

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

УМАНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ САДІВНИЦТВА

Кафедра інформаційних технологій

**КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ ТА ОБРОБКА
ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ В БІОЛОГІЇ**

**Методичні рекомендації до виконання лабораторних занять
здобувачами освітнього рівня «Магістр»
спеціальності 091 – «Біологія та біохімія»**

Умань–2024 р.

Методичні вказівки підготував:

С.М. Концеба, к. е. н., доцент кафедри інформаційних технологій

Розглянуті і затверджені на засіданні кафедри інформаційних технологій (протокол від 06.08.2024 року № 2).

Рецензенти:

кандидат с.-г. наук, доцент кафедри екології та безпеки життєдіяльності
А.В. Балабак;

кандидат с.-г. наук, доцент кафедри генетики, селекції рослин та біотехнології
О.П. Сержук.

Концеба С.М. Комп'ютерне моделювання та обробка експериментальних даних в біології. Методичні рекомендації до виконання практичних занять здобувачами освітнього ступеня «Магістр» спеціальності 091 – «Біологія та біохімія». – Умань, 2024. – 42 с.

ЗМІСТ

Вступ		4
Лабораторне заняття №1	Моделювання та вирішення у програмі <i>Mathematica</i> задач Коші на прикладі простої ферментативної реакції	6
Лабораторне заняття №2	Моделювання та вирішення у програмі <i>Mathematica</i> задач Коші на прикладі ферментативної реакції з конкурентним інгібуванням	8
Лабораторне заняття №3	Моделювання та вирішення у програмі <i>Mathematica</i> задач Коші на прикладі ферментативної реакції з неконкурентним інгібуванням	11
Лабораторне заняття №4	Моделювання та побудова графіків швидкості ферментних реакцій	15
Лабораторне заняття №5	Чисельне знаходження залежності значення стаціонарного стійкого стану від параметру f та при різних значеннях параметру g	17
Лабораторне заняття №6	Розв'язування задач за моделлю Вольтера з обмеженням максимальної чисельності жертв	18
Лабораторне заняття №7	Визначення обсягу вибірки	20
Лабораторне заняття №8	Розрахунок кореляційних зав'язків між ознаками	28
Лабораторне заняття №9	Розрахунок коефіцієнту регресії	31
Лабораторне заняття №10	Розрахунок однофакторного дисперсійного аналізу	38
Рекомендована література		42

ВСТУП

Мета дисципліни – набуття студентами знань про принципи побудови математичних моделей біологічних процесів з використання різних комп'ютерних програм, сучасні підходи у проведенні статистичного аналізу результатів біологічних досліджень, а також про принципи узагальнення та оформлення результатів таких досліджень.

Завдання – освоїти математичні поняття і методи, що використовуються при комп'ютерному моделюванні біологічних процесів, набути навичок обробки, класифікації, аналізу і інтерпретації отриманих результатів досліджень.

Місце дисципліни у структурно-логічній схемі підготовки здобувачів вищої освіти. Навчальна дисципліна «Комп'ютерне моделювання та обробка експериментальних даних в біології» є обов'язковою і займає відповідне місце у структурно-логічній схемі підготовки фахівців і тісно пов'язана з іншими дисциплінами, зокрема: математика, сучасні геоінформаційні системи, математичні методи та інформаційні технології в біології, вступ до фаху та основи наукової діяльності та іншими дисциплінами, знаннями яких студенти повинні оволодіти.

Інтегральна компетентність – здатність розв'язувати складні задачі і проблеми в галузі біології при здійсненні професійної діяльності або у процесі навчання, що передбачає проведення досліджень та/або здійснення інновацій та характеризується невизначеністю умов і вимог.

Загальні компетентності:

- ЗК02. Здатність використовувати інформаційні та комунікаційні технології.
- ЗК05. Здатність розробляти та керувати проектами.
- ЗК06. Здатність проведення досліджень на відповідному рівні.

Спеціальні компетентності:

- СК02. Здатність формулювати задачі моделювання, створювати моделі об'єктів і процесів на прикладі різних рівнів організації живого із

використанням математичних методів інформаційних технологій.

- СК03. Здатність користуватися сучасними інформаційним технологіями та аналізувати інформацію в галузі біології і на межі предметних галузей.

Програмні результати навчання:

- ПР02. Використовувати бібліотеки, інформаційні бази даних, інтернет ресурси для пошуку необхідної інформації.
- ПР07. Описувати й аналізувати принципи структурно-функціональної організації, механізмів регуляції та адаптації організмів до впливу різних чинників.
- ПР10. Представляти результати наукової роботи письмово (у вигляді звіту, наукових публікацій тощо) та усно (у формі доповідей та захисту звіту) з використанням сучасних технологій, аргументувати свою позицію в науковій дискусії.
- ПР11. Проводити статистичну обробку, аналіз та узагальнення отриманих експериментальних даних із використанням програмних засобів та сучасних інформаційних технологій.
- ПР15. Уміти самостійно планувати і виконувати інноваційне завдання та формулювати висновки за його результатами.

Лабораторне заняття №1

Моделювання та вирішення у програмі *Mathematica* задач Коші на прикладі простої ферментативної реакції

Задача Розв'яжіть наведену задачу Коші (2.1.5) прийнявши $k_1=10^8$, $k_2=10^2$, $k_{-1}=10^4$, $F_0(0)=5 \cdot 10^{-5}$, $S(0)=10^{-2}$.

Програма 2.

```
k1=10^8;k2=10^2;km1=10^4;Fo=5*10^-5;So=10^-2;
T=3;(*Час спостереження в секундах*);
Km=(km1+k2)/k1;theta=1/(km1+k2);xi=km1/(km1+k2);
fo=Fo/Km;so=So/Km;solution=NDSolve[
{f1'[tau]==s[tau](fo - f1[tau]) - f1[tau],f1[0]==0,
 s'[tau]==xi*f1[tau] - s[tau](fo - f1[tau]),s[0]==so},
{f1,s},{tau,0,T/theta}];
SetOptions[ParametricPlot,FrameLabel->{"час t,
с","концентрація/Kм"},
Frame->True,Axes->False,AxesOrigin->{0,0},
PlotRange->{Automatic,{0,1.1*so}},
DisplayFunction->$DisplayFunction,AspectRatio->1];
graph1=ParametricPlot[{tau*
theta,Evaluate[{s[tau]}/.solution]},{tau,0,T/theta},
PlotStyle->{Red,Thickness[0.005],Dashing[{0.01,0.01]}}];
graph2=ParametricPlot[{tau*theta,100*Evaluate[{f1[tau]}/.solu
tion]},{tau,0,T/theta},PlotStyle-
>{Blue,Thickness[0.005]};T=10^-5;
(* Будуємо другий графік на меньшому часовому
інтервалі *)
graph3=ParametricPlot[{tau*theta,Evaluate[{s[tau]}/.solution]
},{tau,0,T/theta},
PlotStyle->{Red,Thickness[0.005],Dashing[{0.01,0.01]}}];
graph4=ParametricPlot[{tau*theta,100*Evaluate[{f1[tau]}/.solu
tion]},{tau,0,T/theta},PlotStyle->{Blue,Thickness[0.005]};
Show[GraphicsRow[{Show[graph1,graph2],Show[graph3,gra
ph4]}],
DisplayFunction->$DisplayFunction,ImageSize->900]
```

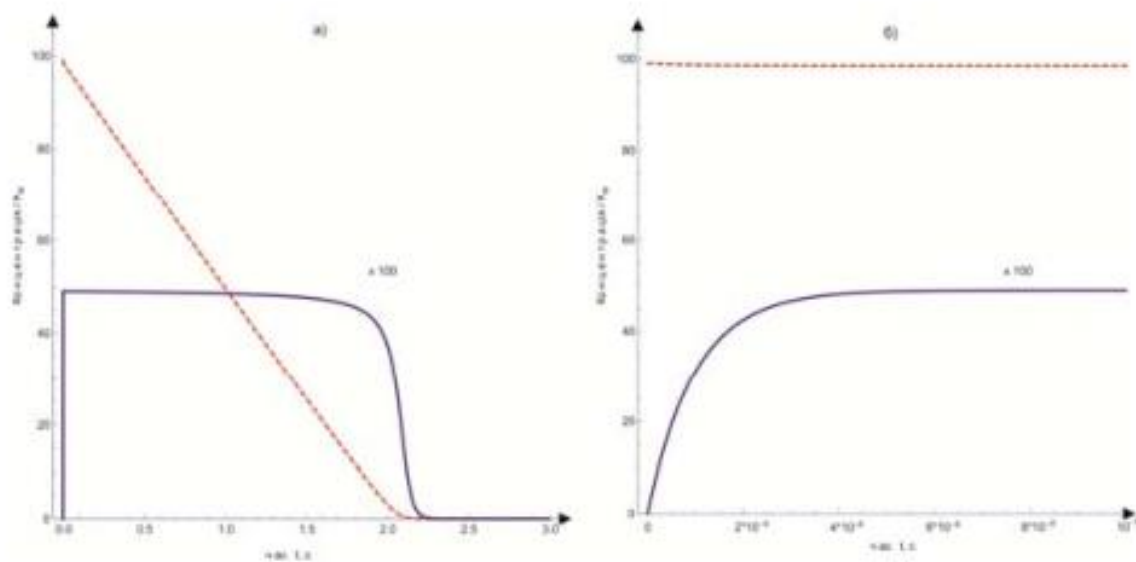


Рис. 3.3 Результат роботи програми. Червона лінія концентрація f_1 - (ФСК), синя лінія концентрація s – субстрату збільшена в сто разів ($\times 100$)

Лабораторне заняття №2

Моделювання та вирішення у програмі *Mathematica* задач Коші на прикладі ферментативної реакції з конкурентним інгібуванням

Задача Розв'яжіть наведену задачу Коші (2.2.5) прийнявши

$$k_1 = 10^8 M^{-1} c^{-1}, k_2 = 10^2 c^{-1}, k_{-1} = 10^4 c^{-1}, k_3 = 10^8 c^{-1},$$

$$k_{-3} = 5 * 10^3 c^{-1}. \text{ Використайте } F_0(0) = 5 * 10^{-5} M,$$

$$I(0) = 6 * 10^{-5} M, S(0) = 10^{-2} M.$$

Програма 3.

```
k1=10^8;k2=10^2;km1=10^4;k3=10^8;km3=5*10^3;
Fo=5*10^-5;Io=6*10^-5;So=10^-2;
T=3;(*Час спостереження в секундах*);
Km=(km1+k2)/k1;theta=1/(km1+k2);xi=km1/(km1+k2);a=km3/(
km1+k2);
b=k3/k1;fo=Fo/Km;io=Io/Km;so=So/Km;
solution=NDSolve[
{{s'[tau]==xi* f1[tau] -s[tau](fo+i[tau]-f1[tau]-io),s[0]==so,
(i'[tau]==a* (io-i[tau]) -b*i[tau](fo+i[tau]-f1[tau]-io),i[0]==io,
(f1'[tau]==s[tau](fo+i[tau]-f1[tau]-io)- f1[tau],f1[0]==0),
{s,i,f1},{tau,0,T/theta}};SetOptions[ParametricPlot,
FrameLabel->{"час t, c","концентрація/Km"},Frame-
>True,Axes->False,
AxesOrigin->{0,0},PlotRange->{Automatic,{0,1.1*so}},
DisplayFunction->$DisplayFunction,AspectRatio->1];
graph1=ParametricPlot[{tau
theta,Evaluate[{s[tau]}/.solution]},{tau,0,T/theta},
```



```

PlotStyle->{Red,Thickness[0.005],Dashing[{0.01,0.01}}];
graph2=ParametricPlot[{tau*theta,100*Evaluate[{i[tau]}/.solution]},
{tau,0,T/theta},PlotStyle->{Green,Thickness[0.005}}];
graph3=ParametricPlot[{tau*theta,100*Evaluate[{f1[tau]}/.solution]},
{tau,0,T/theta},PlotStyle->{Blue,Thickness[0.005}}];
T=10^-5;(*Будуємо другий графік на меньшому часовому інтервалі *)
graph4=ParametricPlot[{tau
theta,Evaluate[{s[tau]}/.solution]},{tau,0,T/theta},
PlotStyle->{Red,Thickness[0.005],Dashing[{0.01,0.01}}];
graph5=ParametricPlot[{tau*theta,100*Evaluate[{i[tau]}/.solution]},
{tau,0,T/theta},PlotStyle->{Green,Thickness[0.005],Dashing[{0.01,0.01}}];
graph6=ParametricPlot[{tau*theta,100*Evaluate[{f1[tau]}/.solution]},
{tau,0,T/theta},PlotStyle->{Blue,Thickness[0.005}}];
Show[GraphicsRow[{Show[graph1,graph2,graph3],Show[graph4,graph5,graph6]}],DisplayFunction->$DisplayFunction,ImageSize->900]

```

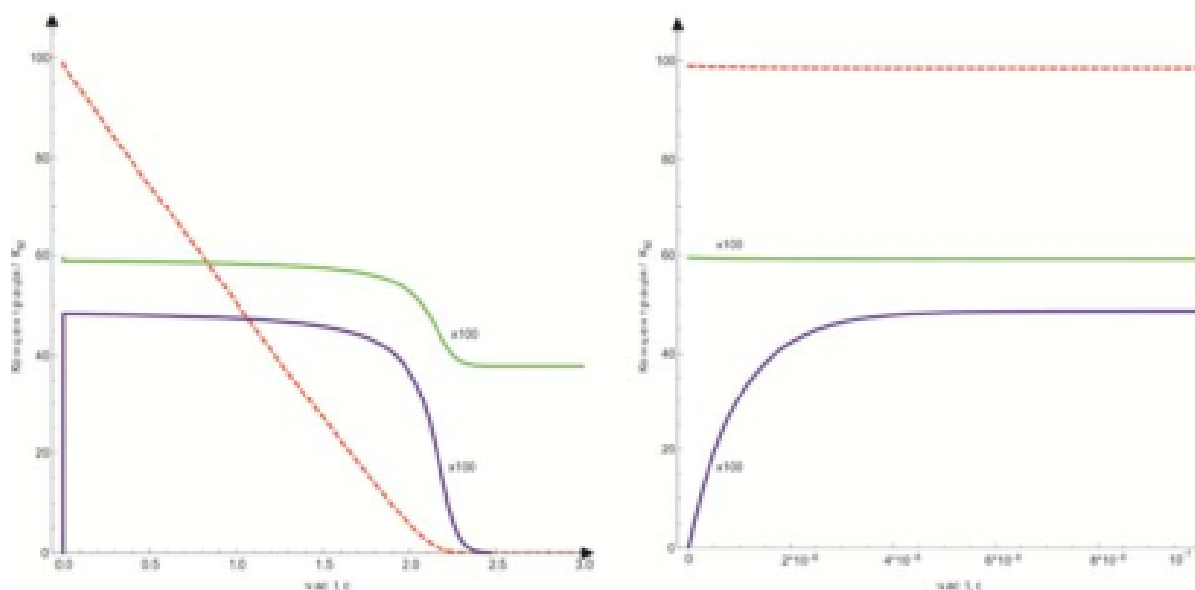


Рис. 3.5 Результат роботи програми. Червона лінія концентрація субстрату $S(t)$, зелена лінія інгібітору $i(t)$ збільшена в сто разів ($\times 100$), синя ФСК $f_1(t)$ збільшена в сто разів ($\times 100$)

Для наведеної задачі $\chi \approx 1$, $\alpha \approx 0.5$, $\beta \approx 1$, $s^0 \approx 100$, $i^0 \approx 0.5$, $f^0 \approx 0.5$. Розглянемо тепер нетипові значення констант $k_1 = 10^8 M^{-1} c^{-1}$, $k_2 = 5 \cdot 10^2 c^{-1}$,

$k_{-1} = 10^4 \text{ c}^{-1}$, $k_3 = 10^9 \text{ c}^{-1}$, $k_{-3} = 5 * 10^2 \text{ c}^{-1}$. Використаємо
 $F_0(0) = 5 * 10^{-5} M$, $I(0) = 10^{-4} M$, $S(0) = 10^{-3} M$. Для цих
 значень: $\chi \approx 1$, $\alpha \approx 0.05$, $\beta \approx 10$, $s^0 \approx 10$, $i^0 \approx 1$, $f^0 \approx 0.5$. Розрахунки, виконані для
 цих значень, дають:

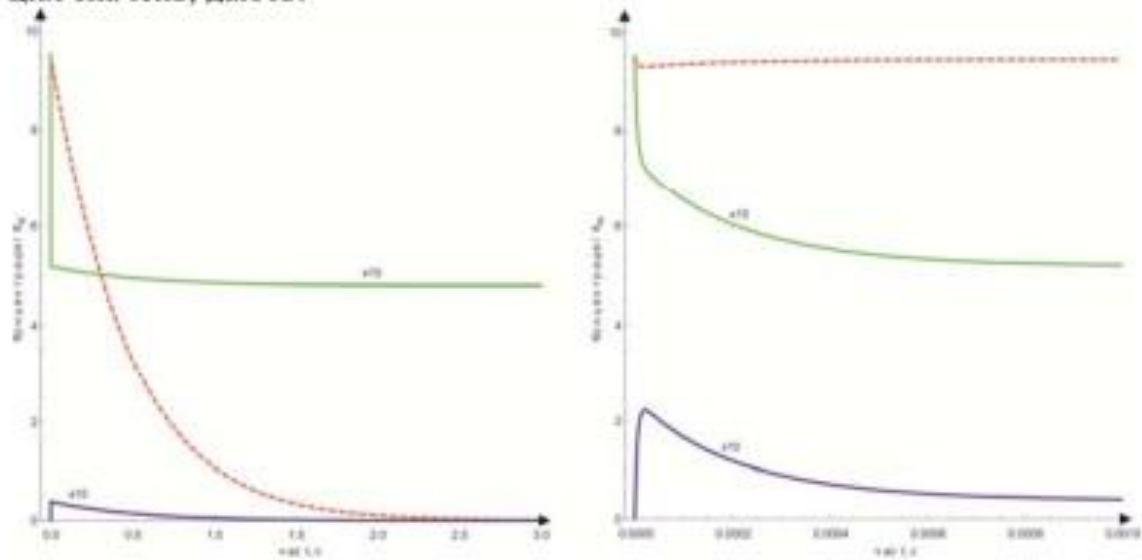


Рис. 3.6 Червона лінія концентрація субстрату $S(t)$, зелена лінія інгібітору $I(t)$ збільшена в десять разів (x10), синя ФСК $f_1(t)$ збільшена в десять разів (x10)

Лабораторне заняття №3

Моделювання та вирішення у програмі *Mathematica* задач Коші на прикладі ферментативної реакції з неконкурентним інгібуванням

Задача Розв'яжіть наведену задачу Коші (2.3.6), прийнявши

$$k_1 = 10^8 M^{-1} c^{-1}, k_2 = 10^2 c^{-1}, k_{-1} = 10^4 c^{-1}, k_3 = 10^8 c^{-1},$$

$$k_{-3} = 5 \cdot 10^3 c^{-1}. \text{ Використайте } F_0(0) = 5 \cdot 10^{-5} M,$$

$$I(0) = 6 \cdot 10^{-5} M, S(0) = 10^{-2} M.$$

Програма 4.

```
k1=10^8;k2=10^2;km1=10^4;k3=10^8;km3=5*10^3;
Fo=5*10^-5;Io=6*10^-5;So=10^-2; Km=(km1+k2)/k1;theta=1/(km1+k2);
xi=km1/(km1+k2);a=km3/(km1+k2);b=k3/k1;
fo=Fo/Km;io=Io/Km;so=So/Km;
T=5;(*Час спостереження в секундах*);
solution=NDSolve[
{f0'[tau]==fo-io-f0[tau]+i[tau]- s[tau]f0[tau],f0[0]==fo,
i'[tau]==a*(io-i[tau]) -b*i[tau](fo-io-f0[tau]+i[tau]),i[0]==io,
s'[tau]==xi*(fo-io- f0[tau]+i[tau]) -s[tau]f0[tau],s[0]==so
},{f0,i,s},{tau,0,T/theta}];
SetOptions[ParametricPlot, FrameLabel->{"час t,
c","концентрація/Km"}, Frame->True,Axes->False, AxesOrigin->{0,0},
PlotRange->{Automatic, {0,1.1*so}}, DisplayFunction-
->$DisplayFunction,AspectRatio->1];
graph1=ParametricPlot[{tau theta,
Evaluate[{s[tau]}/.solution]},{tau,0,T/theta},PlotStyle->
{Red,Thickness[0.005],Dashing[{0.01,0.01}]}];
graph2=ParametricPlot[{tau
theta,100*Evaluate[{i[tau]}/.solution]},{tau,0,T/theta},PlotStyle->
{Green,Thickness[0.005]}];
graph3=ParametricPlot[{tau theta,100*Evaluate[{fo-io-
f0[tau]+i[tau]}/.solution]},{tau,0,T/theta},PlotStyle->
{Blue,Thickness[0.005]}];
T=10^-5;(*Будуємо другий графік на меншому часовому інтервалі *)
graph4=ParametricPlot[{tau
theta,Evaluate[{s[tau]}/.solution]},{tau,0,T/theta}, PlotStyle->
{Red,Thickness[0.005],Dashing[{0.01,0.01}]}];
graph5=ParametricPlot[{tau
theta,100*Evaluate[{i[tau]}/.solution]},{tau,0,T/theta}, PlotStyle->
{Green,Thickness[0.005],Dashing[{0.01,0.01}]}];
graph6=ParametricPlot[{tau theta,100*Evaluate[{fo-io-
f0[tau]+i[tau]}/.solution]},{tau,0,T/theta},PlotStyle-
->{Blue,Thickness[0.005]}];
Show[GraphicsRow[{Show[graph1,graph2,graph3],Show[graph4,graph5,gr
aph6}], DisplayFunction->$DisplayFunction,ImageSize->900]
```

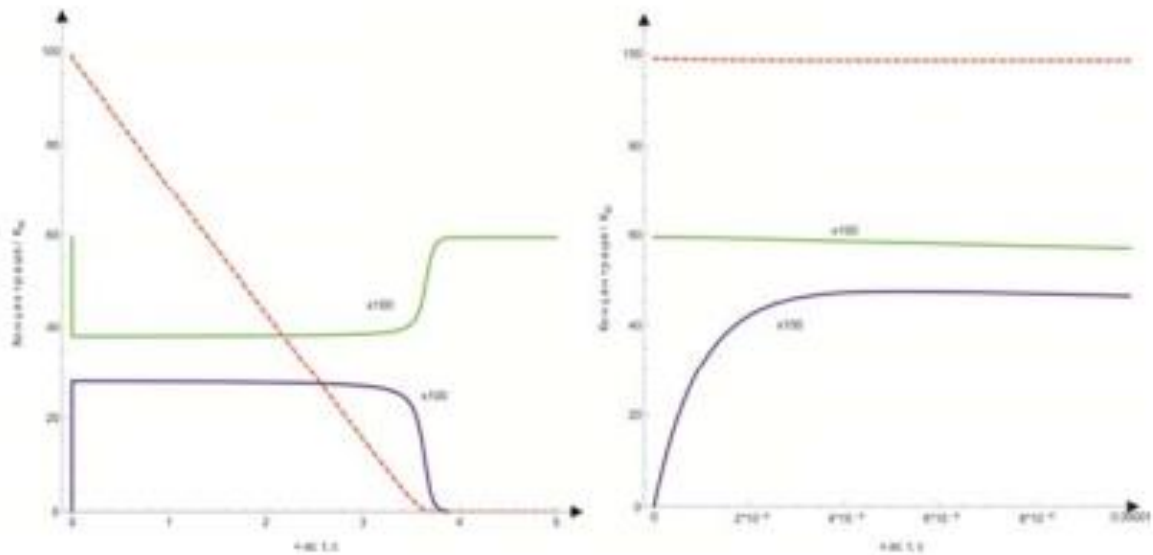


Рис. 3.8 Результат роботи програми. Червона лінія концентрація субстрату $S(t)$, зелена лінія інгібітору $i(t)$ збільшена в сто разів ($\times 100$), синя ФСК збільшена в сто разів ($\times 100$)

Для наведеної задачі $\chi \approx 1$, $\alpha \approx 0.5$, $\beta \approx 1$, $s^0 \approx 100$, $i^0 \approx 0.5$, $f^0 \approx 0.5$. Розглянемо тепер не зовсім типові значення констант $k_1 = 10^8 M^{-1} c^{-1}$,

$$k_2 = 5 * 10^2 c^{-1}, k_{-1} = 10^4 c^{-1}, k_3 = 10^9 c^{-1}, k_{-3} = 5 * 10^2 c^{-1}.$$

Використаємо $F_0(0) = 5 * 10^{-5} M$, $I(0) = 10^{-4} M$, $S(0) = 10^{-3} M$.

Для цих значень: $\chi \approx 1$, $\alpha \approx 0.05$, $\beta \approx 10$, $s^0 \approx 10$, $i^0 \approx 1$, $f^0 \approx 0.5$. Розрахунки, виконані для цих значень, дають:

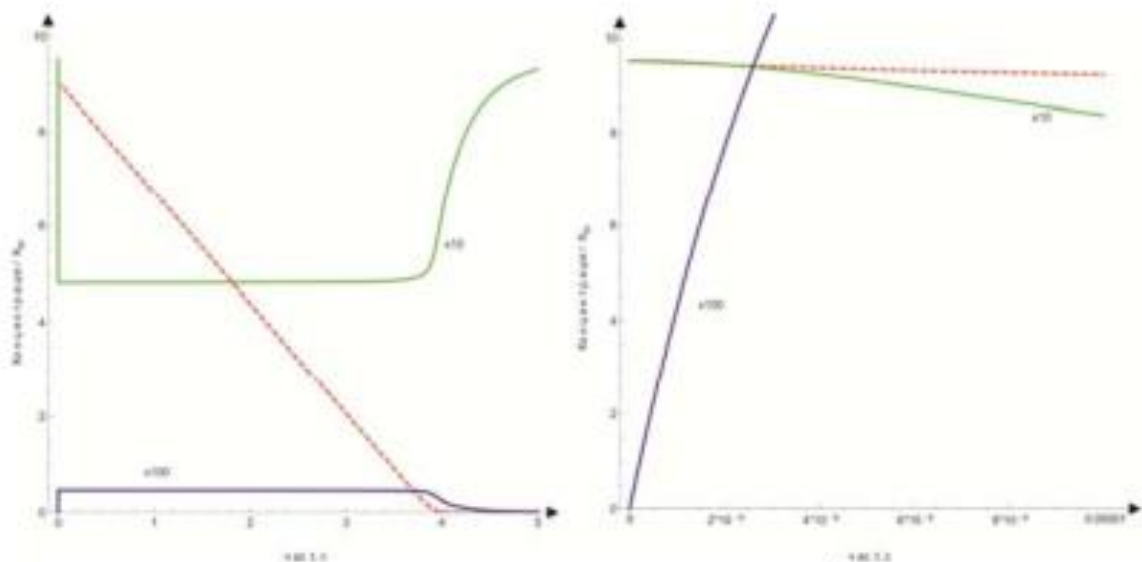


Рис. 3.9 Червона лінія концентрація субстрату $S(t)$, зелена лінія інгібітору $i(t)$ збільшена в десять разів ($\times 10$), синя ФСК $f_1(t)$ збільшена в десять разів ($\times 10$)

Наостанок розглянемо дуже нетипові значення констант $k_1 = 10^8 M^{-1} c^{-1}$, $k_2 = 5 * 10^2 c^{-1}$, $k_{-1} = 10^3 c^{-1}$, $k_3 = 3 * 10^7 c^{-1}$, $k_{-3} = 10^4 c^{-1}$. І використаємо $F_0(0) = 5 * 10^{-5} M$, $I(0) = 10^{-4} M$, $S(0) = 10^{-3} M$. За цих значень: $\chi \approx 0.6$, $\alpha \approx 7$, $\beta \approx 0.3$. Побудуємо швидкість ферментативної реакції для випадку конкурентного та неконкурентного інгібування.

Програма 5.

```

k1=10^8;k2=5*10^2;km1=1000;k3=3*10^7;km3=10^4;
Fo=5*10^-5;Io=10^-4;So=10^-3;T=0.09;(*Час спостереження в
секундах*);
Km=(km1+k2)/k1;theta=1/(km1+k2);
xi=km1/(km1+k2);a=km3/(km1+k2);b=k3/k1;
fo=Fo/Km;io=Io/Km;so=So/Km;
solution=NDSolve[
{s'[tau]==xi*f1[tau]-s[tau](fo+i[tau]-f1[tau]-io),s[0]==so,
i'[tau]==a*(io-i[tau])-b*i[tau](fo+i[tau]-f1[tau]-io),i[0]==io,
f1'[tau]==s[tau](fo+i[tau]-f1[tau]-io)-f1[tau],f1[0]==0},
{s,i,f1},{tau,0,T/theta}];SetOptions[ParametricPlot,
FrameLabel->{"час t, c","концентрація/Km"},Frame->True,Axes->False,
AxesOrigin->{0,0},PlotRange->{Automatic,{0,10*so}},DisplayFunction-
>SDisplayFunction,AspectRatio-
>1];graphKonkurent=ParametricPlot[{tau*
theta,100*Evaluate[{f1[tau]}/.solution]},{tau,0,T/theta},PlotStyle->
{Blue,Thickness[0.005]};
solution=NDSolve[
{f0'[tau]==fo-io-f0[tau]+i[tau]-s[tau]f0[tau],f0[0]==fo,
I'[tau]==a*(io-i[tau])-b*i[tau](fo-io-f0[tau]+i[tau]),i[0]==io,
s'[tau]==xi*(fo-io-f0[tau]+i[tau])-s[tau]f0[tau],s[0]==so
},{f0,i,s},{tau,0,T/theta}];
graphNeKonkurent=ParametricPlot[{tau theta,100*Evaluate[{(fo-io-
f0[tau]+i[tau])}/.solution]},{tau,0,T/theta},PlotStyle-
>{Red,Thickness[0.005]};
Show[{graphKonkurent,graphNeKonkurent}]

```

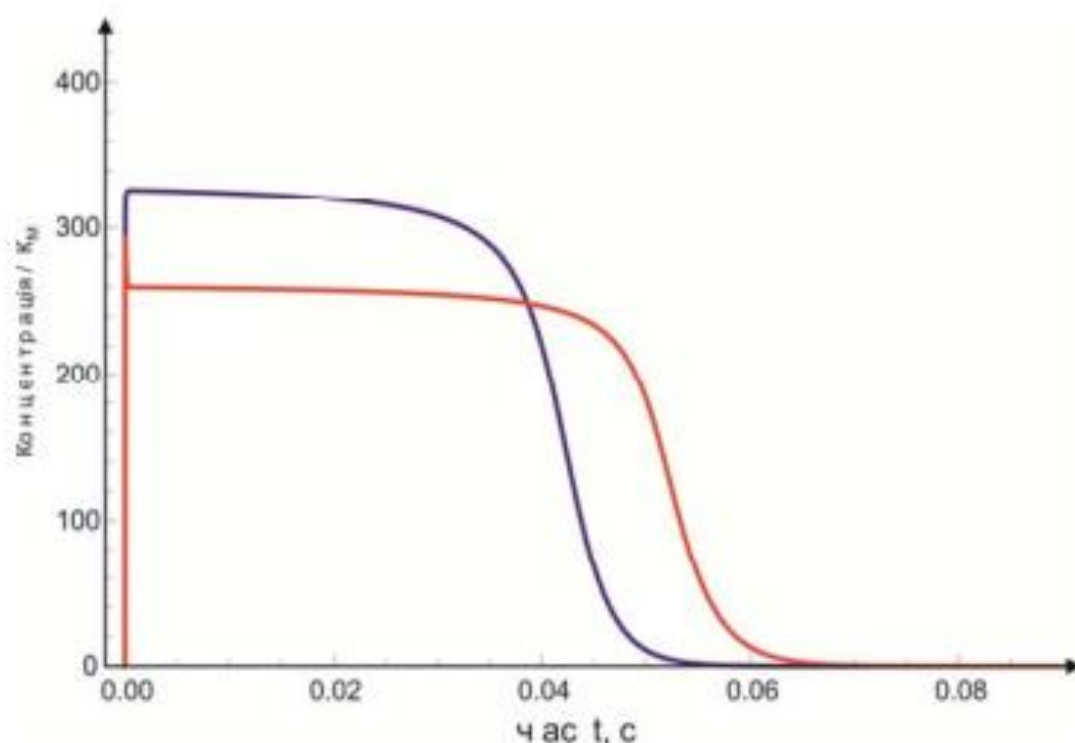


Рис. 3.10 Швидкість ферментативної реакції з конкурентним інгібуванням (синя лінія) та неконкурентним інгібуванням (червона лінія).

З порівняння чисельних результатів Рис. 3.10, отриманих для однакових констант швидкостей реакцій та початкових умов, але різних типів інгібування, бачимо, що максимум графіка швидкості для неконкурентного інгібування має гостріший профіль і максимальна швидкість, що спостерігається в реакції з неконкурентним інгібуванням, є меншою, ніж в реакції з конкурентним інгібуванням, що повністю узгоджується з теоретичними передбаченнями, зробленими для стаціонарної кінетики.

Лабораторне заняття №4

Моделювання та побудова графіків швидкості ферментних реакцій

Задача Побудуйте графіки швидкості реакції, що відбувається по схемі Рис. 3.11, розв'язавши наведену задачу Коші (3.4.6) та прийнявши $k_1 = k_3 = 10^7 \text{ M}^{-1} \text{ c}^{-1}$, $k_2 = 1.3 * 10^3 \text{ c}^{-1}$, $k_{-1} = k_{-3} = 10^4 \text{ c}^{-1}$ та за різних значень $k_4 = 10^2 \text{ c}^{-1}$, $k_4 = 5 * 10^2 \text{ c}^{-1}$, $k_4 = 9 * 10^2 \text{ c}^{-1}$, $k_4 = 1.3 * 10^3 \text{ c}^{-1}$ і переконайтесь, що у всіх випадках, крім останнього швидкість реакції має максимум, оскільки за таких констант $k_4/k_2 = B/A \neq 1$, і тільки коли $k_4 = 1.3 * 10^3$, то $k_4/k_2 = B/A = 1$. Використайте $S(0) = 10^{-2} \text{ M}$, $F^0 = 5 * 10^{-3} \text{ M}$.

Програма 6.

```
km1=10^4;k1=10^7;k2=13*10^2;k3=10^7;k4=10^2;km3=10^4;
Col={Red,Green,Blue,Black};k=1;
Table[Fo=5*10^-3;So=10^-2;
T=0.005>(*Час спостереження в секундах*);
Km=(km1+k2)/k1;theta=1/(km1+k2);
xi=km1/(km1+k2);b=(km3+k4)/(km1+k2);a=k3/k1;c=km3/(km1+k
2);
fo=Fo/Km;so= So/Km;
solution=NDSolve[{f0'[tau]==fo-f0[tau]-f1[tau]-2s[tau] f0[tau]
,f0[0]==fo,
s'[tau]==xi (fo-f0[tau])+(2c-xi)f1[tau]-a s[tau](fo- f1[tau])+(a-2)
s[tau]f0[tau],s[0]==so,f1'[tau]==a s[tau](fo-f0[tau]-f1[tau])-2b
f1[tau],f1[0]==0},
{f0,s,f1},{tau,0,T/theta}];SetOptions[ParametricPlot,AspectRatio-
>1,
FrameLabel->{"час t, c","швидкість"},Frame->True,Axes-
>False,DisplayFunction->Identity,PlotRange-
>{Automatic,{0,10}}];
graph1=ParametricPlot[{t theta,Evaluate[{k2 Km (fo-f0[t ]-f1[
t])+2 k4 Km f1[ t ]}/.solution]},{t,0,T/theta},
AxesOrigin->{-0.0001,0},PlotStyle-
>{Col[[k]],Thickness[0.005]};T=T/50;
```

```

graph2=ParametricPlot[{t theta,Evaluate[{k2 Km (fo-f0[t ]-f1[
t])+2 k4 Km f1[ t ]]/.solution]},{t,0,T/theta},
AxesOrigin->{-0.000001,0},PlotStyle-
>{Col[[k]],Thickness[0.005]};k=k+1;
Show[GraphicsArray[{graph1,graph2}],
DisplayFunction->$DisplayFunction,ImageSize->600],
{k4,13*10^2,10^2,-4*10^2};//Show

```

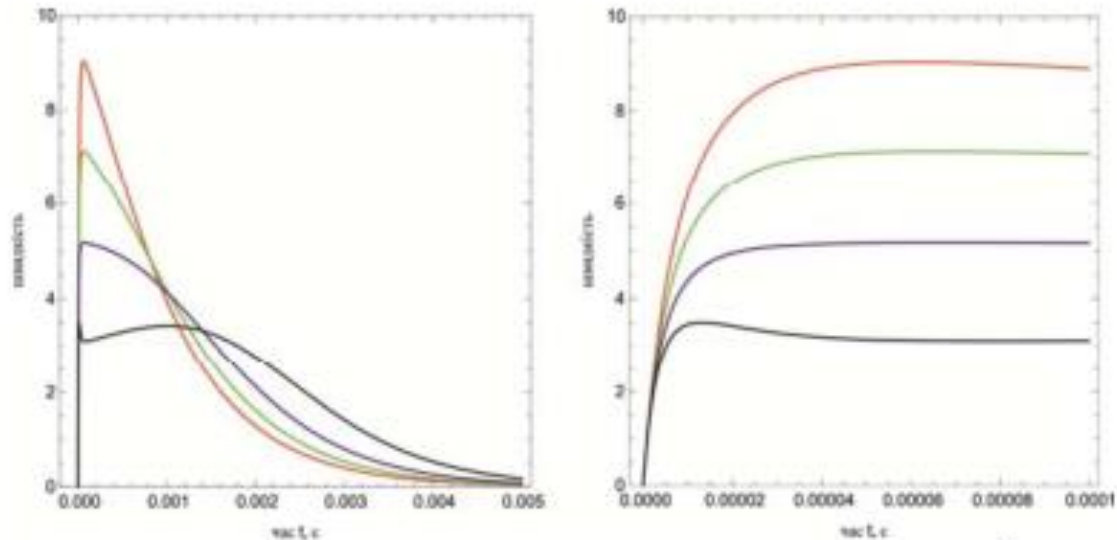


Рис. 3.12 Результат роботи програми за параметрів $k_1 = k_3 = 10^7 \text{ M}^{-1} \text{ c}^{-1}$, $k_2 = 1.3 * 10^3 \text{ c}^{-1}$, $k_{-1} = k_{-3} = 10^4 \text{ c}^{-1}$ та за різних значень $k_4 = 10^2 \text{ c}^{-1}$ (чорна лінія), $k_4 = 5 * 10^2 \text{ c}^{-1}$ (синя), $k_4 = 9 * 10^2 \text{ c}^{-1}$ (зелена), $k_4 = 1.3 * 10^3 \text{ c}^{-1}$ (червона).

Зміна параметрів призводить до кооперативної поведінки (Рис. 3.12). Найбільш яскраво це помітно при сильному відхиленні відношення k_4/k_2 від одиниці, для наведених залежностей найбільша відмінність спостерігається для чорної лінії. Бачимо, що навіть на початку реакції, коли субстрату достатньо, на кривій залежності швидкості утворення продукту (чорна лінія) є максимум.

Лабораторне заняття №5

Чисельне знаходження залежності значення стаціонарного стійкого стану від параметру f та при різних значеннях параметру g

Задача Чисельно знайдіть, за яких значень параметру g при фіксованих початкових умовах $x(0)=2$ буде реалізовуватись X_1 , а за яких X_2 . Візьміть $f = 2$ та $g \in [0.5,1]$.

Програма 7.

```
f = 2; xo = 0.4; T = 15; list = {};  
SetOptions[{Plot, ListPlot}, PlotStyle -> {{Dashing[{0, 0]}}, Dashing[{0.01, 0.01]}},  
FrameLabel -> {"час t", "концентрація"}, Frame -> True, Axes -> False,  
AxesOrigin -> {0, 0}, DisplayFunction -> Identity, PlotRange -> {0, yo}}];  
SetOptions[ListPlot, FrameLabel -> {"g", "x"}];  
Table[solution = NDSolve[{x'[t] == -x[t]^3 + 3*x[t]^2 - f*x[t] + g, x[0] == xo}, {x}, {t, 0, T}]; graph1 = Plot[Evaluate[{x[t]} /. solution], {t, 0, T}, PlotPoints -> 1000];  
Show[graph1, DisplayFunction -> $DisplayFunction]; list = Join[list, {{g, Evaluate[x[T] /. solution][[1]]}}, {g, 0.1, 1, 0.01}]; graph2 = ListPlot[list];  
Show[GraphicsArray[{graph1, graph2}], DisplayFunction -> $DisplayFunction, ImageSize -> 900];
```

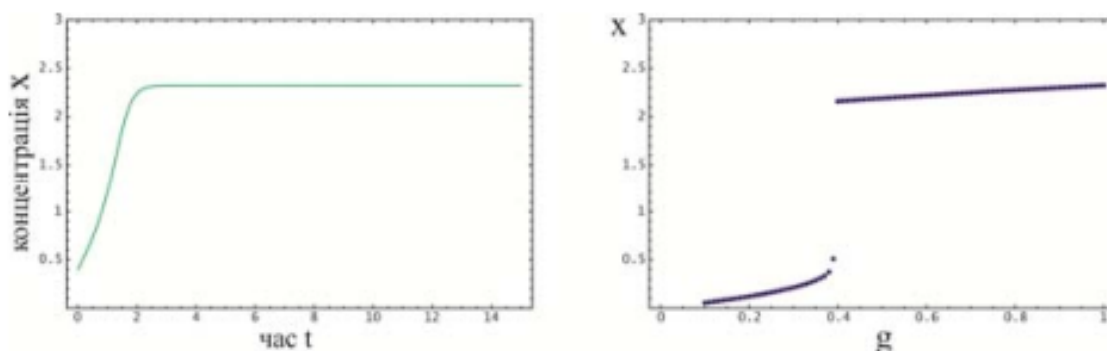


Рис. 4.2 Результат роботи програми. Залежність $x(t)$ (ліворуч) при параметрах $f=2$, $g=1$ та при фіксованому $f=2$. Праворуч залежність $x(15)$ при різних значеннях g . При $g \approx 0.4$ відбувається стрибковий перехід від одного стійкого положення до іншого.

Лабораторне заняття №6

Розв'язування задач за моделлю Вольтера з обмеженням максимальної чисельності жертв

Задача Чисельно розв'яжіть наведену задачу Коші (5.4.1), прийнявши $\alpha = 0.9$, $\beta = 0.8$, $\gamma = 0.9$, $k_4 = 0.7$. Використайте $X(0) = 0.5$, $Y(0) = 2$, також перевірте, що під час руху величина E зберігається.

Програма 9.

```
a = 0.9; b = 0.8; g = 0.9; d = 0.7; Xc = a/b; Yc = d/(b*g);
xo = 0.5/Xc; yo = 2/Yc; T = 30; solution = NDSolve[{
x'[t] == d*x[t]*(y[t] - 1), x[0] == xo, y'[t] == a*y[t]*(1 - x[t]), y[0]
== yo}, {x, y}, {t, 0, T}, Method -> ExplicitRungeKutta];
SetOptions[Plot, FrameLabel -> {Style["час t",FontSize-
>16], Style["кількість",FontSize->16]}, Frame -> True, Axes
-> False,
AxesOrigin -> {0, 0},PlotRange -> {0, 2*yo},
PlotStyle-
>{{Red,Thickness[0.005]},{Green,Thickness[0.005]}}];
graph = Plot[Evaluate[{Xc*x[t], Yc*y[t]} /. solution], {t, 0, T}];
Plot[Evaluate[{{(x[t]/E^x[t])^a*(y[t]/E^y[t])^d}/.solution},{t,0,T
}],
DisplayFunction->$DisplayFunction,PlotRange->{0,0.2}]
Show[graph, DisplayFunction -> Identity, ImageSize -> 900]
```

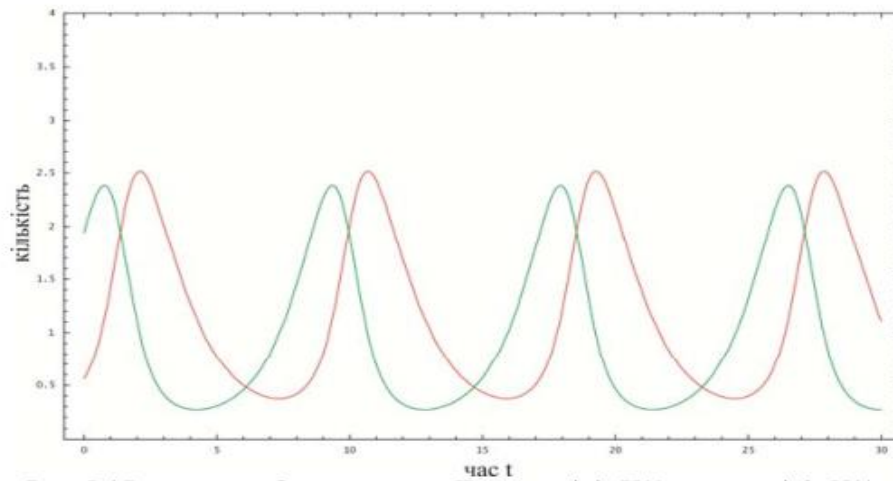


Рис. 5.4 Результат роботи програми. Червона лінія $X(t)$, зелена лінія $Y(t)$

Задача 5* (на самостійну роботу) Чисельно розв'яжіть рівняння з обмеженням максимальної чисельності жертв $Y^m = 2$ та тими ж параметрами і початковими умовами, що в задачі 5.

Лабораторне заняття №7

Визначення обсягу вибірки

Мета: поглиблення теоретичних знань і набуття практичних навичок щодо вибіркового спостереження, розрахунку стандартної та граничної похибок вибірки, визначення обсягу вибірки.

Вибірковим спостереженням називають найбільш науково-обґрунтований вид несучільного спостереження, при якому досліджується не вся сукупність, а лише її частина, відібрана за правилами вибірки і забезпечує отримання даних, що характеризує сукупність в цілому.

Сутність вибіркового спостереження полягає в тому, що з генеральної сукупності способом випадкового вибору відбирається достатньо велика кількість одиниць сукупності й результати їх обстеження поширюються на генеральну сукупність з обумовленою ймовірністю та похибками вибірки.

Розрізняють генеральну та вибірку сукупності.

Вся сукупність одиниць, які підлягають обстеженню за певною ознакою, називається **генеральною**. Частину генеральної сукупності, яку спеціально відібрано для формування змістовних характеристик генеральної сукупності, називають **вибірковою**.

Основні параметри, що характеризують генеральну та вибірку сукупність наведені в таблиці 1.

1. Параметри генеральної і вибіркової сукупностей

№ з/п	Характеристика	Генеральна сукупність	Вибіркова сукупність
1	Обсяг сукупності (чисельність одиниць)	N	n
2	Чисельність одиниць, які мають обстежувану ознаку	M	m
3	Частка одиниць, які не мають обстежувану ознаку	$p = \frac{M}{N}$	$W = \frac{m}{n}$
4	Середній розмір ознаки	$\bar{x} = \frac{\sum x_i}{N}$	$\tilde{x} = \frac{\sum x_i}{n}$
5	Дисперсія кількісної ознаки	$\sigma_x^2 = \frac{\sum (x_i - \bar{x})^2}{N}$	$\sigma_{\tilde{x}}^2 = \frac{\sum (x_i - \tilde{x})^2}{n}$
6	Дисперсія частки	$\sigma_p^2 = pq$	$\sigma_w^2 = W(1 - W)$

Вибірковий метод відрізняється від інших видів не суцільного тим, що відбір одиниць у вибірку сукупність забезпечує рівну можливість

потрапляння кожної одиниці у вибірку шляхом неупередженого суворого відбору. Це здійснюється способами, розробленими математичною статистикою.

Відбір одиниць з генеральної сукупності у вибірку залежно від того, як організований відбір із загального обсягу одиниць спостереження. Розрізняють **два методи відбору**:

- повторний;
- безповторний.

Повторною називається вибірка, за якої кожна раніше відібрана одиниця повертається до генеральної сукупності і може повторно брати участь у вибірці.

Безповторною називається вибірка, за якої кожна раніше відібрана одиниця не повертається до генеральної сукупності і в подальшій вибірці участі не бере.

Безповторний відбір дає більш точні результати порівняно з повторним, оскільки за однокового обсягу вибірки безповторні дослідження охоплюють більше одиниць, ніж повторні.

Розрізняють наступні **види (способи)** вибіркового спостереження:

- власне випадкова вибірка;
- механічна вибірка;
- типова (районована) вибірка;
- серійна (гніздова) вибірка;
- інші види вибірки.

Застосування того чи іншого виду вибірки повинно бути науково обґрунтованим.

Власне випадкова вибірка – це випадковий відбір одиниць з генеральної сукупності за умови однакової можливості для всіх одиниць цієї сукупності потрапити до вибірки. Для забезпечення цього використовують жеребкування або таблицю випадкових чисел.

Випадкова вибірка здійснюється за схемою повторного та без повторного відбору. У процесі **повторного відбору** кожна одиниця бере участь у вибірці стільки разів, скільки відбирається одиниць. За таких умов генеральна сукупність залишається незмінною і для всіх одиниць забезпечується можливість потрапити до вибірки. Але чисельність вибірки в такому випадку може досягти обсягу, який перевищує генеральну сукупність (окремі одиниці враховуються у вибірці не один раз, а декілька). Як наслідок, власне випадкова повторна вибірка практично не використовується в дослідженнях соціально-економічних явищ і процесів.

У випадке безповторного відбору чисельність генеральної сукупності буде змінюватись, оскільки кожна відібрана одиниця в подальшому не повертається до генеральної сукупності. Це підвищує ймовірність решти одиниць потрапити до вибірки, а тому середня похибка вибірка за

безповторного відбору менша, ніж за повторного, і потребує меншого обсягу одиниць спостереження.

Механічна вибірка – це послідовний відбір одиниць через рівні проміжки за їх розташуванням (порядковим, алфавітним, географічним тощо) у генеральній сукупності або в будь-якій іншій послідовності. Зокрема, кожна п'ята одиниця за 20-відсоткового відбору, кожна десята – за 10-відсоткового тощо. Проміжок між відібраними одиницями визначається залежно від прийнятої пропорції і розраховується як частка від ділення чисельності сукупності на обсяг вибірки. Механічний відбір завжди без повторний.

Механічний спосіб забезпечує рівномірність відбору одиниць з усіх частин генеральної сукупності, тобто їх пропорційне представництво (репрезентативність) у вибірковій сукупності.

У статистичні практики досить часто використовують **типову або районовану вибірку**. При цьому досліджувану сукупність спочатку поділяють на однорідні групи (типи, райони) за певною ознакою. Типи можуть утворювати штучно або використовувати ті, що сформувалися природно. З кожної групи випадковим або механічним способом відбирають певну кількість одиниць до вибіркової сукупності.

На практиці типову вибірку можна здійснити наступними методами:

- пропорційно обсягу типових груп;
- непропорційним відбором;
- пропорційно внутрішньо груповій диференціації ознаки.

Обсяг вибірки i -ої типової групи визначають за формулою:

$$n_i = n \frac{N_i}{N}$$

де n_i - обсяг вибірки i -ої типової групи;

n - загальний обсяг вибірки з генеральної сукупності;

N_i - обсяг типової групи;

N - обсяг генеральної сукупності.

Для непропорційної вибірки визначення чисельності здійснюється за формулою:

$$n_i = \frac{n_i}{K}$$

де K – число відокремлених груп.

Для здійснення відбору, пропорційного диференціації ознаки, оптимальний обсяг вибірки i -ої типової групи визначають за формулою:

$$n_i = n \frac{N_i \cdot \sigma_i}{\sum N \cdot \sigma_i}$$

де σ_i - для i -ої групи;

Також часто в практиці статистичного спостереження застосовують **серійну (гніздову) вибірку**, яка передбачає відбір одиниць з генеральної сукупності цілими групами (серіями, гніздами), що потім підлягають суцільному обстеженню. Групи (серії) відбирають за способом власне випадкової неповторної вибірки або способом механічного відбору.

У статистичній практиці серійний відбір одиниць з генеральної сукупності здійснюється в наступних варіантах:

- усі серії мають однакову кількість одиниць;
- усі серії неоднакові за обсягом.

Якщо серії в генеральній сукупності мають однакову чисельність, то середня вибірка визначається як середня арифметична проста із серійних середніх. За неоднакової чисельності серій середня вибірка визначається як середня арифметична зважена.

Більшого поширення набув перший варіант. Серійний відбір простіший в організації та дешевший порівняно з іншими способами, однак за серійного відбору значно порушується рівномірність розподілу відібраних одиниць у межах генеральної сукупності. Крім того, похибка вибірки може бути більша, ніж при інших способах відбору. Для зменшення похибки серійної вибірки збільшують її обсяг.

Точність серійної вибірки залежить від того, наскільки точно середні показники серій будуть репрезентувати генеральну сукупність.

Вибіркова сукупність повинна представляти пропорції генеральної сукупності, тобто бути презентативною (представницькою). З цією метою застосовують різні способи відбору одиниць вибіркової сукупності.

Однак завжди мають **помилки спостереження** - різниця між відповідними характеристиками генеральної і вибіркової сукупностей. Залежно від причин виникнення розрізняють помилки репрезентативності і помилки реєстрації.

Похибки репрезентативності характерні тільки для несцільного спостереження і виникають у зв'язку з тим, що вибірка сукупність за досліджуваною ознакою недостатньо точно відтворює генеральну.

Похибками репрезентативності називають розбіжності між середніми величинами або частками ознаки вибіркової і генеральної сукупностей. Похибки репрезентативності можуть бути систематичними і випадковими.

Систематичні похибки репрезентативності виникають внаслідок порушення принципів відбору одиниць з генеральної сукупності, тобто мають тенденційний характер викривлення величини досліджуваної ознаки в бік її збільшення або зменшення.

Випадкові **похибки репрезентативності** зумовлені тим, що вибірка сукупність не відтворює точно середні й відносні показники генеральної сукупності.

При визначенні помилки вибірки необхідно наблизити показники вибіркової сукупності до показників генеральної сукупності і виявити допустимі граничні відхилення цих показників.

Теоретичною основою вибіркового спостереження є **закон великих чисел**. З точки зору його реалізації відомий статистик **П.Л. Чебишев сформулював теорему**: з ймовірністю, скільки завгодно близькою до одиниці можна стверджувати, що при достатньо великому обсязі вибірки n і обмеженої дисперсії генеральної сукупності σ^2 різниця між вибірковою середньою \tilde{x} і генеральною \bar{x} середніми буде скільки завгодно мала.

Узагальнюючою характеристикою похибки вибірки є **середня (стандартна) похибка репрезентативності μ** , яка виражає середнє квадратичне відхилення вибіркової середньої від генеральної середньої. Величина середньої похибки репрезентативності залежить від σ - показника коливання ознаки в генеральній сукупності, числа відібраних одиниць n , від способу формування вибірки і виражається формулою:

– для середньої:

$$\mu = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

– для частки:

$$\mu = \sqrt{\frac{W(1-W)}{n}}$$

У процесі розрахунку середньої помилки неповторної вибірки необхідно враховувати поправку неповторності відбору:

– для середньої:

$$\mu = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$$

– для частки:

$$\mu = \sqrt{\frac{W(W-1)}{n} \left(1 - \frac{n}{N}\right)}$$

Отже середня помилка вибірки показує, які можливі відхилення характеристик вибіркової сукупності від відповідних характеристик генеральної сукупності.

Поряд із середньою розраховують і **граничну похибку вибірки**, бо стверджувати, що генеральна середня не вийде за межі середньої вибіркової можна стверджувати лише з певним ступенем ймовірності, якій відповідає t -разове значення μ . З урахуванням показника кратності похибки t формула граничної похибки вибірки набуває вигляду:

$$\Delta = t \cdot \mu$$

Звідки

$$t = \frac{\Delta}{\mu}$$

де t – квантиль нормального розподілу, який називають коефіцієнтом довіри. Коефіцієнт довіри залежить від ймовірності, з якою гарантується значення граничної похибки вибірки. Значення ймовірностей P при різних t наводяться в спеціальних таблицях, зокрема деякі з них наведені в таблиці 2.

2. Ймовірність розподілу помилок вибірки

Ймовірність розподілу, P	Значення t	Ймовірність розподілу, P	Значення t
0,890	1,6	0,972	2,2
0,911	1,7	0,979	2,3
0,928	1,8	0,983	2,4
0,942	1,9	0,987	2,5
0,954	2,0	0,997	3,0
0,964	2,1	0,999	4,0

Ці показники означають, що з ймовірністю 0,683 можна стверджувати, що гранична похибка вибірки не перевищує μ , тобто в 68,3% випадків похибка репрезентативності не виходить за межі $\pm\mu$. Отже, в 683 випадках із 1000 похибка репрезентативності не перевищує одного значення середньої похибки. З ймовірністю 0,954 можна стверджувати, що гранична похибка вибірки не перевищує $\pm 2\mu$, з ймовірністю 0,997 можна стверджувати, що гранична похибка вибірки не перевищує $\pm 3\mu$, з ймовірністю 0,999 можна стверджувати, що гранична похибка вибірки не перевищує $\pm 4\mu$.

Гранична похибка вибірки розраховується за вибірковим спостереженням залежно від видів і способів відбору і дає можливість встановити, в яких межах знаходиться значення генеральної середньої або частки. З теореми Чебишева випливає:

$$\bar{x} - \tilde{x} = \pm\Delta$$

$$\bar{x} = \tilde{x} \pm \Delta$$

$$\tilde{x} - \Delta \leq \bar{x} \leq \tilde{x} + \Delta$$

Теорема Бернуллі дає можливість визначити граничну похибку вибірки для альтернативної ознаки. Доведено що, при достатньо великому обсязі вибірки в міру його збільшення ймовірність відхилення між частками ознак у вибірковій і генеральній сукупностях наблизатиметься до одиниці. Тобто з ймовірністю, скільки завгодно близькою до одиниці, можна стверджувати, що при достатньо великому обсязі вибірки вибіркова частка мало відрізняється від її частки в генеральній сукупності.

Межі генеральної частки визначають наступним чином:

$$W - P = \pm\Delta$$

$$W - \Delta \leq P \leq W + \Delta$$

За допомогою формул граничної похибки вибірки визначають:

– довірчі межі генеральної середньої і частки з певною ймовірністю. Для визначення довірчих меж генеральної середньої обчислюють вибірку середню \bar{x} і з прийнятою ймовірністю P граничну похибку вибірки $\Delta_x = t \cdot \mu$;

– ймовірність того, що відхилення між вибірковими й генеральними характеристиками не перевищує визначену величину;

– необхідну чисельність вибірки, яка із заданою ймовірністю забезпечує очікувану точність вибірових показників

Формули для обчислення граничної похибки вибірки для найбільш поширених способів формування вибіркової сукупності наведені в табл.3 .

3. Формули розрахунку граничної помилки вибірки для деяких способів відбору одиниць у вибірку сукупність

Метод відбору Спосіб відбору	Повторний		Безповторний	
	для середньої	для частки	для середньої	для частки
Власне випадковий	$\Delta_x = t \sqrt{\frac{\sigma^2}{n}}$	$\Delta_p = t \sqrt{\frac{W(1-W)}{n}}$	$\Delta_x = t \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$	$\Delta_p = t \sqrt{\frac{W(1-W)}{n} \left(1 - \frac{n}{N}\right)}$
Механічний	*	*	$\Delta_x = t \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$	$\Delta_p = t \sqrt{\frac{W(1-W)}{n} \left(1 - \frac{n}{N}\right)}$
Типовий (пропорційно обсягу типових груп)	$\Delta_x = t \sqrt{\frac{\sigma_i^2}{n}}$	$\Delta_p = t \sqrt{\frac{W_i \cdot (1 - W_i)}{n}}$	$\Delta_x = t \sqrt{\frac{\sigma_i^2}{n} \left(1 - \frac{n}{N}\right)}$	$\Delta_{wx} = t \sqrt{\frac{W_i(1 - W)_i}{n} \left(1 - \frac{n}{N}\right)}$
Серійний (гніздовий)	$\Delta_x = t \sqrt{\frac{\delta_x^2}{r}}$	$\Delta_x = t \sqrt{\frac{\delta_w^2}{r}}$	$\Delta_x = t \sqrt{\frac{\delta_x^2}{r} \left(1 - \frac{r}{R}\right)}$	$\Delta_x = t \sqrt{\frac{\delta_w^2}{r} \left(1 - \frac{r}{R}\right)}$

*механічний відбір завжди без повторний

де Δ_x - гранична похибка вибірки для середньої;

Δ_p - гранична похибка вибірки для частки;

n - кількість одиниць вибіркової сукупності;

N - кількість одиниць генеральної сукупності;

σ^2 - дисперсія вибірки;

t - коефіцієнт довіри;

σ_i^2 - середня з групових дисперсій;

R - кількість серій (гнізд) генеральної сукупності;

r - кількість серій (гнізд) вибіркової сукупності;

δ_i^2 - міжгрупова дисперсія

У процесі здійснення вибіркового спостереження досить часто виникає необхідність визначення мінімального обсягу вибірки, яка з відповідною ймовірністю забезпечує встановлену точність результатів спостереження, бо надмірна чисельність вибірки призводить до затягнення строків дослідження,

додаткових витрат часу і коштів, недостатня ж дає результати з великою похибкою репрезентативності.

Визначення необхідної чисельності вибірки залежить від алгебраїчного перетворення формул граничної похибки вибірки при різних способах відбору.

Для власне випадкової і механічної вибірки виведення формули необхідної чисельності вибірки здійснюється наступним чином: з формули

граничної похибки вибірки для середньої при повторному відборі $\Delta_x = t \sqrt{\frac{\sigma^2}{n}}$

необхідно визначити чисельність вибіркової сукупності n . Для цього необхідно обидві частини даного рівняння підносимо до квадрата, тобто $\Delta_x^2 = \frac{t^2 \cdot \sigma^2}{n}$, звідки

необхідна чисельність вибірки буде дорівнювати: $n = \frac{t^2 \cdot \sigma^2}{\Delta_x^2}$. Формула доводить

залежність чисельності вибірки від граничної похибки вибірки Δ_x , величини коефіцієнта довіри t і варіації (дисперсії) ознаки σ^2 .

Аналогічно виводять формули необхідної чисельності вибірки у випадку обчислення частки ознаки при повторному і без повторному відборах.

У випадку відбору одиниць до вибіркової сукупності з окремих типово однорідних груп, виділених за певною ознакою, варіації групових середніх не існує, і відповідно похибка типової вибірки залежить від середньої величини з групових дисперсій. Таким чином, у випадку типового відбору в формулах похибок вибірки замість загальної дисперсії необхідно використовувати середню з групових $\overline{\sigma_i^2}$ для середньої і $\overline{W(1-W)}$ для частки.

Таким чином, необхідну чисельність вибірки за типовим відбором визначають за формулами:

- для повторного відбору:

$$n = \frac{t^2 \cdot \overline{\sigma^2}}{\Delta^2}$$

- для без повторного відбору:

$$n = \frac{t^2 \cdot \overline{\sigma^2} \cdot N}{\Delta^2 \cdot N + t^2 \cdot \overline{\sigma^2}}$$

У випадку, коли у вибірку сукупність потрапляють утворені в генеральній сукупності однакові за обсягом серії (гнізда), здійснюють суцільне спостереження одиниць у відібраних серіях, і похибка вибірки залежатиме не від числа обстежених одиниць сукупності, а від кількості відібраних серій. Тобто, похибка вибірки залежатиме не від варіації ознаки в усій сукупності, а від варіації серійних середніх, яка вимірюється міжсерійною (між груповою) дисперсією δ^2 та показниками:

- S - число серій у генеральній сукупності,
- s - число відібраних серій.

Необхідну чисельність вибірки для серійного відбору визначають як відбір певної кількості серій, які забезпечують з відповідною ймовірністю необхідну точність дослідження:

– для повторного відбору:

$$s = \frac{t^2 \cdot \delta^2}{\Delta^2}$$

– для без повторного відбору:

$$s = \frac{t^2 \cdot \delta^2 \cdot S}{\Delta^2 \cdot S + t^2 \cdot \delta^2}$$

Формули необхідної чисельності для різних вибірок, отримані в результаті алгебраїчного перетворення формул граничної похибки вибірки наведені в таблиці 4.

4. Формули розрахунку чисельності вибірки від деяких способів відбору одиниць у вибірку сукупність

Метод відбору Спосіб відбору	Повторний		Безповторний	
	для середньої	для частки	для середньої	для частки
Власне випадковий	$n = \frac{t^2 \cdot \overline{\sigma_x^2}}{\Delta_x^2}$	$n = \frac{t^2 \cdot W(1-W)}{\Delta_w^2}$	$n = \frac{t^2 \cdot \overline{\sigma_x^2} \cdot N}{\Delta_x^2 \cdot N + t^2 \cdot \overline{\sigma_x^2}}$	$n = \frac{t^2 \cdot W(1-W) \cdot N}{\Delta_w^2 \cdot N + t^2 \cdot W(1-W)}$
Механічний	*	*	$n = \frac{t^2 \cdot \overline{\sigma_x^2}}{\Delta_x^2}$	$n = \frac{t^2 \cdot W(1-W)}{\Delta_w^2}$
Типовий (пропорційно обсягу типових груп)	$n = \frac{t^2 \cdot \overline{\sigma_x^2}}{\Delta_x^2}$	$n = \frac{t^2 \cdot \overline{W(1-W)}}{\Delta_x^2}$	$n = \frac{t^2 \cdot \overline{\sigma_x^2} \cdot N}{\Delta_x^2 \cdot N + t^2 \cdot \overline{\sigma_x^2}}$	$n = \frac{t^2 \cdot \overline{W(1-W)} \cdot N}{\Delta_x^2 \cdot N + t^2 \cdot \overline{W(1-W)}}$
Серійний (гніздовий)	$r = \frac{t^2 \cdot \delta_x^2}{\Delta_x^2}$	$r = \frac{t^2 \cdot \overline{W_r(1-W_r)}}{\Delta_x^2}$	$r = \frac{t^2 \cdot \delta_x^2 \cdot R}{\Delta_x^2 \cdot R + t^2 \cdot \delta_x^2}$	$r = \frac{t^2 \cdot \overline{W(1-W)} \cdot R}{\Delta_x^2 \cdot R + t^2 \cdot \overline{W_r(1-W_r)}}$

Лабораторне заняття №8

Розрахунок кореляційних зав'язків між ознаками

Мета: оволодіти методами обчислення показників зв'язку між Ознаками.

Навчитись обчислювати коефіцієнт кореляції (r) методом малих вибірок.

З метою підвищення ефективності селекції одночасно за кількома ознаками рекомендується насамперед враховувати їхню взаємну зумовленість, тобто кореляцію. Для цього виникла необхідність математичного усвідомлення

такого явища та вираження його числовим показником, коефіцієнтом, за величиною якого можна було б говорити про тісноту чи силу зв'язку між окремими ознаками.

Коефіцієнт кореляції - числовий показник простої лінійної кореляції, який описує напрям і тісноту зв'язку між досліджуваними величинами, вимірює зв'язок лише при лінійній формі залежності, а його абсолютне значення знаходиться в межах від -1 до $+1$. при значенні $r = 0$ – зв'язок відсутній; при $+1$ – пряма кореляційна залежність; а при -1 – зворотня.

Прийнято вважати, що залежність низька, якщо $r=0,1-0,3$; середня при $r=0,4-0,6$ і висока – при $0,7$ і більше. при коефіцієнті кореляції $r=0,1$ мінливість другої ознаки лише на 1% залежить від мінливості першої, а в 99% випадках – від випадкових факторів.

Розрізняють *фенотипову, генотипову та середовищну* кореляції.

Фенотипова кореляція дає змогу встановити зв'язок між різними господарськи корисними ознаками і використати їх з метою селекції.

Генотипова кореляція – це форма зв'язку між двома ознаками, зумовлена адитивною дією і взаємодією генів (плейотропія, епістаз тощо). генотипова кореляція визначає вплив спадкових факторів на ступінь фенотипової кореляції.

Паратипова (середовищна) кореляція зумовлена силою і спрямованістю впливу умов середовища на дві ознаки, що вивчаються.

Розрахунок коефіцієнта кореляції для малих вибірок.

Приклад. Визначено ураженість льону фузаріозом (ряд y) у залежності від інтервалу між посівами (у роках) на одному і тому ж полі сприйнятливих до грибних патогенів (фузаріозу) сортів льону (ряд x).

Провести кореляційний аналіз даних.

Для цього використовуємо допоміжну таблицю 5, шаблон якої поданий нижче.

5. Форма запису для розрахунку коефіцієнта кореляції між інтервалом посівів (x) льону на одному і тому ж полі та ураженістю фузаріозом (y)

Номери пар	Значення ознак		x ²	y ²	xy
	x, роки	y, %			
1.	1	88	1	7744	88
2.	2	76	4	5776	152
3.	2	70	4	4900	140
4.	7	5	49	25	35
5.	6	12	36	144	72
6.	5	28	25	784	140
7.	3	45	9	2025	135
8.	4	45	16	2025	180
9.	6	9	36	81	54
10.	3	62	9	3844	186
n=10	$\sum x = 39$	$\sum y = 440$	$\sum x^2 = 189$	$\sum y^2 = 27348$	$\sum xy = 1182$

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{C_x \cdot C_y}};$$

$$\text{де: } C_x = \sum x^2 - \frac{(\sum x)^2}{n} = 189 - \frac{(39)^2}{10} = 36,9$$

$$C_y = \sum y^2 - \frac{(\sum y)^2}{n} = 27348 - \frac{(440)^2}{10} = 7988$$

$$r = \frac{1182 - \frac{39 \cdot 440}{10}}{\sqrt{36,9 \cdot 7988}} = \frac{-534}{543} = -0,98$$

Розраховуємо помилку коефіцієнта кореляції (s_r), критерій значущості (t_r), довірчий інтервал та перевіряємо нульову гіпотезу (H_0) про відсутність Кореляційного зв'язку між явищами.

$$s_r = \sqrt{\frac{1 - r^2}{n - 2}} = \sqrt{\frac{1 - 0,98^2}{10 - 2}} = 0,07$$

$$t_r = \frac{r}{s_r} = \frac{0,98}{0,07} = 14,00$$

$$v = n - 2 = 10 - 2 = 8 \quad t_{05} = 2,31$$

$$r \pm t_{05}s_r = -0,98 \pm 2,31 \times 0,07 = -0,98 \pm 0,16$$

$$r \pm t_{05}s_r = (-1,14 \div 0,82)$$

За критерієм t_r ($t_{факт} > t_{05}$) і довірчим інтервалом, який не ключає нульового значення, кореляція суттєва, отже, нульова гіпотеза на 5%-му рівні відкидається.

Контрольні запитання:

1. Що таке кореляція ?
2. Яка різниця між позитивною і негативною кореляціями
3. Які можливі значення коефіцієнта кореляції ?
4. Які значення коефіцієнта кореляції слід вважати високими, середніми і чому ?
5. Чому дорівнює коефіцієнт кореляції при повному кореляційному зв'язку?
6. Як треба розуміти нульову гіпотезу стосовно коефіцієнту кореляції?

Лабораторне заняття №9

Розрахунок коефіцієнту регресії

Мета: оволодіти методами розрахунку коефіцієнту регресії. Навчитись обчислювати коефіцієнт регресії.

Основи регресійного аналізу

При обробці експериментальних даних і побудові математичних моделей використовують теорію кореляційно-регресійного аналізу (theory of correlation and regression analysis).

Регресійний аналіз (regression analysis) встановлює математичну модель, що зв'язує залежну змінну y з досліджуваною змінною x , тобто дозволяє отримати залежність виду $y = f(x)$ - рівняння парної регресії.

Значення змінної y може залежати відразу від декількох змінних x_1, x_2, \dots, x_n . У результаті обробки таких статистичних даних можна отримати залежність виду $y = f(x_1, x_2, \dots, x_n)$ - рівняння множинної регресії. Всі рівняння регресії поділяються на лінійні і нелінійні.

Функцію, апроксимуючу досвідчені дані, називають теоретичною функцією. При парній залежності експериментальні дані можуть бути апроксимовані за допомогою наступних функцій:

- прямою лінією $y = a \cdot x + b$;
- параболою другого порядку $y = a \cdot x^2 + b \cdot x + c$;
- гіперболою $y = \frac{a \cdot x + b}{x}$;
- логарифмічною функцією $y = a \ln x + b$;
- степеневою функцією $y = a \cdot x^b$;
- показовою функцією $y = a \cdot x^x$;
- арифметичною прогресією $y = a + (n - 1)d$;
- геометричною прогресією $y = a \cdot q^{n-1}$;
- алгебраїчним поліномом, тобто рядом Маклорена:

$$y = A_0x^0 + A_1x^1 + A_2x^2 + A_3x^3 + \dots,$$

в якому коефіцієнти ряду визначаються за формулами:

$$A_0 = f(x), A_1 = \frac{f'(0)}{1!}, A_2 = \frac{f''(0)}{2!}, \dots \text{ і т.д.}$$

- тригонометричним рядом, тобто рядом Фур'є

$$y = \frac{a}{2} + \sum_{n=0}^{\infty} a_n \cos nx + b_n \sin nx,$$

та іншими функціями.

Для двухфакторної регресійної залежності досвідчені дані можуть апроксимувати наступними функціями:

- площиною $z = ax + by + c$;

- параболоїдом другого порядку $z = ax^2 + by^2 + cx + dy + c$;

- гіперболоїдом $z = \frac{a}{bx + cy + d}$ і т.д.

У загальному випадку для n -мірного простору і n змінних рівняння регресії другого порядку виглядає так:

$$y = B_0 + \sum_{i=1}^n B_i x_i + \sum_{i < j} B_{ij} x_i x_j + \sum_{i=1}^n B_{ii} x_i^2 + \dots,$$

де y - досліджувана ознака (параметр) як функція багатьох змінних;

x_i - фактори, що роблять вплив на параметр;

B_i - приватні коефіцієнти регресії, що показують вплив фактора x_i на досліджувану ознаку;

B_{ij} - коефіцієнти, що характеризують подвійний (парний) вплив факторів x_i і x_j на досліджувану ознаку.

Множинна регресія

Ряд завдань автомобільного транспорту вимагають побудови множинних регресій (multiple regression) виду:

$$Y = f(x_1, x_2, \dots, x_n)$$

Так, якщо будемо розглядати діяльність АТП, то одним з основних показників його роботи є коефіцієнт випуску автомобілів на лінію - , на формування якого впливає велика кількість факторів.

Основними з них будуть:

X_1 - забезпеченість водіями;

X_2 - забезпеченість ремонтними робітниками;

X_3 - забезпеченість інженерними кадрами;

X_4 - забезпеченість запасними частинами і т.д.

Для розглянутого випадку:

$$\alpha_B = y = f(x_1, x_2, x_3, x_4)$$

тобто маємо справу з побудовою множинної регресії.

У більш загальному вигляді лінійна множинна регресія записується так:

$$y = b_0 x_0 + b_1 x_1 + b_2 x_2 + \dots + b_m x_m = \sum_{i=0}^m b_i x_i,$$

де y - теоретичне значення результативної ознаки;

x_i - аргументи (фактори);

m - число досліджуваних факторів;

b_i - коефіцієнти регресії, що показують ступінь впливу кожного з факторів на функцію;

b_0 - залишковий член, що характеризує середнє значення функції.

Для отримання таких регресій необхідно скористатися статистичними (експериментальними) даними, які в загальному випадку можна представити таблицею 6. Процес побудови множинної регресії пояснимо на прикладі.

6. Регресія зі статистичними даними

		$J = 1, 2, \dots, n$ – (стовпець)						
		1	2	...	J	...	n	Σ
$I = 1, 2, \dots, m$ (рядок)	X_1	X_{12}	X_{12}	...	X_{1J}	...	X_{1n}	ΣX_{1J}
	X_2	X_{21}	X_{22}	...	X_{2J}	...	X_{2n}	ΣX_{2J}

	X_i	X_{i1}	X_{i2}	...	X_{iJ}	...	X_{in}	ΣX_{iJ}

	X_m	X_{m1}	X_{m2}	...	X_{mJ}	...	X_{mn}	ΣX_{mJ}
	Y	Y_1	Y_2	...	Y_J	...	Y_n	ΣY_J

Приклад. Для випадку $y=f(x_1, x_2)$ лінійна регресія має вигляд $y=a+bx_1+cx_2$. При обчисленні коефіцієнтів a, b, c скористаємося методом найменших квадратів. Тоді будемо мати:

$$z = \sum_{j=1}^n (y_j - a - bx_{1j} - cx_{2j})^2 \Rightarrow \min$$

Розрахуємо приватні похідні виразу 2.26 за параметрами :

$$\left. \begin{aligned} \frac{\partial z}{\partial a} &= 2 \sum_{j=1}^n (y_j - a - bx_{1j} - cx_{2j})(-1) = 0; \\ \frac{\partial z}{\partial b} &= 2 \sum_{j=1}^n (y_j - a - bx_{1j} - cx_{2j})(-x_{1j}) = 0; \\ \frac{\partial z}{\partial c} &= 2 \sum_{j=1}^n (y_j - a - bx_{1j} - cx_{2j})(-x_{2j}) = 0. \end{aligned} \right\}$$

Перетворюючи систему, отримаємо

$$\left. \begin{aligned} na + b \sum_{j=1}^n x_{1j} + c \sum_{j=1}^n x_{2j} &= \sum_{j=1}^n y_j; \\ a \sum_{j=1}^n x_{1j} + b \sum_{j=1}^n x_{1j}^2 + c \sum_{j=1}^n x_{1j} x_{2j} &= \sum_{j=1}^n x_{1j} y_j; \\ a \sum_{j=1}^n x_{2j} + b \sum_{j=1}^n x_{1j} x_{2j} + c \sum_{j=1}^n x_{2j}^2 &= \sum_{j=1}^n x_{2j} y_j. \end{aligned} \right\}$$

Система представляє систему лінійних рівнянь, вирішуючи яку одним з відомих методів, знайдемо значення невідомих параметрів a, b, c .

Методом Крамера параметри a, b, c визначаються з виразів:

$$a = \frac{D_a}{D}; \quad b = \frac{D_b}{D}; \quad c = \frac{D_c}{D},$$

де D - головний визначник системи лінійних рівнянь;

D_a - визначник системи рівнянь, в якому стовпець коефіцієнтів при a замінений стовпцем вільних членів;

D_b - визначник системи, в якій стовпець коефіцієнтів при b замінений стовпцем вільних членів;

D_c - визначник системи, в яких стовпець коефіцієнтів при c замінений стовпцем вільних членів.

Якщо потрібно побудувати математичну модель виду $y = \sum_{i=0}^m b_i x_i$, то для обчислення коефіцієнтів b_i ($i = 0, 1, 2, \dots, m$) знову скористаємося методом найменших квадратів, отримуємо систему нормальних рівнянь виду

$$\left. \begin{aligned} nb_0 + b_1 \sum x_{1j} + \dots + b_m \sum x_{mj} &= \sum y_i; \\ b_0 \sum x_{1j} + b_1 \sum x_{1j}x_{2j} + \dots + b_m \sum x_{1j}x_{mj} &= \sum x_{1j}y_i; \\ \dots &\dots \\ b_0 \sum x_{mj} + b_1 \sum x_{1j}x_{mj} + \dots + b_m \sum x_{mj}x_{mj} &= \sum x_{mj}y_i. \end{aligned} \right\}$$

Вирішуючи систему відомим методом, знайдемо значення параметрів b_i рівняння регресії. Найбільш поширеним методом вирішення системи (2.26) є метод звернення матриць, запрограмований у ряді стандартних програм апроксимації.

У матричній формі система запишеться

$$XB = Y, \quad (2.30)$$

де X - вихідна матриця;

B - матриця вихідних значень параметрів;

Y - матриця вільних членів.

$$X = \begin{pmatrix} n & \sum x_{1j} & \dots & \sum x_{mj} \\ \sum x_{1j} & \sum x_{1j}x_{2j} & \dots & \sum x_{1j}x_{mj} \\ \dots & \dots & \dots & \dots \\ \sum x_{mj} & \sum x_{1j}x_{mj} & \dots & \sum x_{mj}x_{mj} \end{pmatrix}; \quad B = \begin{pmatrix} B_0 \\ B_1 \\ \dots \\ B_m \end{pmatrix}; \quad Y = \begin{pmatrix} \sum y_i \\ \sum x_{1j}y_i \\ \dots \\ \sum x_{mj}y_i \end{pmatrix}.$$

При вирішенні системи лінійних рівнянь матричним способом необхідно обчислити зворотну матрицю (X^{-1}) і помножити її на праву і ліву частини рівняння (2.31), отримаємо

$$\bar{X}^{-1}XB = X^{-1}Y.$$

Так як $\bar{X}X = E$ і $EB = B$, то остаточний вираз для визначення матричним способом параметрів B_i прийме вигляд:

$$B = X^{-1}Y. (2.33)$$

Контрольні запитання:

1. Призначення регресійного аналізу.
2. Види парних регресій.
3. Способи лінеаризації нелінійних залежностей.
4. Множинна лінійна регресія.
5. Послідовність обчислення параметрів множинної регресії методом Крамера.
6. Послідовність обчислення параметрів множинної регресії матричним способом.

Лабораторне заняття №10

Розрахунок однофакторного дисперсійного аналізу

Мета: отримати навички вивчення статистичного впливу одного або декількох факторів на результативну ознаку.

У випадку, коли число вибірок в дослідженнях більше двох, метод попарних порівнянь вимагає великої обчислювальної роботи. У цьому випадку використовують метод дисперсійного аналізу, заснований на розкладанні загальної дисперсії статистичного комплексу на її складові компоненти.

При утворенні дисперсійних комплексів необхідно дотримуватися двох умов:

- 1) діючі на ознаку регульовані фактори повинні бути незалежні один від одного (багатофакторні комплекси);
- 2) вибірки, що групуються в статистичний комплекс, повинні здійснюватися за принципом рендомізації.

Дисперсійний аналіз використовує властивість адитивності дисперсії випадкової величини, що обумовлено дією незалежних факторів. У залежності від числа джерел дисперсії розрізняють *однофакторний* та *багатофакторний* дисперсійний аналіз.

Однофакторний дисперсійний аналіз дозволяє виявити вплив тільки одного фактора. При цьому загальне варіювання результативної ознаки розкладається на два компоненти – варіювання варіантів та випадкове варіювання: $C_y = C_v + C_z$

Статистичний аналіз даних однофакторного експерименту проводять у три етапи:

1 етап. Складають розрахункову таблицю, розміщуючи у ній вихідні дані за рядками та стовпчиками, визначають суми та середні за варіантами, загальну суму та середнє значення результативної ознаки в експерименті:

Варіанти	Вихідні дані, X	Число спостережень, n	Суми за варіантами, V	Середні за варіантами, \bar{x}
1	$X_1, X_2, X_3, \dots, X_{1n}$	n_1	V_1	\bar{x}_1
2	$X_{11}, X_{12}, X_{13}, \dots, X_{2n}$	n_2	V_2	\bar{x}_2
3	$X_{21}, X_{22}, X_{23}, \dots, X_{3n}$	n_3	V_3	\bar{x}_3
...
l	$X_{l1}, X_{l2}, X_{l3}, \dots, X_{ln}$	n_l	V_l	\bar{x}_l
Загальна сума		$N = \sum n$	$\sum X = \sum V$	$\bar{x} = \sum X / N$

2 етап. Розраховують суми квадратів відхилень та f-критерій за такими формулами та заносять до таблиці:

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	$F_{\text{факт.}}$	$F_{\text{теор.}}$
Загальна C_y	$\sum X^2 - C$	$N-1$	-	-	-
Варіантів C_v	$\sum \frac{v^2}{n} - C$	$l-1$	$s_v^2 = \frac{C_v}{l-1}$	$\frac{s_v^2}{s^2}$	за таблицею
Остаток C_z	$C_y - C_v$	$N-l$	$s^2 = \frac{C_z}{N-l}$	-	-

Коригуючий фактор (C) розраховують за формулою:

$$C = \frac{(\sum X)^2}{N}$$

3 етап. Визначають помилку досліду та істотність окремих різниць.

помилка середньої: $S_x = \sqrt{\frac{s^2}{n}}$

помилка різниці середніх: $S_d = \sqrt{\frac{2s^2}{n}}$

$$HP_{05} = t_{05} \times S_d$$

4 етап. Результати експерименту та статистичної обробки даних записують до підсумкової таблиці.

Приклад. Встановлення селекційної цінності п'яти сортів за врожайністю зерна.

1. Складаємо розрахункову таблицю:

Варіанти	Врожайність зерна, г/ділянку				Число спостережень, n	Суми за варіантами, V	Середні за варіантами, \bar{x}
1(st)	454	470	430	500	4	1854	463,5
2	502	550	490	507	4	2049	512,2
3	601	670	550	607	4	2428	607,0
4	407	412	475	402	4	1696	424,0
5	418	470	460	412	4	1760	440,0
Загальна сума					$N = \sum n = 20$	$\sum X=9787$	$\bar{x} = 489,4$

2. Розраховуємо суми квадратів відхилень та F-критерій.

$$C = \frac{(\sum X)^2}{N} = \frac{(9787)^2}{20} = 4789268$$

$$\sum X^2 - C = (454^2 + 470^2 + \dots + 412^2) - 4789268 = 104941$$

$$\sum \frac{V^2}{n} - C = \frac{(1854^2 + 2049^2 + \dots + 1760^2)}{4} - 4789268 = 86961$$

$$C_y - C_v = 104941 - 86961 = 17980$$

$$F_{\text{факт.}} = \frac{21740}{1199} = 18,13$$

Одержані результати заносимо до таблиці дисперсійного аналізу.

Дисперсія	Сума квадратів	Ступені свободи	Середній квадрат	$F_{\text{факт.}}$	$F_{\text{теор.}}$
Загальна C_y	104941	19	-	-	-
Варіантів C_v	86961	4	21740	18,13	3,06
Залишок C_z	17980	15	1199	-	-

Здійснюємо оцінку нульової гіпотези. Оскільки $F_{\text{факт.}} > F_{\text{теор.}}$, то нульова гіпотеза відкидається, тобто у досліді є істотні відмінності за варіантами на 5%-му рівні значущості (t_{05}).

3. Встановлюємо помилку досліду та істотність різниць:

помилка середньої: $S_x = \sqrt{\frac{S^2}{n}} = \sqrt{\frac{1199}{4}} = 17.3 \text{ г}$

$$\text{помилка різниці середніх: } S_d = \sqrt{\frac{2S^2}{n}} = \sqrt{\frac{2 \times 1199}{4}} = 24.5 \text{ г}$$

$$\text{НІР}_{05} = t_{05} \times S_d = 2,13 \times 24,5 = 52,2 \text{ г,}$$

$$\text{або НІР у відсотках: } \text{НІР}_{05} = \frac{t_{05} \times S_d}{x} \times 100 = \frac{52.2}{489.4} \times 100 = 10.7\%$$

4. Результати експерименту та статистичної обробки даних запишемо до підсумкової таблиці.

Сорт	Врожайність зерна, г/ділянку	Різниця зі стандартом		Група за врожайністю зерна
		г/ділянку	%	
Саратівська 55 (st)	463,5	-	-	1 (st)
Саратівська 29	424,0	-39,5	-8,5	1
Саратівська 46	440,0	-23,5	-5,1	1
Саратівська 58	512,2	48,7	10,5	1
К 503	607,0	143,5	30,9	2

Висновок: у сортовипробуванні встановлені істотні різниці за врожайністю зерна між сортами м'якої пшениці (варіантами). Істотно перевищив за врожайністю зерна сорт-стандарт лише сорт Л 503.

Контрольні запитання:

1. Що вивчають методом дисперсійного аналізу?
2. Що таке дисперсія, та на які види вона розподіляється?
3. Що таке дисперсійний комплекс, і які види комплексів бувають?
4. Що таке градація, і які комплекси бувають у залежності від розмірів градацій?
5. За якою формулою визначається показник сили впливу фактора, дія якого вивчається?
6. Як визначити достовірність розрахованого показника сили впливу фактора?

Рекомендована література

1. Барсов В.І., Костерна О.Ю. Методи обчислення та моделювання на EOM=Computational Methods and Simulation Techniques: навч. посіб. до лаб. Практикуму. Нац. аерокосм. ун-т ім. М. Є. Жуковського "Харків. авіац. ін-т". Харків, 2020. 168 с.
2. Осадча Ю.В. Математичні методи в біології. Київ, 2017. 601 с.
3. Прилуцький Ю.І., Ільченко О.В., Цимбалюк О.В., Костерін С.О. [Статистичні методи в біології](#). Київ: Наукова думка, 2017. 211 с.
4. Присажнюк О.І., Клименко Н.М., О.В. Полуніна. Методологія і організація наукових досліджень в сільському господарстві та харчових технологіях. Вінниця: Нілан-ЛТД. 2021. 300 с.
5. Швець Е.А., Кісарін О.О. Комп'ютерне моделювання фізіологічних систем організму. Запоріжжя: ЗДІА, 2009. 175 с.
6. Малюк В.Г., Борзенков Б.І. Моделювання в біології та медицині. – Харків: Наук.-метод. центр вищ. освіти, 2005. 212 с.
7. Оглобля О.В., Мірошніченко М.С., Костерін С.О. Комп'ютерне моделювання в біології. К.: Видавничий центр «Азбука», 2012. 120 с.
8. Оглобля О.В., Мірошніченко М.С., Костерін С.О. «Комп'ютерне моделювання в біології», К.: Фітосоціоцентр, 2006. 66 с.
9. Прилуцький Ю.І., Оглобля О.В., Склярів Ю.П., Богуцька К.І. «Математичні моделі в біології», К.: ВПЦ КНУ, 2002. 64 с.
10. Хусаїнов Д.Я., Харченко І.І., Шатирко А.В. Введення в моделювання динамічних систем. К.: Київський національний університет імені Тараса Шевченка, 2010. 130 с.