень між двома середніми проводиться на підставі формули:

$$t = \frac{|\bar{x} - \bar{y}|}{\sqrt{\dfrac{\sigma_x^2}{n_x} + \dfrac{\sigma_y^2}{n_y}}}\,. \qquad (4.6)$$

Тоді розраховане значення необхідно порівняти із табличним значенням критерію Ст'юдента із числом ступенів свободи:

$$df = (n_x + n_y - 2)\cdot\left(\frac{1}{2} + \frac{\sigma_x^2\sigma_y^2}{\sigma_x^4 + \sigma_y^4}\right). \qquad (4.7)$$

Приклад 3. Середня жива маса корів герефордської породи в череді чисельністю 150 голів (при аутбридинге) складала 600 кг із середнім квадратичним відхиленням 20 кг, а середня живаючи маса корів цієї ж породи в іншій череді чисельністю 200 голів (при тісному інбридингу) – 580 кг із середнім квадратичним відхиленням 30 кг.

Чи впливає тип схрещування на живу масу корів?

Спочатку знову, використовуючи критерій Фішера-Снедекора, з'ясуємо, чи вірогідно розрізняються варіанси порівнюваних вибірок чи ні. Підставляємо вихідні значення у формулу 4.2:

$$F = \frac{30^2}{20^2} = 2{,}25.$$

Розраховане по формулі 4.3 значення u-критерію для даного значення критерію Фішера-Снедекора складає 5,10, що набагато перевищує критичне (1,645); таким чином, вибіркові варіанси розрізняються вірогідно і для розрахунку оцінки критерію Ст’юдента потрібно використовувати формулу 4.6:

$$t = \frac{|600 - 580|}{\sqrt{\frac{20^2}{150} + \frac{30^2}{200}}} = 7{,}46.$$

Використовуючи формулу 4.7, розрахуємо число ступенів свободи для даного значення критерію Ст’юдента складає:

$$df = (150 + 200 - 2) \cdot \left(\frac{1}{2} + \frac{20^2 \cdot 30^2}{20^4 + 30^4} \right) \approx 303.$$

У табл. 3.1 не приведено критичних значень для такого значення числа ступенів волі; однак при *df* > 200 можна в якості табличних використовувати значення 1,96. Оскільки розраховане значення критерію Ст’юдента набагато перевищує табличне, можна зробити висновок, що тісний інбридинг приводить до достовірного зниження живої маси корів.

►**Завдання**:1. Визначте, чи вірогідно розрізняється вміст жиру в молоці 20 корів та їх матерів. (Значення “*i”* – це цифра, яка має вигляд “0,0А”, де *А* – остання цифра залікової книжки. Наприклад, якщо остання цифра залікової книжки дорівнює 5, то *i* = 0,05. Тоді для першої дочки в таблиці необхідно брати не 3,77, а 3,77 + 0,05 = 3,82; а для другої матері замість 4,06 необхідно брати 4,06 - 0,05 = 4,01.)

№ п/п	Вміст жиру в молоці матерів (%)	Вміст жиру в молоці дочок (%)
1	3,72	3,77 + *i*
2	4,06 – *i*	3,84
3	4,12 – *i*	3,77
4	3,93	3,84 + *i*
5	3,57	3,81
6	4,19 – *i*	3,98 – *i*
7	3,51 + *i*	3,91 – *i*
8	3,70	3,73
9	4,04	4,22 – *i*
10	3,77 + *i*	4,08
11	3,68	3,60 + *i*
12	3,52 + *i*	3,79
13	3,80	3,72 + *i*
14	3,97 – *i*	3,59 + *i*
15	3,80	3,87
16	3,85	4,28 – *i*
17	3,77 + *i*	3,72
18	3,73	3,69 + *i*
19	3,91 – *i*	3,82
20	3,80	3,76 + *i*

2. Розрахуйте вірогідність відмінностей між цими двома рядами значень, вважаючи їх незалежними вибірками. ◄

Література. **1**: с.50-53, с.60-72; **2**: с.32-38, с.41-44; **3**: с.78-87; **4**: с.253-260.

Питання для самоперевірки:

1. Що таке статистична гіпотеза?
2. Від чого залежить вид формули критерію Ст'юдента?
3. Що таке залежні і незалежні ознаки?
4. Для чого використовується критерій Фішера-Снедекора?

ПРАКТИЧНА РОБОТА № 5

Тема: Кореляція
Мета: Набути навичок оцінки вибіркового коефіцієнта кореляції

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ

Функціональним називається такий зв'язок між двома ознаками, коли кожному значенню одної змінної відповідає строго визначене значення іншої. Але такий тип зв'язку в медико-біологічних дослідженнях майже ніколи не зустрічається. В зоотехнії найчастіше має місце такий тип зв'язку між ознаками, коли чисельному значенню однієї з них відповідає декілька значень іншої. Такий зв'язок має назву *кореляційного* або *статистичного*.

Кореляційний зв'язок є неповним, він має місце лише при великій кількості спостережень. За допомогою метода кореляційного аналізу можна вирішити дві задачі:

- виміряти тісноту зв'язку;
- визначити форму та параметри рівняння зв'язку.

Перша задача рішається на підставі оцінки вибіркового значення коефіцієнта кореляції (власне кореляційний аналіз), а друга – при використанні методів регресійного аналізу (див. Практичну роботу № 6).

В залежності від набору вихідних даних у вибірці можна визначити наступні типи коефіцієнту кореляції:

- *коефіцієнт парної лінійної кореляції* Пірсона-Браве визначає силу й напрямок лінійного зв'язку між двома ознаками;
- *коефіцієнт множинної кореляції* визначає силу зв'язку між залежною змінною й набором незалежних змінних на підставі рівняння лінійної множинної регресії;
- *коефіцієнт часткової кореляції* визначає силу й напрямок лінійного зв'язку між залежною ознакою й будь-якою з набору незалежних при допущенні, що вплив інших незалежних ознак відсутній.

1. Розрахунок коефіцієнту парної лінійної кореляції Пірсона-Браве.

Оцінку вибіркового коефіцієнту парної лінійної кореляції Пірсона-Браве можна отримати за формулою:

$$r = \frac{n \cdot \sum xy - \sum x \cdot \sum y}{\sqrt{[n \cdot \sum y^2 - (\sum y)^2] \cdot [n \cdot \sum x^2 - (\sum x)^2]}}. \qquad (5.1)$$

Для зручності, всі проміжні розрахунки краще проводити у табличному вигляді.

Наприклад, розрахуємо коефіцієнт парної лінійної кореляції між показниками вмісту жиру у корів та їх матерів (n = 15). Всі необхідні попередні розрахунки наведено в табл. 5.1.

Таким чином, вибіркове значення коефіцієнта кореляції для даних тварин складатиме:

$$r = \frac{15 \cdot 217{,}235 - 56{,}75 \cdot 57{,}38}{\sqrt{[15 \cdot 220{,}161 - (57{,}38)^2] \cdot [15 \cdot 215{,}235 - (56{,}75)^2]}} = 0{,}318.$$

Таблиця 5.1

№ п/п	Вміст жиру в молоці матерів, % (x)	Вміст жиру в молоці корів, % (y)	x^2	$x \cdot y$	y^2
1	3,77	3,72	14,213	14,024	13,838
2	3,84	4,06	14,746	15,590	16,484
3	4,05	4,12	16,403	16,686	16,974
4	3,84	3,93	14,746	15,091	15,445
5	3,81	3,57	14,516	13,602	12,745
6	3,98	4,19	15,840	16,676	17,556
7	3,70	3,51	13,690	12,987	12,320
8	3,73	3,70	13,913	13,801	13,690
9	3,40	4,04	16,484	16,402	16,322
10	3,80	3,77	14,440	14,326	14,213
11	3,60	3,68	12,960	13,248	13,542
12	3,79	3,52	14,364	13,341	12,390
13	3,72	3,80	13,838	14,136	14,440
14	3,85	3,97	14,823	15,285	15,761
15	3,87	3,8	14,977	14,706	14,440
Суми	*56,75*	*57,38*	*215,028*	*217,235*	*220,161*

Статистична помилка коефіцієнта кореляції розраховується за формулою:

$$SE_r = \sqrt{\frac{1 - r^2}{n - 2}}\ . \tag{5.2}$$

Вірогідність відхилення отриманого вибіркового значення коефіцієнту кореляції від нуля можна оцінити на підставі критерію Ст'юдента:

$$t = \frac{r}{SE_r}\ . \tag{5.3}$$

Статистичну значущість коефіцієнта кореляції можна вважати доведеною, якщо отримане значення критерію Ст'юдента перевищує табличне (див. табл. 3.1) для числа ступенів свободи: $df = n - 2$.

Для даних з нашого прикладу статистична помилка коефіцієнта кореляції складає:

$$SE_r = \sqrt{\frac{1 - 0{,}318^2}{15 - 2}} = 0{,}263.$$

Тоді оцінка критерію Ст'юдента буде дорівнювати: $t = \frac{0{,}318}{0{,}263} = 1{,}21$, що набагато менше, ніж табличне значення критерію Ст'юдента для 13 ступенів свободи (2,18). Таким чином, можна зробити висновок, що вибіркове значення коефіцієнту кореляції вірогідно не відхиляється від нуля, а тому зв'язок між вмістом жиру в молоці корів та їх матерів відсутній.

Для зручності в таблиці 5.2 наведено критичні значення коефіцієнту кореляції для різних рівнів ступенів свободи.

Таким чином, для того, щоб зв'язок між вмістом жиру в молоці корів та їх матерів вважати статистично доведеним, значення коефіцієнту кореляції між ними повинно складати якнайменше 0,514.

Таблиця 5.2

df	$r_{\text{крит}}$	*df*	$r_{\text{крит}}$	*df*	$r_{\text{крит}}$	*df*	$r_{\text{крит}}$
10	0,576	19	0,433	28	0,361	80	0,217
11	0,553	20	0,423	29	0,365	90	0,205
12	0,532	21	0,413	30	0,349	100	0,195
13	0,514	22	0,404	35	0,325	125	0,174
14	0,497	23	0,396	40	0,304	150	0,159
15	0,482	24	0,388	45	0,288	200	0,138
16	0,468	25	0,381	50	0,273	300	0,113
17	0,456	26	0,374	60	0,250	400	0,098
18	0,444	27	0,367	70	0,232	500	0,088

В загальному вигляді для будь-якого числа ступенів свободи критичне значення коефіцієнта кореляції можна розрахувати за формулою:

$$r_{крит} = \frac{t}{\sqrt{df + t^2}}, \tag{5.4}$$

де відповідні значення критерію Ст'юдента необхідно брати з таблиці 3.1. Наприклад, якщо у обсяг вибірки складає 113 особин, то відповідне табличне значення критерію Ст'юдента складає 1,98, тоді критичне значення коефіцієнту кореляції дорівнюватиме:

$$r_{крит} = \frac{1{,}98}{\sqrt{(113-2)+1{,}98^2}} = 0{,}185.$$

2. <u>Оцінка довірчого інтервалу для коефіцієнта кореляції</u>.

Розподіл вибіркових коефіцієнтів кореляції відповідає нормальному лише поблизу нульового значення, тому для побудови довірчого інтервалу вибіркового коефіцієнта кореляції необхідно використовувати його відповідну z-трансформацію:

$$z = \frac{1}{2}\ln\frac{1+r}{1-r}, \tag{5.5}$$

статистична помилка якої не залежить від значення коефіцієнту кореляції й залежить тільки від обсягу вибірки:

$$SE_z = \frac{1}{\sqrt{n-3}}. \tag{5.6}$$

Оскільки значення z вже мають нормальний розподіл, 95 % довірчий інтервал для цього показника можна побудувати на підставі формули:

$$z - \frac{1{,}96}{\sqrt{n-3}} \le z \le z + \frac{1{,}96}{\sqrt{n-3}}. \tag{5.7}$$

Зворотний перехід від значень z до r проводиться за формулою:

$$r = \frac{e^{2\cdot z} - 1}{e^{2\cdot z} + 1}. \tag{5.8}$$

Для зручності наводимо таблицю переводу значень коефіцієнта кореляції в значення z та навпроти (табл. 5.3).

Таблиця 5.3

r	Значення z для сотих часток r									
	0	1	2	3	4	5	6	7	8	9
0,0	0,000	0,010	0,020	0,030	0,040	0,050	0,060	0,070	0,080	0,090
0,1	0,100	0,110	0,121	0,131	0,141	0,151	0,161	0,172	0,182	0,192
0,2	0,203	0,213	0,224	0,234	0,245	0,255	0,266	0,277	0,288	0,299
0,3	0,310	0,321	0,332	0,343	0,354	0,365	0,377	0,388	0,400	0,412
0,4	0,424	0,436	0,448	0,460	0,472	0,485	0,497	0,510	0,523	0,536
0,5	0,549	0,563	0,576	0,590	0,604	0,618	0,633	0,648	0,662	0,678
0,6	0,693	0,709	0,725	0,741	0,758	0,775	0,793	0,811	0,829	0,848
0,7	0,867	0,887	0,908	0,929	0,950	0,973	0,996	1,020	1,045	1,071
0,8	1,099	1,127	1,157	1,188	1,221	1,256	1,293	1,333	1,376	1,422
0,9	1,472	1,528	1,589	1,658	1,738	1,832	1,946	2,092	2,298	2,647

Наприклад, оцінимо довірчий інтервал для оцінки коефіцієнта кореляції, отриманої між вмістом жиру в молоці корів та їх матерів. На цьому прикладі продемонструємо методику використання табл. 5.3.

Як дано за умовами, коефіцієнт кореляції дорівнює r = 0,318. В табл. 5.3 таке значення відсутнє. Обираємо такі два найближчі значення r, для яких є табличні значення z, так щоб $r_0 < r < r_1$. Таким чином, r_0 = 0,31 та r_1 = 0,32. Відповідні табличні значення для них складають: z_0 = 0,321 та z_1 = 0,332.

Тоді, використав формулу лінійної інтерполяції, шукане значення можна отримати за формулою:

$$z = \pm [z_0 + u \cdot (z_1 - z_0)], \quad (5.9)$$

де $u = \dfrac{r - r_0}{r_1 - r_0}$. Знак перед значенням z такий же, який має коефіцієнт корреляції.

Для нашого прикладу допоміжна величина дорівнює: $u = \dfrac{0{,}318 - 0{,}31}{0{,}32 - 0{,}31} = 0{,}8$. Підставляємо отримане значення в формулу 5.9, оцінюємо значення $z = 0{,}321 + 0{,}8 \cdot (0{,}332 - 0{,}321) = 0{,}330$.

Далі, розрахуємо довірчий інтервал для отриманої оцінки за формулою 5.7. Тоді нижня межа 95 % довірчого інтервалу дорівнюватиме:

$$z_H = 0{,}330 - \frac{1{,}96}{\sqrt{15-3}} = -0{,}236,$$

а верхня:

$$z_B = 0{,}330 + \frac{1{,}96}{\sqrt{15-3}} = +0{,}896.$$

Для того, щоб від оцінок z зараз повернутися до оцінок r знову використаємо табл. 5.3.

В табл. 5.3 немає такого значення r, для якого $z = 0{,}236$ (знак поки не враховуємо). Обираємо два найближчі табличні значення z таким чином, щоб $z_0 < z < z_1$. Тоді, $z_0 = 0{,}234$ та $z_1 = 0{,}245$. Відповідні значення r для них складають: $r_0 = 0{,}23$ та $r_1 = 0{,}24$.

Знову використав лінійну інтерполяцію, шукане значення можна отримати за формулою:

$$r = \pm (r_0 + 0{,}01 \cdot v), \quad (5.10)$$

де $v = \dfrac{z - z_0}{z_1 - z_0}$. Знак перед значенням коефіцієнта кореляції ставиться такий, який має значення z.

Підставляємо наші дані та отримуємо значення допоміжної величини: $v = \dfrac{0{,}236 - 0{,}234}{0{,}245 - 0{,}234} = 0{,}182$. Тоді шукана оцінка нижньої довірчої межі коефіцієнта кореляції складає: $r_H = -(0{,}23 + 0{,}01 \cdot 0{,}182) = -0{,}232$. (Оскільки відповідне значення z мало від'ємний знак, від'ємний знак також ставимо й перед значенням r.)

Використав аналогічну процедуру, визначаємо верхню межу 95 % довірчого інтервалу складає $r_B = 0{,}714$. Таким чином, з ймовірністю 95 % можна стверджувати, що генеральне значення коефіцієнта кореляції між вмістом жиру в молоці корів та їх матерів знаходиться у межах:

$$-0{,}232 \le r \le 0{,}714.$$

(Оскільки нульове значення опинилося в межах довірчого інтервалу, знову підтверджується наш висновок, щодо відсутності кореляційного зв'язку між вмістом жиру в молоці корів та їх матерів.).

►**Завдання**: 1. Розрахувати вибіркове значення коефіцієнта кореляції між вмістом жиру та надоєм корів:

Надій, кг (x)	Вміст жиру, % (y)
6111 + i	3,70
4818 + i	4,11
3719 - i	4,31
5200 - i	3,90
5524 + i	4,07
5634 + i	3,94
4020 + i	4,03
4476 - i	4,20
6479 - i	3,94
3657 + i	4,40
3666 - i	4,17
5064 + i	3,60
4431 - i	3,86
5215 + i	3,80
4017 - i	3,79

де i – останні дві цифри залікової книжки.

2. Розрахувати межі 95 % довірчого інтервалу для коефіцієнта кореляції та оцінити його статистичну значущість. ◄

Література. **1**: с.83-93; **2**: с.45-48; **3**: с.143-149; **4**: с.333-346.

Питання для самоперевірки:

1. Які є типи зв'язку між ознаками?
2. Які є типи коефіцієнтів кореляції?
3. Що таке z-трансформація й для чого вона використовується?

ПРАКТИЧНА РОБОТА № 6

Тема: Проста лінійна регресія
Мета: Набути навичок оцінки вибіркових коефіцієнтів простої лінійної регресії

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ

Регресія – це лінія, вид залежності середньої величини результативної ознаки від факторної. З погляду математики регресія являє собою функцію $y = f(x_1, x_2, \ldots, x_n)$, яка описує залежність умовного математичного очікування залежної змінної (y) від заданих фіксованих значень незалежних значень ($x_1, x_2, \ldots, x_n$).

Рівняння регресії – аналітичне рішення, за допомогою якого відображується зв'язок між досліджуваними ознаками. Розрізняють *прямолінійне* рівняння зв'язку (пряма лінія) і *криволінійне* (парабола, гіпербола, експонента, логарифмічна крива тощо).

При прямолінійній залежності однієї ознаки від іншої рівняння регресії має вигляд:

$$y = a + b \cdot x, \tag{6.1}$$

де b – коефіцієнт лінійної регресії, тангенс кута нахилу лінії регресії відносно вісі ОХ; a – точка перетину лініює регресії вісі ОУ.

Коефіцієнт лінійної регресії визначає середній рівень зміни залежної ознаки при зміні незалежної змінної на одиницю.

Визначення коефіцієнтів лінійної регресії відбувається із застосуванням методу найменших квадратів (МНК), розроблений Гаусом ще у XIX сторіччі. Відповідно до цього методу, коефіцієнти лінійної регресії є рішенням системи лінійних рівнянь:

$$\begin{cases} a \cdot n + b \cdot \sum x = \sum y; \\ a \cdot \sum x + b \cdot \sum x^2 = \sum xy. \end{cases} \tag{6.2}$$

Параметри a та b рівняння лінійної регресії можна визначити по іншим робочим формулам:

$$b = \frac{n \cdot \sum xy - \sum x \cdot \sum y}{n \cdot \sum x^2 - \left(\sum x\right)^2}; \tag{6.3}$$

$$a = \frac{\sum y \cdot \sum x^2 - \sum xy \cdot \sum x}{n \cdot \sum x^2 - \left(\sum x\right)^2}. \tag{6.4}$$

Як бачимо, для розрахунку коефіцієнтів парної лінійної регресії необхідні теж самі проміжні розрахунки, що й при розрахунку коефіцієнту парної регресії. Тому при визначення форми аналітичної залежності жирномолочності дочок від вмісту жиру в молоці їх матерів (Практична робота № 5) використовуємо дані, що отримані у табл. 5.1.

Таблиця 6.1

№ п/п	Вміст жиру в молоці матерів, % (x)	Вміст жиру в молоці корів, % (y)	x^2	$x \cdot y$	y^2
1	3,77	3,72	14,213	14,024	13,838
2	3,84	4,06	14,746	15,590	16,484
3	4,05	4,12	16,403	16,686	16,974
4	3,84	3,93	14,746	15,091	15,445
5	3,81	3,57	14,516	13,602	12,745
6	3,98	4,19	15,840	16,676	17,556
7	3,70	3,51	13,690	12,987	12,320
8	3,73	3,70	13,913	13,801	13,690
9	3,40	4,04	16,484	16,402	16,322
10	3,80	3,77	14,440	14,326	14,213
11	3,60	3,68	12,960	13,248	13,542
12	3,79	3,52	14,364	13,341	12,390
13	3,72	3,80	13,838	14,136	14,440
14	3,85	3,97	14,823	15,285	15,761
15	3,87	3,8	14,977	14,706	14,440
Суми	*56,75*	*57,38*	*215,028*	*217,235*	*220,161*

Тоді, коефіцієнти парної лінійної регресії дорівнюватимуть:

$$b = \frac{15 \cdot 217{,}235 - 56{,}75 \cdot 57{,}38}{15 \cdot 215{,}028 - (56{,}75)^2} = 0{,}455$$

та

$$a = \frac{57{,}38 \cdot 215{,}028 - 217{,}235 \cdot 56{,}75}{15 \cdot 215{,}028 - (56{,}75)^2} = 2{,}104 .$$

Таким чином, можна зробити висновок, що при зміні вмісту жиру в молоці корів-матерів на 1 % у їх дочок рівень жирномолочності підвищується на 0,455 %. Аналітична форма зв’язку має наступний вигляд:

$$y = 2{,}104 + 0{,}455 \cdot x.$$

Коефіцієнти парної лінійної регресії, як всі вибіркові показники, мають визначений рівень випадковості, який визначається їх статистичними помилками:

$$SE_b = \sqrt{\frac{\sigma^2_{зал}}{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}}, \tag{6.5}$$

$$SE_a = SE_b \cdot \sqrt{\frac{\sum x^2}{n}}, \tag{6.6}$$

де $\sigma^2_{зал}$ - варіанса залишок, яка визначається за формулою:

$$\sigma^2_{зал} = \frac{\sum y^2 - a \cdot \sum y - b \cdot \sum xy}{n-2}. \tag{6.7}$$

Перевірка рівня вірогідності відмінності кожного з коефіцієнтів парної лінійної кореляції від нуля відбувається із застосуванням критерію Ст'юдента:

$$t_b = \frac{b}{SE_b}, \tag{6.8}$$

$$t_a = \frac{a}{SE_a}. \tag{6.9}$$

Коефіцієнти парної лінійної регресії вважаються значущими, якщо відповідні їм значення критерію Ст'юдента перевищують табличні значення (табл. 3.1) для числа ступенів свободи $df = n - 2$.

Таким чином, для розрахунку оцінок статистичних помилок коефіцієнтів парної лінійної регресії по-перше необхідно розрахувати значення варіанси залишок:

$$\sigma^2_{зал} = \frac{220{,}161 - 2{,}104 \cdot 57{,}38 - 0{,}455 \cdot 217{,}235}{15-2} = 0{,}046\,.$$

Тоді оцінки помилок коефіцієнтів регресії становитимуть відповідно:

$$SE_b = \sqrt{\frac{0{,}046}{215{,}028 - \frac{(56{,}75)^2}{15}}} = 0{,}377$$

та

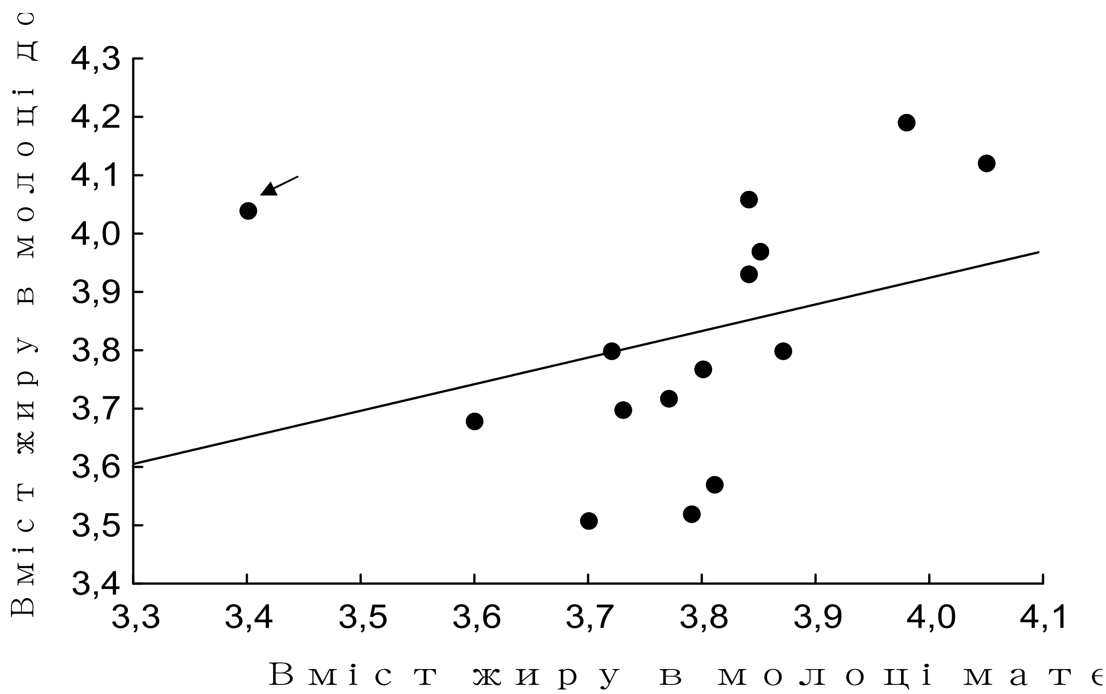
$$SE_a = 0{,}377 \cdot \sqrt{\frac{215{,}028}{15}} = 1{,}427\,.$$

Таким чином відповідні значення критерію Ст'юдента дорівнюють:

$$t_b = \frac{0{,}455}{0{,}377} = 1{,}21; \quad t_a = \frac{2{,}104}{1{,}427} = 1{,}47.$$

В табл.3.1 знаходимо критичні значення критерію Ст'юдента для числа ступенів свободи $df = 15 - 2 = 13$; воно дорівнює 2,16. Таким чином ані для коефіцієнту *a*, ані для коефіцієнту *b* рівняння лінійної регресії розрахункові значення критерію Ст'юдента не перевищують табличне значення. Ще раз підтвердився висновок, отриманий в Практичній роботі № 5 про відсутність зв'язку між жирномолочністю корів та їх матерів.

В графічній формі ця залежність має вигляд:



Графічний аналіз дозволяє візуалізовати залежність, яка аналізується й краще з'ясувати характер цього зв'язку. Це стосується насамперед виявленню "викидів", тобто, значень, які значно відхиляються від загальної залежності, але впливають на хід розрахунків. В нашому випадку точка, яка на графіку позначена стрілкою, значно відхиляється від інших пар "матері-дочки". Якщо її видалити, зв'язок між вмістом жиру в молоці матерів та їх дочок стане вірогідним.

►**<u>Завдання</u>**: 1. Розрахувати вибіркові значення коефіцієнта парної лінійної кореляції між вмістом жиру та надоєм корів із Практичної роботи № 5.

2. Визначити рівень значущості отриманих значень коефіцієнтів лінійної регресії. ◄

Література. **1**: с.93-100; **2**: с.59-62; **3**: с.140-143; **4**: с.316-319, с.337-346.

Питання для самоперевірки:

1. Для чого використовується регресійний аналіз?
2. Які є форми ліній регресії?

3. Хто й коли розробив метод найменших квадратів для оцінці коефіцієнтів регресії?
4. Що визначають коефіцієнти a та b лінії регресії?

ПРАКТИЧНА РОБОТА № 7

Тема: Аналіз якісних ознак
Мета: Набути навичок розрахунку вибіркових показників для якісних ознак

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ

Фенами називаються будь-які дискретні альтернативні варіації ознак та властивостей особин. Фени завжди відбивають генетичну конституцію даної особини, а своєю частотою – генетичну структуру популяції чи інших груп особин даного виду.

Фенетика популяцій – це поширення генетичних підходів і принципів на види і форми, власне генетичне вивчення яких утруднено чи неможливо.

Предмет фенетики – внутрішньовидова мінливість, яка в остаточному підсумку доводиться до розглядання дискретних, альтернативних ознак-маркерів генотипового складу популяції (тобто фенів).

Методи фенетики полягають у вичленовуванні різних фенів, характерних для мінливості досліджуваних форм, їх кількісне та якісне вивчення.

Теоретичною основою фенетичного підходу є правило гомологічних рядів у спадковій мінливості, що сформульоване М.І.Вавиловим у 1920 р. Він показав, що генетично та філогенетичне близькі форми характеризуються подібними рядами спадкоємної мінливості.

1. Оцінка частот фенів та побудова її довірчого інтервалу.

Якщо у вибірці обсягом n відзначено n_A особин, що мають фен A, то частота його в даній вибірці складає:

$$p_A = \frac{n_A}{n}. \tag{7.1}$$

Статистична помилка цієї величини:

$$SE_{p_A} = \sqrt{\frac{p_A(1-p_A)}{n}}. \tag{7.2}$$

Якщо обсяг вибірки невеликий, то величина помилки може зрівнятися з оцінкою частоти. Практично можна мати справа з оцінками лише в тих випадках, якщо SE_{PA} не менше, чим у 3 рази менше p_A чи $1 - p_A$ (у залежності від величини p_A — менше вона 0,5 чи більше).

Якщо у вибірці зустрічається m фенів (варіацій) однієї ознаки, то оцінки їхніх частот і помилок обчислюються за аналогічними формулаами. Ці формули придатні, якщо частота будь-якого фена лежить у границях від 0,2 до 0,8.

Величина помилки SE_{PA} часто використовується для розрахунку довірчого інтервалу для визначеного значення частоти: $p \pm 1{,}96 \cdot SE_{PA}$.

У випадках, якщо оцінки частот дуже великі чи дуже маленькі (менш 0,2 чи більш 0,8), для більш точної оцінки частоти фена та її статистичної помилки необхідно користуватися φ-перетворенням Р.Фішеру:

$$\varphi_A = 2 \cdot \arcsin \sqrt{\frac{n_A}{n}}, \tag{7.3}$$

$$SE\,\varphi_A = \frac{1}{\sqrt{n}}. \tag{7.4}$$

Приклад 1. У вибірці обсягом 78 особин виявлено 13 тварин, які мають фен *А*. Визначити частоту зустрічальності цього фена та його статистичну помилку.

Частота зустрічальності фена *А* в даній вибірці складає:

$$p_A = 13:78 = 0{,}167,$$

а її помилка:

$$SE_{p_A} = \sqrt{\frac{0{,}167 \cdot (1 - 0{,}167)}{78}} = 0{,}042\ .$$

Довірчий інтервал для даної оцінки складає від: 0,167 - 1,96·0,042 до 0,167 + 1,96·0,042, тобто від 0,085 до 0,249.

Однак, оскільки оцінка частоти фена *А* близька до 0, більш коректно використовувати φ-перетворення. Тоді:

$$\varphi_A = 2 \cdot \arcsin \sqrt{\frac{13}{78}} = 0{,}840\ ;$$

$$SE\,\varphi_A = \sqrt{\frac{1}{78}} = 0{,}113\ .$$

Довірчі інтервали:

$$\varphi_{A'} = 0{,}840 - 1{,}96 \cdot 0{,}113 = 0{,}619;$$
$$\varphi_{A''} = 0{,}840 + 1{,}96 \cdot 0{,}113 = 1{,}061.$$

Для того щоб повернутися від φ-перетворення до оцінок частот, використовується формула:

$$p = \sin^2 \frac{\varphi}{2}. \qquad (7.5)$$

Таким чином, для даного прикладу 95 % довірчий інтервал частоти фену *A* буде складати: $0{,}093 \le p_A \le 0{,}256$.

Для зручності в таблиці 7.1 наведені значення перетворення Р.Фішеру.

Таблиця 7.1

	0	1	2	3	4	5	6	7	8	9
0,00	0,000	0,063	0,089	0,110	0,127	0,142	0,155	0,168	0,179	0,190
0,0	0,000	0,200	0,284	0,348	0,403	0,451	0,495	0,536	0,574	0,609
0,1	0,644	0,676	0,707	0,738	0,767	0,795	0,823	0,850	0,876	0,902
0,2	0,927	0,953	0,976	1,000	1,024	1,047	1,070	1,093	1,115	1,137
0,3	1,159	1,181	1,203	1,224	1,245	1,266	1,287	1,308	1,328	1,349
0,4	1,369	1,390	1,410	1,430	1,451	1,471	1,491	1,511	1,531	1,551
0,5	1,571	1,591	1,611	1,631	1,651	1,671	1,691	1,711	1,731	1,752
0,6	1,772	1,793	1,813	1,834	1,855	1,875	1,897	1,918	1,939	1,961
0,7	1,982	2,004	2,026	2,049	2,071	2,094	2,118	2,141	2,165	2,190
0,8	2,214	2,240	2,265	2,292	2,319	2,346	2,375	2,404	2,434	2,465
0,9	2,498	2,532	2,568	2,606	2,647	2,691	2,739	2,793	2,858	2,941
0,99	2,941	2,952	2,963	2,974	2,987	3,000	3,015	3,032	3,052	3,078

Значення φ для вибіркових оцінок частот, які відсутні в табл. 7.1 можна розрахувати за допомогою лінійної інтерполяції за формулою:

$$\varphi = \varphi_н + u\cdot(\varphi_в - \varphi_н), \qquad (7.6)$$

де $u = \dfrac{p - p_н}{p_в - p_н}$; $p_н$ та $p_в$ – найближчі до вибіркової оцінки значення частот, для яких в таблиці 7.1 наведені відповідні значення арксинус-перетворення Р.Фішеру – $\varphi_н$ та $\varphi_в$, відповідно.

Наприклад, для значення $p_A = 0{,}167$ із прикладу 1 в таблиці 7.1 немає точного значення. Найближчі до нього $p_н = 0{,}16$ (для нього $\varphi_н = 0{,}823$) та $p_в = 0{,}17$ (для нього $\varphi_в = 0{,}850$). Тоді, використовуємо формулу 7.6, отримуємо значення: $\varphi = 0{,}823 + \left(\dfrac{0{,}167 - 0{,}16}{0{,}17 - 0{,}16}\right)\cdot(0{,}850 - 0{,}823) = 0{,}842$, які відрізняється від точного у третьому знаку після коми.

Для зворотного перетворення використовується формула:

$$p = p_н + v\cdot(p_в - p_н), \qquad (7.7)$$

де $v = \dfrac{\varphi - \varphi_н}{\varphi_в - \varphi_н}$.

Наприклад, для значення $\varphi_{A'}$ = 0,619 найближчі табличні складають $\varphi_н$ = 0,609 (для нього $p_н$ = 0,09) та $\varphi_в$ = 0,644 (для нього $p_в$ = 0,10). Тоді шукане значення нижньої межі 95 % довірчого інтервалу складатиме:

$p' = 0{,}09 + \left(\dfrac{0{,}619 - 0{,}609}{0{,}644 - 0{,}609}\right) \cdot (0{,}10 - 0{,}09) = 0{,}093$, яке не відрізняється від розрахованого за формулою 7.5. (Розраховане за формулою 7.7 верхня межа 95 % довірчого інтервалу тоді складатиме 0,254, а точне значення 0,256.)

2. Оцінки фенетичного розмаїття.

Мірою внутрішньопопуляційного фенетичного розмаїття у відношенні якісних ознак, які мають m різних альтернативних варіацій (тобто фенів), є **середнє число фенів** (μ):

$$\mu = \left(\sum_{i=1}^{m} \sqrt{p_i}\right)^2, \tag{7.8}$$

тобто квадрат суми квадратних коренів вибіркових оцінок частот фенів. Він показує, скільки у вибірці різних фенотипів (з урахуванням їх частот зустрічальності).

При рівності частот усіх фенів, що зустрічаються у вибірці, цей показник прагне до m; при нерівномірному розподілі частот — $\mu < m$; якщо популяція є *мономорфною*(тобто зустрічається лише один фен) — $\mu = 1$.

Помилка середнього числа фенів розраховується за формулою:

$$SE_\mu = \sqrt{\frac{\mu(m-\mu)}{n}}. \tag{7.9}$$

На основі середнього числа фенів обчислюється й інший показник – **частка рідких морф** (h_μ) **та** її статистична помилка:

$$h_\mu = 1 - \frac{\mu}{m}, \tag{7.10}$$

$$SE_{h_\mu} = \sqrt{\frac{h_\mu(1-h_\mu)}{n}}. \tag{7.11}$$

Якщо розподіл частот морф у вибірці має рівномірний характер, то $h_\mu \approx 0$; при нерівномірності ж розподілу частот — $h_\mu > 0$.

Таким чином, μ оцінює ступінь розмаїття, а показник h_μ дає визначену характеристику цього розмаїття в розумінні співвідношення між частотами найменш численних та найбільш численних у цій вибірці (популяції) фенів.

Приклад 2. У череді великої рогатої худоби симентальської породи племінного заводу "Дубов'язівський" виявлено 10 фенів типу чола первісток (за 1976-1990 рр.); чисельність тварин з різним феном чола наведено в таблиці 7.2.

Необхідно оцінити рівень внутрішньопопуляційного розмаїття корів у відношенні даної якісної ознаки.

Як видно з таблиці 7.2, чисельності і, відповідно, частоти зустрічальності особин з тим чи іншим феном значно коливаються. Найбільш розповсюдженим є «подовжений» фен чола корів симентальської породи в даному племзаводі.

Для того щоб оцінити показники розмаїття (середнє число фенів і частку рідких морф), необхідно попередньо розрахувати частоту зустрічальності кожного фена (третій стовпець таблиці) і корені квадратні з цих частот (четвертий стовпець). Оцінка середнього числа фенів тоді розраховується як квадрат суми цих величин:

$$\mu = (0{,}356 + 0{,}313 + ... + 0{,}149)^2 = 3{,}095^2 = 9{,}579.$$

Помилка цього показника складає:

$$SE_\mu = \sqrt{\frac{9{,}58 \cdot (10 - 9{,}58)}{6723}} = 0{,}024.$$

Таблиця 7.2

Фени типу чола (i)	Кількість особин (n_i)	Частота фена (p_i)	$\sqrt{p_i}$
Широкий	852	0,127	0,356
Вузький	660	0,098	0,313
Плоский	880	0,131	0,362
Опуклий	758	0,113	0,336
Чотирикутний	410	0,061	0,247
Гостровершинний	700	0,104	0,323
Подовжений	1022	0,152	0,390
Високий	683	0,102	0,319
Середній	608	0,090	0,301
Низький	150	0,022	0,149
Разом	6723	1,000	3,095

Як бачимо, розраховане значення середнього числа фенів дуже близько до числа виділених фенів і, отже, популяція у відношенні аналізованого ознака високо поліморфна.

Про це ж свідчить і низьке значення частки рідких морф:

$$h_{\mu} = 1 - \frac{9{,}58}{10} = 0{,}042,$$

$$SEh_{\mu} = \sqrt{\frac{0{,}042 \cdot (1 - 0{,}042)}{6723}} = 0{,}002.$$

3. Порівняння двох вибірок у відношенні частот фенів.

Якщо оцінки частот зустрічальності аналізованого фена знаходяться в межах від 0,2 до 0,8, а обсяги вибірок досить великі (кілька сотень особин), то для їхнього порівняння використовується стандартний *t*-критерій Ст'юдента, де вибіркові середні арифметичні заміняються оцінками частот ознаки в двох вибірках, а помилки середніх арифметичних — помилками оцінок частот:

$$t = \frac{|p_1 - p_2|}{\sqrt{SEp_1{}^2 + SEp_2{}^2}}. \tag{7.12}$$

Розраховане значення критерію Ст'юдента порівнюється з табличним для числа ступенів свободи $df = n_1 + n_2 - 2$, де n_1 і n_2 — обсяги першої і другої вибірок, відповідно. Якщо $t > t_{крит}$, то частота зустрічальності ознаки у вибірці 1 вірогідно відрізняється від частоти зустрічальності цієї ж ознаки у вибірці 2. Табличні значення критерію Ст'юдента наведені в табл.3.1.

Якщо оцінки частот близькі до 1 чи 0, то використовується φ-перетворення (див. вище) і тоді оцінкою вірогідності розбіжностей між двома вибірками служить величина u-критерію:

$$u = \frac{|\varphi_1 - \varphi_2|}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \tag{7.13}$$

де φ_1 і φ_2 — φ-перетворені значення частот аналізованої ознаки у вибірках обсягом n_1 і n_2. Якщо $u > 1{,}96$, то вибірки вірогідно розрізняються за частотою зустрічальності даної ознаки.

Приклад 3. У першій вибірці, що містить 25 особин великої рогатої худоби виявлено п'ять, що мають конічні дійки, а в другій вибірці обсягом 15 особин лише дві.

Чи вірогідно розрізняється частота зустрічальності корів з конічними дійками в цих двох вибірках?

Оскільки обсяги вибірок мали, використовуємо формулу 7.13. Частота зустрічальності корів з конічним типом дійок у першій вибірці складає 5:25 = 0,20, відповідно $\varphi_1 = 0{,}927$; частота зустрічальності тварин з аналогічною ознакою в другій вибірці — 2:15 = 0,133, і, відповідно φ_2 = 0,747.

Тоді значення u-критерію складає $u = \dfrac{0{,}927 - 0{,}747}{\sqrt{\dfrac{1}{25} + \dfrac{1}{15}}} = 0{,}55$, що набагато нижче критичного. Таким чином, нуль-гіпотеза щодо рівної частоти зустрічальності особин з даним типом дійок в двох вибірках підтверджується.

4. Асоціація ознак.

Для статистичного підтвердження наявності зв'язку між двома якісними ознаками критерій Хі-квадрат Пірсона:

$$\chi^2 = \sum \frac{(\Phi - T)^2}{T}. \qquad (7.14)$$

Оцінка вірогідності відхилення фактичних частот сполучень ознак від теоретичних проводиться на підставі рівня значимості критерію Хі-квадрат при відповідному числі ступенів свободи $df = (v - 1) \cdot (c - 1)$, де v – кількість стовпців таблиці сполученості, а c – кількість її рядків.

Критичні значення критерію Хі-квадрат для різного числа ступенів свободи наведено в таблиці 7.3.

Таблиця 7.3

df	1	2	3	4	5	6	7	8	9	10
χ^2	3,84	5,99	7,82	9,49	11,07	12,59	14,07	15,51	16,92	18,31
df	11	12	13	14	15	16	17	18	19	20
χ^2	19,68	21,03	22,36	23,69	25,00	26,30	27,59	28,87	30,14	31,41
df	21	22	23	24	25	26	27	28	29	30
χ^2	32,67	33,92	35,17	36,42	37,65	38,89	40,11	41,34	42,56	43,77

Приклад 4. При аналізі прояву асоціації у відношенні фенів крупа та постановки хвоста у 3300 корів симентальської породи, що належать племзаводові «Терезино» (1975-1985 рр.) виявлені наступні абсолютні частоти зустрічальності різних пар сполучень цих двох ознак (табл. 7.4).

Таблиця 7.4

Фен постановки хвоста	Фен крупа		Разом
	прямий	скошений	
Високопоставлений	700 / *474,0*	1140 / *1366,0*	1840
Низькопоставлений	150 / *376, 0*	1310 / *1084,0*	1460
Разом	850	2450	3300

Необхідно з'ясувати, чи має місце зв'язок між цими двома ознаками, тобто чи незалежно зустрічаються різні фени крупа і постановки хвоста у тварин даного підприємства? Якщо ні, то які сполучення більш ймовірні?

Для цього необхідно розрахувати теоретично очікувані абсолютні частоти всіх чотирьох можливих сполучень двох ознак і порівняти їх з фактичними з використанням критерію Хі-квадрат. Для розрахунку теоретичних частот можна користатися простим правилом: теоретична абсолютна частота по якому-небудь осередку таблиці спряженості дорівнює добутку суми частот по даному стовпці і суми частот по даному рядку, діленому на обсяг вибірки. Причому це правило справедливе для будь-яких таблиць сполученості, незалежно від кількості рядків та стовпців.

Таким чином, розраховуємо теоретичні частоти для всіх чотирьох пар сполучень:

для сполучення «прямий круп – високопоставлений хвіст»:

$$n_{ПВ} = \frac{850 \cdot 1840}{3300} = 474{,}0 \ ,$$

для сполучення «прямий круп – низькопоставлений хвіст»:

$$n_{ПН} = \frac{2450 \cdot 1840}{3300} = 1366{,}0 \ ,$$

і т.д. Усі розраховані теоретичні частоти приведені в таблиці 7.4 курсивом. Як бачимо, навіть візуально видно, що фактичні і теоретичні частоти досить сильно відрізняються друг від друга, але наскільки вірогідно цю відмінність з'ясуємо за допомогою критерію Хі-квадрат:

$$\chi^2 = \frac{(700 - 474{,}0)^2}{474{,}0} + ... + \frac{(1310 - 1084{,}0)^2}{1084{,}0} = 328{,}1 \ .$$

Число ступенів волі в даному випадку дорівнює df = (2-1)·(2-1) = 1. Як бачимо, розраховане значення критерію Хі-квадрат набагато перевищує табличне (3,84), що свідчить про те, що гіпотеза про незалежну зустрічальність цієї пари ознак у даній групі тварин повинна бути відкинута. Тому приймається альтернативна їй гіпотеза про те, що має місце вірогідний зв'язок між появою у корови того чи іншого фенотипу крупа з характером постановки хвоста. Якщо проаналізувати дані табл.7.4 й порівняти фактичні частоти з теоретичними, то можна відзначити, що сполучення «прямий круп – високопоставлений хвіст» і «скошений круп – низькопоставлений хвіст» зустрічаються частіше, ніж очікувалося.

Після того як доведено, що асоціація між парою ознак має місце, можна вирішувати другу задачу – задачу оцінки сили цього зв'язку. Напрямок зв'язку може бути враховано тільки у випадку чотирипільної таблиці (тобто 2 х 2). Для багатопільних таблиць зв'язок завжди має позитивний знак.

Методи оцінки сили зв'язку для чотирипільної таблиці базуються на т.зв. «перехресному відношенні» (cross ratio) чи «відношенні переваги» (odd ratio). Якщо ми маємо таблицю спряженості розмірності 2 х 2, де всі її осередкі позначені в такий спосіб:

a	***b***
c	***d***

то відношення переваги має вид: $\frac{ad}{bc}$. Відомо, що у випадку незалежності ознак, це відношення дорівнює 1. На основі цієї закономірності був розроблено цілий ряд коефіцієнтів для оцінки сили зв'язку.

Найбільш простими і широко розповсюдженими є **коефіцієнт асоціації Юла** (Q_r):

$$Q_r = \frac{ad - bc}{ad + bc}, \tag{7.15}$$

$$SE_{Qr} = \frac{1 - Q_r^2}{2}\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}, \tag{7.16}$$

та **коефіцієнт контингенції** Шарл'є (r_β):

$$r_\beta = \frac{\left|ad - bc\right| - \frac{n}{2}}{\sqrt{(a+b)(b+d)(a+c)(b+c)}} = \sqrt{\frac{\chi^2}{n}}, \tag{7.17}$$

$$SEr_\beta = \frac{1 - r_\beta^2}{\sqrt{n}}. \tag{7.18}$$

Коефіцієнти Юла та Шарл'є варіюють від –1 до +1. Чим ближче величина показника до 1, тим сильніше виражений зв'язок. Знак перед оцінкою коефіцієнта показує напрямок цього зв'язку – якщо відношення переваги більше одиниці, то зв'язок позитивна, якщо ж менше – зв'язок негативна.

Необхідно відзначити, що коефіцієнт Юла завжди дає більш високе значення, чим коефіцієнт Шарл'є (на тих самих даних).

Оцінки статистичної помилки коефіцієнтів сили зв'язку дозволяють оцінити вірогідність відхилення величини коефіцієнта від нуля і, у випадку потреби, порівняти два коефіцієнти, розрахованих для різних вибірок.

Приклад 5. Оцінимо тепер силу зв'язку між ознаками, розглянутими в прикладі 4.

Використовуючи формули 7.15 і 7.16, розрахуємо значення коефіцієнта асоціації Юла та його статистичної помилки:

$$Q_r = \frac{700 \cdot 1310 - 150 \cdot 1140}{700 \cdot 1310 + 150 \cdot 1140} = +0,686,$$

$$SE_{Qr} = \frac{1 - 0,686^2}{2} \sqrt{\frac{1}{700} + \frac{1}{1140} + \frac{1}{150} + \frac{1}{1310}} = 0,026.$$

Аналогічно, за формулами 7.17 і 7.18 розрахуємо оцінки коефіцієнта контингенції Шарл'є та його статистичної помилки:

$$r_\beta = \frac{|700 \cdot 1310 - 150 \cdot 1140| - \frac{3300}{2}}{\sqrt{850 \cdot 2450 \cdot 1840 \cdot 1460}} = +0,315,$$

$$SEr_\beta = \frac{1 - 0,315^2}{\sqrt{3300}} = 0,016.$$

Як бачимо, в обох випадках зв'язок виявляється високо вірогідною й позитивною.

У тих випадках, коли таблиця сполученості має більше рядків і стовпців, методи оцінки сили зв'язку між ознаками базуються безпосередньо на величині розрахованого критерію Хі-квадрат. Сила такого зв'язку може бути оцінена за допомогою **міри Чупрова** (*T*) чи **міри Крамера** (*V*). Формули для обчислення цих показників та їхніх помилок наступні:

$$T = \sqrt{\frac{\chi^2}{n\sqrt{df}}}, \tag{7.19}$$

$$SE_T = \sqrt{\frac{1}{n\sqrt{df}}}; \tag{7.20}$$

$$V = \sqrt{\frac{\chi^2}{n \cdot \min\{v-1; c-1\}}}; \tag{7.21}$$

$$SE_V = \sqrt{\frac{1}{n \cdot \min\{v-1; c-1\}}}. \tag{7.22}$$

Вираз "min $\{v - 1;\ c - 1\}$" являє собою мінімальне значення з кількості стовпців таблиці сполученості мінус одиниця і кількості рядків таблиці мінус одиниця. Обидва показники варіюють від 0 до 1.

Приклад 6. У череді корів симентальської породи було проаналізовано зв'язок між фенами ушей та тривалістю тільності (виділені три градації – максимальна, середня і мінімальна). Всі вихідні дані наведено в таблиці 7.5.

Необхідно з'ясувати – чи має місце зв'язок між тривалістю тільності й особливостями в будівлі ушей тварин дослідної групи.

Таблиця 7.5

Тривалість тільності	Фен ушей					Разом
	загострені	тупі	вузькі	широкі	м'ясисті	
Максимальна	44 / *64,6*	67 / *48,7*	45 / *57,0*	45 / *37,8*	54 / *46,8*	255
Середня	42 / *44,8*	31 / *33,8*	33 / *39,6*	35 / *26,2*	36 / *32,5*	177
Мінімальна	85 / *61,6*	31 / *46,5*	73 / *54,5*	20 / *36,0*	34 / *44,7*	243
Разом	171	129	151	100	124	675

Першою справою необхідно розрахувати теоретичні частоти для кожної пари сполучень і порівняти ці розраховані частоти з фактичними, використовуючи критерій Хі-квадрат. Розрахунок проводиться стандартним методом (див. приклад 4). Теоретичні частоти приведені в табл. 7.5 курсивом.

Далі, використовуючи формулу 7.14, визначаємо, чи має місце вірогідний зв'язок між ознаками:

$$\chi^2 = \frac{(44-64{,}6)^2}{64{,}6} + \ldots + \frac{(34-44{,}7)^2}{44{,}7} = 53{,}38 .$$

У даному випадку число ступенів волі дорівнює df = (5– 1)·(3 –1) = 8. Табличне значення критерію Хі-квадрат для даного числа ступенів волі набагато нижче – 15,51 (табл. 7.3), отже, можна вважати доведеним, що має місце вірогідний зв'язок між формою ушей корів симентальської породи даної череди та тривалістю їх тільності.

Оцінку сили цього зв'язку розрахуємо, використовуючи міри Чупрова та Крамера:

$$T = \sqrt{\frac{53{,}38}{675 \cdot \sqrt{8}}} = 0{,}167; \quad SE_T = \sqrt{\frac{1}{675 \cdot \sqrt{8}}} = 0{,}023;$$

$$V = \sqrt{\frac{53{,}38}{675 \cdot 2}} = 0{,}199; \quad SE_V = \sqrt{\frac{1}{675 \cdot 2}} = 0{,}027.$$

Як бачимо, оцінка цього зв'язку хоча і вірогідна, але все-таки невелика.

►**Завдання**:

1. За даними, що наведено в таблиці, розрахуйте частоти всіх фенів та їх 95 % довірчий інтервал для тварин, що належать племзаводові "Матусово" (де i – дві останні цифри залікової книжки):

Фени типа спини	Племзавод	
	«Матусово»	«Тростянець»
Пряма	70 + *i*	240
Горбата	62	20 + *i*
Провисла	113	55 + *i*
Рівна	325 + *i*	812
Широка	308	654 + *i*
М'яка	12 + *i*	489 - *i*
Виконана	383 - *i*	298 - *i*
Не виконана	183	15 + *i*

2. Розрахуйте вибіркові оцінки фенетичного розмаїття за даною ознакою для тварин, що належать двом різним племінним заводам.

3. Порівняйте частоти зустрічальності фенів "пряма спина" та "горбата спина" у вибірках особин, що належать різним племзаводам.

4. Оцінити вірогідність та оцінку зв'язку між захворюваністю матерів та їх дочок на лейкоз (де *i* – дві останні цифри залікової книжки):

	Матері	
Дочки	здорові	хворі
здорові	391- *i*	114
хворі	70	46 + *i*

(Розрахуйте значення коефіцієнтів Юла та Шарл'є та їх помилки).

5. Необхідно з'ясувати – чи має місце зв'язок між тривалістю тільності й особливостями в будівлі ушей тварин наступної групи (де *i* – дві останні цифри залікової книжки):

Тривалість тільності	Фен ушей				
	загострені	тупі	вузькі	широкі	м'ясисті
Максимальна	44 + *i*	67	45	145 - *i*	54
Середня	42	31 + *i*	33	35	36
Мінімальна	85	31	73	20 + *i*	134 - *i*

(Розрахуйте значення мір Чупрова та Крамера та їх помилки). ◄

Література. **1**: с.45-47, с.54-57; **4**: с.354-355.

Питання для самоперевірки:

1. Що таке фенетика та які її основні завдання?
2. Для чого використовується арксинус-перетворення Фішера?

3. Що визначають показники фенетичного різноманіття?
4. Які показники використовуються для асоціації ознак у таблицях 2х2?

ПРАКТИЧНА РОБОТА № 8

Тема: Дисперсійний аналіз Р.Фішера
Мета: Набути навичок проведення класичного дисперсійного аналізу Р.Фішера

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ

Алгоритм *дисперсійного аналізу* було розроблено англійським вченим Р.Фішером в 1925 р. та набув суттєвого розвитку в працях його учня – Йєтса.

Основна мета дисперсійного аналізу – розкладання загальної мінливості ознаки на мінливість часткову, що виникає між членами популяції під впливом багатьох різноманітних факторів. В ході аналізу встановлюють частку мінливості, що обумовлена кожним фактором в експерименті, частку мінливості, що обумовлена сумісною дією цих факторів, а також частку мінливості, що є результатом впливу багатьох неорганізованих (випадкових) факторів.

Дисперсійний аналіз широко використовується зоотехнії та селекції с.-г. тварин:

- при визначенні оцінки коефіцієнтів успадкування (h^2) та повторюваності (w);
- при оцінці плідників за якістю нащадків;
- при оцінці частки генотипової та паратипової мінливості в загальній мінливості тварин, тощо.

Дисперсійний аналіз базується на понятті *дисперсії*, тобто суми квадратів відхилень кожної варіанти вибірки від середнього арифметичного значення:

$$C = \sum (x_i - \bar{x})^2 \; . \qquad (8.1)$$

Р.Фішер показав, що якщо аналізується мінливість в декількох групах, загальна мінливість такого комплексу (Cy) може бути розкладена на дві складові – мінливість ознаки між групами (міжгрупова дисперсія – Cx) та мінливість ознаки в середині груп (внутрішньогрупова дисперсія – Cz). Залежність між цими джерелами варіювання можна виразити рівнянням:

$$\sum (x_{ij} - \bar{x})^2 = n \cdot \sum (\bar{x}_i - \bar{x})^2 + \sum \sum (x_{ij} - \bar{x}_i)^2 , \qquad (8.2)$$

тобто, $Cy = Cx + Cz$.

1. <u>Однофакторний дисперсійний аналіз</u>. Алгоритм однофакторнго дисперсійного аналізу краще продемонструвати на прикладі.

Приклад 1. Необхідно визначити, чи впливає вік свиноматок на їх крупноплідність.

Всі вихідні та проміжні дані наведено в табл. 8.1.

Таблиця 8.1

	Градації фактору *A* (вік у опоросах)			Суми
	перший	другий	третій	
x_{ij}	0,9	1,1	1,0	
	1,1	1,1	1,3	
	1,0	1,3	1,4	
	1,0	1,2	1,2	
	1,0	1,1	1,3	
n_i	5	5	5	$\Sigma n_i = N = 15$
Σx_i	5,0	5,8	6,2	$\Sigma\Sigma x_i = 17,0$
$\frac{\left(\sum x_i\right)^2}{n_i}$	5,000	6,728	7,688	$\sum \frac{\left(\sum x_i\right)^2}{n_i} = 19,416$
Σx_i^2	5,02	6,76	7,78	$\Sigma\Sigma x_i^2 = 19,56$

Дисперсійний комплекс має три градації фактору *A*, тобто $l = 3$. Всього в аналізі було використано дані щодо 15 свиноматок ($N = 15$).

Для зручності розрахунку окремих компонент загальної дисперсії попередньо необхідно розрахувати три допоміжні величини:

- загальну суму значень ознаки по всім градаціям фактора: $\Sigma\Sigma x_i = 17,0$;
- суму відношень квадратів сум значень ознаки по кожній градації до відповідного об’єму груп: $\sum \frac{\left(\sum x_i\right)^2}{n_i} = 19,416$;
- загальну суму квадратів значень ознаки по всім градація: $\Sigma\Sigma\ x_i^2 = 19,56$.

Для розрахунку міжгрупової та внутрішньогрупової дисперсій необхідно також розрахувати допоміжну величину:

$$H = \frac{\left(\sum\sum x_i\right)^2}{N}. \qquad (8.3)$$

Для даних з прикладу 1 значення цієї допоміжної величини складає:

$$H = \frac{17{,}0^2}{15} = 19{,}267 .$$

Тоді значення загальної, міжгрупової та внутрішньогрупової дисперсії можна розрахувати за формулами:

$$Cy = \sum\sum {x_i}^2 - H ; \qquad (8.4)$$

$$Cx = \sum \frac{\left(\sum x_i\right)^2}{n_i} - H ; \qquad (8.5)$$

$$Cz = Cy - Cx . \qquad (8.6)$$

Для даних із прикладу 1 ці значення складають, відповідно:

Cy = 19,560 - 19,267 = 0,293;
Cx = 19,416 - 19,267 = 0,149;
Cz = 0,293 - 0,149 = 0,144.

Далі необхідно визначити числу ступенів свободи для кожної з дисперсій:

$$k_y = N - 1 ; \qquad (8.7)$$

$$k_x = l - 1 ; \qquad (8.8)$$

$$k_z = N - l . \qquad (8.9)$$

Таким чином, для даних із нашого прикладу число ступенів свободи для загальної дисперсії складає: $k_y = 15 - 1 = 14$; для міжгрупової дисперсії: $k_x = 3 - 1 = 2$; для внутрішньогрупової: $k_z = 15 - 3 = 12$.

Розраховуємо значення варіанс, тобто середніх квадратів відхилень; для цього необхідно оцінки дисперсій віднести до відповідних значень числа ступенів свободи:

$$\sigma_y^2 = \frac{Cy}{k_y} ; \qquad (8.10)$$

$$\sigma_x^2 = \frac{Cx}{k_x} ; \qquad (8.11)$$

$$\sigma_z^2 = \frac{Cz}{k_z} . \qquad (8.12)$$

Таким чином, загальна варіанса складатиме: $\sigma_y^2 = \frac{0{,}293}{14} = 0{,}0209$; міжгрупова варіанса: $\sigma_x^2 = \frac{0{,}149}{2} = 0{,}0745$; внутрішньогрупова варіанса: $\sigma_z^2 = \frac{0{,}144}{12} = 0{,}0120$.

Частка впливу фактору *A* (в нашому випадку – віку свиноматки) на загальну мінливість ознаки (тобто, їх крупноплідність) та її статистична помилка визначаються за формулами:

$$\eta^2 = \frac{Cx}{Cy}; \tag{8.13}$$

$$SE_{\eta^2} = (1 - \eta^2) \cdot \left(\frac{l - 1}{N - l} \right). \tag{8.14}$$

Таким чином, для даних із прикладу 1 частка мінливості крупноплідності, що залежить від віку свиноматки складає:

$$\eta^2 = 0{,}149 : 0{,}293 = 0{,}509$$

із статистичною помилкою:

$$SE_{\eta^2} = (1 - 0{,}509) \cdot \left(\frac{3 - 1}{15 - 3} \right) = 0{,}082.$$

Тобто, на 50,9 % крупноплідність свиноматок залежить від їх віку (номеру опоросу), а на неорганізовані (випадкові) фактори припадає 49,1 % загальної мінливості.

Рівень значущості одержаної величини можна оцінити на підставі рівня значущості критерію Р.Фішера:

$$F = \frac{\sigma_x^2}{\sigma_z^2}. \tag{8.15}$$

Для цього розраховане за формулою 8.15 значення дисперсійного відношення необхідно порівняти із табличним значенням критерію Р.Фішера із відповідними числами ступенів свободи, які наведені в Додатку Б. (В верхній частині таблиці наведено число ступенів свободи для чисельнику формули 8.15, а у лівій – для знаменнику.)

В нашому випадку дисперсійне відношення дорівнює: $F = \frac{0{,}0745}{0{,}0120} = 6{,}208$. Це отримане значення необхідно порівняти із табличним із числом ступенів свободи: для чисельника – $df_1 = l - 1 = 2$, для знаменника – $df_2 = N - l = 12$.

Але в таблиці Додатку Б не наведено значення критерію Р.Фішера для таких чисел ступенів свободи ($F_{2,\,12}$). Для числа ступенів свободи $df_1 = 2$ (верхня строка таблиці) присутні тільки значення для числа ступенів свободи знаменника $df_2 = 10$ ($F_{2,\,10} = 4{,}103$) та $df_2 = 15$ ($F_{2,\,15} = 3{,}682$). Тому необхідно використати лінійну гармонійну інтерполяцію по вертикалі.

В цьому випадку шукане табличне значення критерію Р.Фішера можна розрахувати за формулою:

$$F = F_0 + u \cdot (F_1 - F_0), \tag{8.16}$$

де $u = \dfrac{df_1 \cdot (df_0 - df)}{df \cdot (df_0 - df_1)}$; F_0 – табличне значення критерію Р.Фішера для df_0; F_1 – табличне значення критерію Р.Фішера для df_1; при цьому $df_0 < df < df_1$.

В нашому випадку df_0 = 10, df = 12, df_1 = 15, тому: $u = \dfrac{15 \cdot (10-12)}{12 \cdot (10-15)} = 0{,}5$. Тоді шукане табличне значення критерію Р.Фішера складатиме: $F_{2,\,12}$ = 4,103 + 0,5·(3,682 - 4,103) = 3,893.

Оскільки розраховане значення дисперсійного відношення (F) набагато вище, ніж табличне, вплив фактору A (віку свиноматки) на їх крупноплідність вважається вірогідним з рівнем значущості $p < 0{,}05$.

Всі розрахункові дані оформлюються у вигляді таблиці дисперсійного аналізу (табл. 8.2).

Таблиця 8.2

Джерело мінливості	Дисперсія (C)	Число ступенів свободи (k)	Варіанса (σ^2)	Дисперсійне відношення (F)	Сила впливу (h^2)
Фактор A	0,144	2	0,0745	***6,208***	0,509
Залишкова (Z)	0,149	12	0,0120		0,491
Сумарна (Y)	0,293	14			

2. Повний двохфакторний дисперсійний аналіз. У селекції двохфакторний дисперсійний аналіз використовується для більш детального аналізу фенотипової варіації – розкладанні її на гено- і паратипову компоненти. Таким чином, у якості факторів найчастіше використовується генотип особин (фактор A) і особливості утримання, вплив середовища і т.п. (фактор B). Розберемо алгоритм повного двохфакторного дисперсійного аналізу на наступному прикладі.

Приклад 2. Вивчався вплив генотипу (порода) і раціону (контроль і з преміксом) на багатоплідність свиней.

Усі допоміжні величини розраховані в таблиці 8.3. Першою чергою необхідно розрахувати допоміжну величину H:

$$H = \frac{\left(\sum\sum x_i\right)^2}{N} = \frac{181^2}{20} = 1638{,}05.$$

Далі, використовуючи цю величину, розраховуємо загальну дисперсію:

$$C_Y = \sum\sum x_i^2 - H = 1681{,}0 - 1638{,}05 = 42{,}95,$$

і міжгрупову (факторіальну) дисперсію:

$$C_X = \sum \frac{\left(\sum x_i\right)^2}{n} = 1662{,}6 - 1638{,}05 = 24{,}55.$$

Оцінку внутрішньогрупової (залишкової) дисперсії знаходимо як різницю між сумарною і факторіальною дисперсіями, тобто

$$C_Z = C_Y - C_X = 42{,}95 - 24{,}55 = 18{,}40.$$

Однак факторіальна дисперсія в такому виді являє собою суму дисперсій одночасно для факторів *A* та *B*, а також для їхнього сполучення. Тому необхідно знову розкласти факторіальну дисперсію на окремі компоненти:

$$C_A = \sum \frac{\left(\sum x_A\right)^2}{n_A} - H = 1656{,}1 - 1638{,}05 = 18{,}05,$$

$$C_B = \sum \frac{\left(\sum x_B\right)^2}{n_B} - H = 1644{,}1 - 1638{,}05 = 6{,}05.$$

Таблиця 8.3

	*A*1		*A*2		Суми
	*B*1	*B*2	*B*1	*B*2	
	7	8	8	12	
	6	9	10	10	
	8	8	11	9	
	7	10	10	10	
	9	9	9	11	
n	5	5	5	5	$\Sigma n = N = 20$
Σx_i	37	44	48	52	$\Sigma \Sigma x_i = 181$
$\frac{\left(\sum x_i\right)^2}{n}$	273,8	387,2	460,8	540,8	$\sum \frac{\left(\sum x_i\right)^2}{n} = 1662{,}6$
Σx_i^2	279	390	466	546	$\Sigma \Sigma x_i^2 = 1681{,}0$
n_A	5 + 5 = 10		5 + 5 = 10		
Σx_A	37 + 44 = 81		48 + 52 = 100		
$\frac{\left(\sum x_A\right)^2}{n_A}$	656,1		1000,0		$\sum \frac{\left(\sum x_A\right)^2}{n_A} = 1656{,}1$
n_B	5 + 5 = 10		5 + 5 = 10		
Σx_B	37 + 48 = 85		44 + 52 = 96		
$\frac{\left(\sum x_B\right)^2}{n_B}$	722,5		921,6		$\sum \frac{\left(\sum x_B\right)^2}{n_B} = 1644{,}1$

Нарешті, дисперсія спільної дії факторів розраховується за формулою:

$$C_{AB} = C_X - (C_A + C_B) = 24{,}55 - (18{,}05 + 6{,}05) = 0{,}45.$$

Переходимо до розрахунку числа ступенів свободи для кожного джерела варіювання. Для

- сумарної дисперсії: $k = N - 1 = 20 - 1 = 19$;
- дисперсії по фактору A: $k = a - 1 = 2 - 1 = 1$;
- дисперсії по фактору B: $k = b - 1 = 2 - 1 = 1$;
- дисперсії спільного впливу AB: $k_{AB} = (a - 1)\cdot(b - 1) = 1$;
- залишкової дисперсії: $k = N – a\cdot b = 20 - 4 = 16$,

де a - число градацій по фактору A; b - число градацій по фактору B; N - загальне число особин, що використані у дисперсійному комплексі.

Розрахунок варіанс проводиться віднесенням значень дисперсій до відповідного числа ступенів свободи:

$$\sigma_A^2 = \frac{C_A}{k_A} = \frac{18{,}05}{1} = 18{,}05;$$

$$\sigma_B^2 = \frac{C_B}{k_B} = \frac{6{,}05}{1} = 6{,}05;$$

$$\sigma_{AB}^2 = \frac{C_{AB}}{k_{AB}} = \frac{0{,}45}{1} = 0{,}45;$$

$$\sigma_Z^2 = \frac{C_Z}{k_Z} = \frac{18{,}4}{16} = 1{,}15.$$

Для того, щоб оцінити вірогідність впливу факторів A, B та їхнього сполучення, необхідно розрахувати дисперсійні відношення:

$$F_A = \frac{\sigma_A^2}{\sigma_Z^2} = \frac{18{,}05}{1{,}15} = 15{,}70;$$

$$F_B = \frac{\sigma_B^2}{\sigma_Z^2} = \frac{6{,}05}{1{,}15} = 5{,}26;$$

$$F_{AB} = \frac{\sigma_{AB}^2}{\sigma_Z^2} = \frac{0{,}45}{1{,}15} = 0{,}39.$$

Далі, кожне дисперсійне відношення (F) порівнюємо з табличним значенням F-критерію Р.Фішера (див. Додаток Б). У всіх трьох випадках для F_A, F_B та F_{AB} табличне значення критерію повинне мати однакові числа ступенів свободи ($df_1 = 1$, $df_2 = 16$). В таблиці Додатку Б не наведено значення критерію $F_{1,16}$. Але його можна розрахувати на підставі алгоритму лінійної гармонійної інтерполяції за формулою 8.16.

В цьому випадку $u = 0{,}25$, а шукане значення критерію Р.Фішера:

$$F_{1,19} = 4{,}543 + 0{,}25\cdot(4{,}351 - 4{,}543) = 4{,}495.$$

Таким чином, результати аналізу довели, що як порода (фактор *A*), так й раціон годівлі (фактор *B*) мають вірогідний вплив на багатоплідність свиней. При цьому породні відмінності на 42 % обумовлюють варіювання ознаки, а тип раціону – лише на 14 %. Сумісної дії породи та раціону не доведено. На випадкові (неорганізовані) фактори приходить 43 % загальної мінливості багатоплідності свиней.

Усі результати двохфакторного дисперсійного аналізу оформляються у вигляді таблиці (табл. 8.4).

Таблиця 8.4

Джерело мінливості	Дисперсія (С)	Число ступенів свободи (k)	Варіанса (σ^2)	Дисперсійне відношення (*F*)	Сила впливу (h^2)
Фактор *A*	18,05	1	18,05	***15,70***	0,42
Фактор *B*	6,05	1	6,05	***5,26***	0,14
Сполучення *AB*	0,45	1	0,45	0,39	0,01
Залишкова (*Z*)	18,40	16	1,15		0,43
Сумарна (*Y*)	42,95	19			

►**Завдання**: 1. Від трьох бугаїв герефордської породи було отримано по сім телят, жива маса (в кг) який у шестимісячному віці була наступна:

Клевер 874	178 + *i*	186	174	162	216 - *i*	230
Жемчуг 191	146	185 + *i*	192	166 - *i*	176	179
Гром 325	199	210	187 - *i*	199	175	200 + i

де *i* – остання цифра залікової книжки.

З'ясувати наявність та ступінь впливу генотипу бугаїв на живу масу їх нащадків.

2. Оцінити вплив типу розведення (*A1* – внутрішньопородне схрещування; *A2* – міжпорідне схрещування) і раціону (*B1* – середній рівень поживних речовин; *B2* – високий рівень) на багатоплідність у свиней великої білої породи:

A1		*A2*	
B1	*B2*	*B1*	*B2*
9	10	9	12
10	11	10	11
11	11	9	13

9	10	11	15
11	9	12	12
9	10	11	10
8	12	12	13
10	12	11	11
12	10	8	9
9	10	12	10

◄

Література. **1**: с.72-82; **2**: с.49-58; **3**: с.121-128; **4**: с.279-289.

Питання для самоперевірки:

1. Хто й коли розробив дисперсійний аналіз?
2. Для вирішення яких задач використовується дисперсійний аналіз в зоотехнії?
3. Що визначає правило розкладання дисперсій?

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

ОСНОВНА:

1. Аничин В. Л. Математическая статистика : учебное пособие. Харьков : Харьк. гос. аграр. ун-т им. В. В. Докучаева, 1992. 114 с.
2. Комп'ютерні методи в сільського господарстві та біології: Навчальний посібник / О. М. Царенко, Ю. А. Злобін, В. Г. Скляр, С. М. Панченко. Суми : Видавництво "Університетська книга", 2000. 203 с.
3. Мармоза А. Т. Практикум по математической статистике : Учебное пособие. Киев : Выща шк., 1990. 191 с.
4. Опря А. Т. Статистика: з програмованою формою контролю знань. Київ : Урожай, 1996. 448 с.

ДОДАТКОВА:

1. Животовский Л. А. Популяционная биометрия. Москва : Наука, 1991. 271 с.
2. Лакин Г. Ф. Биометрия. Москва : Высшая школа, 1980. 293 с.
3. Меркурьева Е. К. Биометрия в селекции и генетике сельскохозяйственных животных. Москва : Колос, 1970. 423 с.
4. Плохинский Н. А. Руководство по биометрии для зоотехников. Москва : Колос, 1969. 256 с.
5. Терентьев П. В., Ростова Н. С. Практикум по биометрии. Ленинград : Изд-во ЛГУ, 1977. 152 с.
6. Урбах В. Ю. Статистический анализ в биологических и медицинских исследованиях. Москва : Медицина, 1975. 296 с.

Додаток А

Ординати нормальної кривої Гауса-Лапласа $f(t) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}$

t	Соті частки *t*									
	0	1	2	3	4	5	6	7	8	9
0,0	5000	5040	5080	5120	5160	5199	5239	5279	5319	5359
0,1	5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
0,2	5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0,3	6179	6217	6255	6293	6331	6368	6406	6442	6480	6517
0,4	6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0,5	6915	6950	6985	7019	7054	7088	7123	7157	7190	7224
0,6	7257	7291	7324	7357	7389	7422	7454	7486	7517	7549
0,7	7580	7611	7642	7673	7703	7734	7764	7794	7823	7852
0,8	7881	7910	7939	7967	7995	8023	8051	8078	8106	8133
0,9	8159	8186	8212	8238	8264	8289	8315	8340	8365	8389
1,0	8413	8438	8461	8485	8508	8531	8554	8577	8599	8621
1,1	8643	8665	8686	8708	8729	8749	8770	8790	8810	8830
1,2	8849	8869	8888	8907	8925	8944	8962	8980	8997	9015
1,3	9032	9049	9066	9082	9099	9115	9131	9147	9162	9177
1,4	9192	9207	9222	9236	9251	9265	9278	9292	9306	9319
1,5	9332	9345	9357	9370	9382	9394	9406	9118	9430	9441
1,6	9452	9463	9474	9484	9495	9505	9515	9525	9535	9545
1,7	9554	9564	9573	9582	9591	9599	9608	9616	9625	9633
1,8	9641	9648	9656	9664	9671	9778	9686	9693	9700	9706
1,9	9713	9719	9726	9732	9738	9744	9750	9756	9762	9767
2,0	9772	9778	9783	9788	9793	9798	9803	9808	9812	9817
2,1	9821	9826	9830	9834	9838	9842	9846	9850	9854	9857
2,2	9861	9865	9868	9871	9875	9878	9881	9884	9887	9890
2,3	9893	9896	9898	9901	9904	9906	9909	9911	9913	9916
2,4	9918	9920	9922	9925	9927	9929	9931	9932	9934	9936
2,5	9938	9940	9941	9943	9945	9946	9948	9949	9951	9952
2,6	9953	9955	9956	9957	9959	9960	9961	9962	9963	9964
2,7	9965	9966	9967	9968	9969	9970	9971	9972	9973	9974
2,8	9974	9975	9976	9977	9977	9978	9979	9979	9980	9981
2,9	9981	9982	9982	9983	9984	9984	9985	9985	9986	9986
3,0	9987	9987	9987	9988	9988	9989	9989	9989	9990	9990
3,1	9990	9991	9991	9991	9992	9992	9992	9992	9993	9993
3,2	9993	9993	9994	9994	9994	9994	9994	9995	9995	9995
3,3	9995	9995	9995	9996	9996	9996	9996	9996	9996	9997
3,4	9997	9997	9997	9997	9997	9997	9997	9997	9997	9998
3,5	9998	9998	9998	9998	9998	9998	9998	9998	9998	9998

Примітка. В таблиці нуль цілих та кома не наведені.

Додаток Б

Критичні значення F-критерію Фішера
(для $\alpha = 0,05$)

df_2	df_1									
	1	2	3	4	5	6	7	8	9	10
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	3,020	2,978
15	4,543	3,682	3,287	3,056	2,901	2,790	2,707	2,641	2,588	2,544
20	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,393	2,348
25	4,242	3,385	2,991	2,759	2,603	2,490	2,405	2,337	2,282	2,236
30	4,171	3,316	2,922	2,690	2,534	2,421	2,334	2,266	2,211	2,165
35	4,121	3,267	2,874	2,641	2,485	2,372	2,285	2,217	2,161	2,114
40	4,085	3,232	2,839	2,606	2,449	2,336	2,249	2,180	2,124	2,077
45	4,057	3,204	2,812	2,579	2,422	2,308	2,221	2,152	2,096	2,049
50	4,034	3,183	2,790	2,557	2,400	2,286	2,199	2,130	2,073	2,026
55	4,016	3,165	2,773	2,540	2,383	2,269	2,181	2,112	2,055	2,008
60	4,001	3,150	2,758	2,525	2,368	2,254	2,167	2,097	2,040	1,993
65	3,989	3,138	2,746	2,513	2,356	2,242	2,154	2,084	2,027	1,980
70	3,978	3,128	2,736	2,503	2,346	2,231	2,143	2,074	2,017	1,969
80	3,960	3,111	2,719	2,486	2,329	2,214	2,126	2,056	1,999	1,951
90	3,947	3,098	2,706	2,473	2,316	2,201	2,113	2,043	1,986	1,938
100	3,936	3,087	2,696	2,463	2,305	2,191	2,103	2,032	1,975	1,927
120	3,920	3,072	2,680	2,447	2,290	2,175	2,087	2,016	1,959	1,910
140	3,909	3,061	2,669	2,436	2,279	2,164	2,076	2,005	1,947	1,899
160	3,900	3,053	2,661	2,428	2,271	2,156	2,067	1,997	1,939	1,890
180	3,894	3,046	2,655	2,422	2,264	2,149	2,061	1,990	1,932	1,884
200	3,888	3,041	2,650	2,417	2,259	2,144	2,056	1,985	1,927	1,878
250	3,879	3,032	2,641	2,408	2,250	2,135	2,046	1,976	1,917	1,869
300	3,873	3,026	2,635	2,402	2,244	2,129	2,040	1,969	1,911	1,862
350	3,868	3,022	2,630	2,397	2,240	2,125	2,036	1,965	1,907	1,858
400	3,865	3,018	2,627	2,394	2,237	2,121	2,032	1,962	1,903	1,854
450	3,862	3,016	2,625	2,392	2,234	2,119	2,030	1,959	1,901	1,852
500	3,860	3,014	2,623	2,390	2,232	2,117	2,028	1,957	1,899	1,850
600	3,857	3,011	2,620	2,387	2,229	2,114	2,025	1,954	1,895	1,846
700	3,855	3,009	2,618	2,385	2,227	2,112	2,023	1,952	1,893	1,844
800	3,853	3,007	2,616	2,383	2,225	2,110	2,021	1,950	1,892	1,843
900	3,852	3,006	2,615	2,382	2,224	2,109	2,020	1,949	1,890	1,841
1000	3,851	3,005	2,614	2,381	2,223	2,108	2,019	1,948	1,889	1,840

Додаток Б (продовження)

df_2	df_1									
	11	12	13	14	15	16	17	18	19	20
10	2,943	2,913	2,887	2,865	2,845	2,828	2,812	2,798	2,785	2,774
15	2,507	2,475	2,448	2,424	2,403	2,385	2,368	2,353	2,340	2,328
20	2,310	2,278	2,250	2,225	2,203	2,184	2,167	2,151	2,137	2,124
25	2,198	2,165	2,136	2,111	2,089	2,069	2,051	2,035	2,021	2,007
30	2,126	2,092	2,063	2,037	2,015	1,995	1,976	1,960	1,945	1,932
35	2,075	2,041	2,012	1,986	1,963	1,942	1,924	1,907	1,892	1,878
40	2,038	2,003	1,974	1,948	1,924	1,904	1,885	1,868	1,853	1,839
45	2,009	1,974	1,945	1,918	1,895	1,874	1,855	1,838	1,823	1,808
50	1,986	1,952	1,921	1,895	1,871	1,850	1,831	1,814	1,798	1,784
55	1,968	1,933	1,903	1,876	1,852	1,831	1,812	1,795	1,779	1,764
60	1,952	1,917	1,887	1,860	1,836	1,815	1,796	1,778	1,763	1,748
65	1,939	1,904	1,874	1,847	1,823	1,802	1,782	1,765	1,749	1,734
70	1,928	1,893	1,863	1,836	1,812	1,790	1,771	1,753	1,737	1,722
80	1,910	1,875	1,845	1,817	1,793	1,772	1,752	1,734	1,718	1,703
90	1,897	1,861	1,830	1,803	1,779	1,757	1,737	1,720	1,703	1,688
100	1,886	1,850	1,819	1,792	1,768	1,746	1,726	1,708	1,691	1,676
120	1,869	1,834	1,803	1,775	1,750	1,728	1,709	1,690	1,674	1,659
140	1,858	1,822	1,791	1,763	1,738	1,716	1,696	1,678	1,661	1,646
160	1,849	1,813	1,782	1,754	1,729	1,707	1,687	1,669	1,652	1,637
180	1,842	1,806	1,775	1,747	1,722	1,700	1,680	1,661	1,645	1,629
200	1,837	1,801	1,769	1,742	1,717	1,694	1,674	1,656	1,639	1,623
250	1,827	1,791	1,759	1,732	1,707	1,684	1,664	1,645	1,628	1,613
300	1,821	1,785	1,753	1,725	1,700	1,677	1,657	1,638	1,621	1,606
350	1,816	1,780	1,748	1,720	1,695	1,672	1,652	1,633	1,616	1,601
400	1,813	1,776	1,745	1,717	1,691	1,669	1,648	1,630	1,613	1,597
450	1,810	1,774	1,742	1,714	1,689	1,666	1,646	1,627	1,610	1,594
500	1,808	1,772	1,740	1,712	1,686	1,664	1,643	1,625	1,607	1,592
600	1,805	1,768	1,736	1,708	1,683	1,660	1,640	1,621	1,604	1,588
700	1,802	1,766	1,734	1,706	1,681	1,658	1,637	1,619	1,601	1,586
800	1,801	1,764	1,732	1,704	1,679	1,656	1,636	1,617	1,600	1,584
900	1,799	1,763	1,731	1,703	1,678	1,655	1,634	1,615	1,598	1,582
1000	1,798	1,762	1,730	1,702	1,676	1,654	1,633	1,614	1,597	1,581

Навчальне видання

СТАТИСТИЧНІ МЕТОДИ У БІОТЕХНОЛОГІЇ

Методичні рекомендації

Укладач: **Крамаренко Сергій Сергійович**

Формат 60×84.1/16.Ум. друк. арк. 0,9

Тираж ___ прим. Зам № ____

Надруковано у видавничому відділі

Миколаївського національного аграрного університету.

54020, м. Миколаїв, вул. Георгія Гонгадзе, 9

Свідоцтво суб’єкта видавничої справи ДК № 4490 від 20.02.2013 р.