

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВАДИМА ГЕТЬМАНА

Навчально-науковий інститут
«Інститут інформаційних технологій в економіці»

Кафедра математичного моделювання та статистики

Освітньо-професійна програма

Економічна кібернетика

Галузь знань

05 Соціальні та поведінкові науки

Спеціальність

051 Економіка

Форма навчання: очна (денна)

КВАЛІФІКАЦІЙНА БАКАЛАВРСЬКА РОБОТА

на тему «Моделювання прибутковості культур в агропромисловому комплексі»

(назва теми)

здобувача Кузьменко Марини Олександрівни

(ПІБ, підпис)

Науковий керівник: кандидат фізико-математичних наук,
доцент Юнькова О. О.

(науковий ступінь, учене звання, ПІБ)

(підпис)

Робота допущена до захисту перед екзаменаційною комісією
з атестації здобувачів вищої освіти (ЕК)

Завідувач кафедри кандидат фізико-математичних наук,
професор Великоіваненко Г.І.

(підпис)

Київ 2023

ЗМІСТ

<u>ВСТУП</u>	3
<u>РОЗДІЛ 1. СТАН СІЛЬГОКОСПОДАРСЬКОГО КОМПЛЕКСУ УКРАЇНИ У ВОЄННИЙ ПЕРІОД</u>	6
<u>1.1 Сільське господарство та економічний розвиток країни</u>	6
<u>1.2 Агропромисловий комплекс в економіці України</u>	11
<u>1.3 Структура землі та основні сільськогосподарські товари України</u>	15
<u>РОЗДІЛ 2. ПОБУДОВА ТА АНАЛІЗ ЕКОНОМЕТРИЧНИХ МОДЕЛЕЙ</u>	25
<u>2.1 Побудова та перевірка значущості економетричної моделі</u>	25
<u>2.2 Діагностика лінійної регресії та відбір предикторів</u>	29
<u>2.3 Дослідження прогностичних та аналітичних якостей моделей</u>	33
<u>РОЗДІЛ 3. МОДЕЛЮВАННЯ ТА ОЦІНЮВАННЯ ПРИБУТКОВОСТІ КУЛЬТУР В АГРОПРОМИСЛОВОМУ КОМПЛЕКСІ У ВОЄННИЙ ЧАС</u>	37
<u>3.1 Побудова економетричних моделей за трьома культурами та оцінка їх ефективності</u>	37
<u>3.2 Дослідження аналітичних якостей моделей прибутковості трьох культур</u>	43
<u>ВИСНОВКИ</u>	55
<u>ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ</u>	58
<u>ДОДАТКИ</u>	66

ВСТУП

Агропромисловий комплекс, як невід’ємна складова економіки України, що забезпечує її економічний розвиток, є також частиною глобального постачання продовольства. Однією з ключових і основних компонент успіху в цій сфері є ефективне використання наявних ресурсів, насамперед у рослинництві, яке за вартістю продукції перевищує тваринництво і є його сировинною базою. При цьому врахування прибутковості культур є вирішальним для прийняття обґрунтованих управлінських рішень та оптимізації виробничих процесів. Найбільш дієвим засобом обґрунтування таких рішень є економіко-математичне моделювання.

Існує безліч факторів, які впливають на економічні результати вирощування різних сільськогосподарських культур, тому моделювання прибутковості культур в агропромисловому комплексі є складною та багатогранною задачею. В умовах війни не лише скорочуються посівні площі, руйнується сільськогосподарська інфраструктура, знищується та викрадається техніка, воєнні події ускладнюють процеси вирощування та збуту готової продукції, викликають додаткові затрати та негативно впливають не лише на економічні показники сільгоспідприємств, але й на економіку нашої держави. Підприємства, які зазнали значних втрат у воєнний період, потребують додаткової підтримки з боку фінансових установ. Тому особливо актуальною задачею для кредиторів є надійність позичальника, яка гарантується результатами його економічної діяльності. У зв’язку з цим побудова моделей прогнозування прибутковості окремих культур є актуальною з точки зору оцінювання потенціалу позичальника і прийняття обґрунтованого рішення щодо надання йому кредиту. З іншого боку, моделі з високими аналітичними якостями можуть стати підґрунтям для правильного вибору культур і підвищення їх прибутковості сільгоспідприємцем. Зазначені обставини зумовлюють актуальність теми даної роботи.

Оскільки продовольча безпека є першочерговою задачею будь-якого суспільства, то моделюванню проблем агропромислового сектору присвячено численні наукові дослідження. В багатьох публікаціях проаналізовано врожайності та прибутковості культур в місцях з не дуже сприятливими кліматичними умовами, а саме: у горській місцевості [1], у дощовій та спекотній Індії [2] та Ефіопському нагір'ї [3]. Роберт Фінгер та Наді Ел Бенні у [4] проводять аналіз доходу від агропромисловості в європейських країнах. Основний напрям наукових робіт у цій сфері присвячений саме аналізу та різним методам прогнозування подальшого розвитку агропромислового комплексу [5], [6]. Наукове видання від Жан-Клод Кадер та Мішель Дельсені, яке займається поширенням бізнес-літератури, опублікувало роботу, що прямо корелює з даною темою. В одному із розділів цієї роботи розглядається моделювання врожайності культур, описано фактори, які, на думку авторів, найбільше впливають на обсяг врожаїв, та спрогнозовано можливі врожаї різних культур [7]. В нашій країні питання моделювання та прогнозування врожайності сільськогосподарських культур розглядали Грицюк П.М. та Бачишина Л.Д. [8], які застосовували трендові моделі. В роботах Грицюка П.М. [9] розглядалися також інші методи середньострокового та довгострокового прогнозування врожайності зернових. Кобець С.П. та Тесьолкін О.І. [9], а також Сотніков В.С. та Львовчикіна О.О. [10] моделювали залежність врожайності від метеоумов, Климко О.Г. [11] прогнозував врожайність культур адаптивними методами динамічної регресії. В роботі Лещенко О.Л. [12] на підставі економіко-математичного дослідження були отримані висновки щодо оптимізації структури посівних площ, асортименту та обсягів реалізації продукції для підвищення прибутковості овочівництва при наявних виробничих ресурсах.

Метою кваліфікаційної бакалаврської роботи є визначення методів, моделей та інструментів, які допоможуть банкам приймати обґрунтовані рішення щодо кредитування агропромислових підприємств, а сільгоспвиробникам збільшувати прибутки за рахунок правильного вибору культур та підвищення їхньої прибутковості.

Відповідно до поставленої мети визначені такі завдання:

- 1) розкрити сутність агропромислового комплексу;
- 2) проаналізувати вплив факторів, таких як врожайність, вартість виробництва, ринкові ціни, витрати на вирощування та інші, на прибутковість культур;
- 3) розробити математичну модель для дослідження прибутковості різних культур в агропромисловому комплексі;
- 4) Перевірити прогностичні та аналітичні якості побудованих моделей.

Об'єктом дослідження кваліфікаційної бакалаврської роботи є агропромисловий сектор України.

Предметом дослідження є методи економіко-математичного моделювання для аналізу та прогнозування прибутковості пшениці, кукурудзи та соняшнику.

У даній роботі були використанні методи аналізу та моделювання прибутковості культур в агропромисловому комплексі, а саме аналітичні та статистичні методи (збирання та систематизація інформації про вирощування культур в агропромисловому секторі, проведення її статистичного аналізу), економіко-математичне моделювання (розроблення економетричних моделей, які враховують вплив різних факторів на прибутковість культур).

Інформаційна база дослідження. Наукові праці вітчизняних та зарубіжних фахівців з економіко-математичного моделювання задач агропромисловості, інформація з інтернет-сайтів. Статистична база дослідження надана ПроКредит банком при проходженні автором виробничої практики та включає в себе 70 підприємств, які вирощують пшеницю, кукурудзу та соняшник.

Структура роботи. Робота складається зі вступу, трьох розділів, висновків, списку використаних джерел та додатків.

РОЗДІЛ 1

СТАН СІЛЬГОКОСПОДАРСЬКОГО КОМПЛЕКСУ УКРАЇНИ У ВОЄННИЙ ПЕРІОД

1.1 Сільське господарство та економічний розвиток країни

Економічне зростання країни, що характеризується розвитком промисловості, супроводжується зменшенням значення сільського господарства. Основну причину цього вказав німецький статистик 19-го століття Ернст Енгель. Він виявив, що при збільшенні доходів населення, частка доходу, яка витрачається на їжу, зменшується [13]. Звідси випливає, що зі збільшенням доходів, менша частка загальних ресурсів суспільства потрібна для виробництва тієї кількості їжі, яка потрібна населенню.

Цей факт здивував би більшість економістів на початку 19 століття, які хвилювалися, що обмежена доступність землі в густонаселених районах Європи визначатиме здатність континенту прогодувати зростаюче населення. Їхній страх ґрунтувався на так званому законі спадної віддачі: за певних умов збільшення кількості праці та капіталу, застосованих до фіксованої кількості землі, призведе до непропорційного збільшення виробництва їжі. Цей принцип справедливий, але класичні економісти не можуть передбачити, наскільки зміняться сучасні умови та методи виробництва. Деякі зміни відбулися в сільському господарстві, інші відбулися в інших секторах економіки, але мали великий вплив на постачання продовольства.

Озираючись на історію більш розвинених країн, можна побачити, що сільське господарство відігравало важливу роль у їхньому процвітанні. По-перше, якщо розвиток має відбутися, сільське господарство має мати можливість виробляти надлишок їжі для підтримки зростаючої робочої сили поза фермами. Оскільки їжа важливіша для життя, ніж послуги, які надають торговці, банкіри чи

фабрики, економіка не може перенаправлятися на таку діяльність, якщо їжі немає в достатній кількості для обміну чи продажу для підтримки людей, які займаються цією діяльністю. Без доступу до продовольства через міжнародну торгівлю країна, як правило, не розвиває промисловість, доки її сільськогосподарські райони не зможуть забезпечувати свої міста продовольством в обмін на продукцію своїх заводів.

Економічний розвиток також вимагає зростання робочої сили [14]. У сільськогосподарській країні більшість необхідної робочої сили має надходити від сільського населення. Отже, сільське господарство має не лише забезпечувати надлишки їжі для міст, воно також має бути здатним виробляти більше їжі з відносно невеликою кількістю праці. Цього можна досягти шляхом заміни людської сили силою тварин або шляхом поступового впровадження працевзберігаючих машин.

Сільське господарство також може бути джерелом капіталу для промислового розвитку, оскільки воно забезпечує надлишок, який можна перетворити на кошти, необхідні для купівлі промислового обладнання або будівництва доріг і надання суспільних послуг.

З цих причин країнам, які прагнуть розвивати свою економіку, можна порадити приділяти значну увагу сільському господарству. Досвід країн, що розвиваються, показує, що продуктивність сільського господарства можна значно підвищити, якщо робити адекватні інвестиції в іригаційні системи, дослідження, добрива, пестициди та гербіциди.

На щастя, багато досягнень прикладної науки не вимагають великих капіталовкладень, хоча можуть знадобитися розширені можливості збуту та транспортування, щоб донести продукцію ферми до широкого загалу.

Визначення пріоритетів у сільському господарстві ускладнюється тим, що переважна частина сільськогосподарського виробництва та прибутків від нього зосереджена в певних областях, а не розкидана по всій країні. Фермери, що залишилися, не змогли збільшити своє виробництво і опинилися фактично в невігідному становищі через зниження цін на сільськогосподарську продукцію. На

це запитання немає простих відповідей, але країни, що розвиваються, повинні це усвідомити; економічний прогрес збігається з довготривалою відсталістю, як це можна побачити в деяких частинах південної Італії чи Аппалачів у Сполучених Штатах.

Характерною ознакою невикористаного дрібного сільського господарства є самозабезпечення. При цьому фермери споживають значну частину виробленого. Хоча деяку частину їхньої продукції можна продати на ринку, їх загальний обсяг виробництва зазвичай не перевищує того, що необхідно для годування сім'ї. У цьому разі низька не тільки продуктивність на душу населення, але й продуктивність одиниці землі. Навіть якщо земля в іншому родюча, її родючість може бути виснажена десятиліттями безперервного вирощування сільськогосподарських культур.

Оскільки економіка продовжує розвиватися, значну частину сільськогосподарської робочої сили необхідно вивільнити від сільського господарства для інших видів діяльності. Звичайно, ця фундаментальна зміна робочої сили стала можливою завдяки величезному збільшенню виробництва на душу населення після модернізації сільського господарства. Збільшення виробництва відбувається за рахунок різних факторів, зокрема, більшого використання добрив і техніки, що підвищує продуктивність окремого працівника.

Модернізація сільського господарства зменшила залежність врожаю від землі та людської праці. Сила тварин і машини замінили силу людини; потім механічна сила замінила силу тварин. Заміна механічної енергії силою тварин зменшує потребу в землі. З модернізацією збільшення використання хімічних добрив також може замінити землю та працю; те саме стосується гербіцидів та пестицидів. Кількість землі та праці, необхідних для виробництва певного обсягу продукції, є меншою, оскільки можна збільшити виробництво на одиницю площі та на годину праці [15].

Незважаючи на загальне підвищення ефективності сільського господарства, зростання врожайності сповільнилося.

Збільшення використання землі, іригації та пестицидів відіграло важливу роль у зростанні сільськогосподарського виробництва під час Зеленої революції. Однак зараз визнано, що сільське господарство часто має негативний вплив на природні ресурси, такі як деградація земель, засолення зрошуваних територій, надмірний видобуток підземних вод, підвищена стійкість до шкідників та ерозія біорізноманіття. Сільське господарство також завдає шкоди навколишньому середовищу через вирубку лісів, викиди парникових газів і забруднення водою нітратами.

Існує також величезна різниця у врожайності сільськогосподарських культур між країнами з високим і низьким рівнем доходу. Врожайність пшениці та рису в країнах з низьким рівнем доходу наразі вдвічі нижча, ніж у країнах з високим рівнем доходу. Врожайності основних сільськогосподарських культур (зернові, коренеплоди, бобові, цукор, олійні культури та овочі) також сильно різняться в різних регіонах.

Практики збереження можуть допомогти підвищити продуктивність сільського господарства.

Ключем до сталого зростання сільського господарства є більш ефективне використання земельних, трудових та інших ресурсів завдяки технологічному прогресу, соціальним інноваціям та новим бізнес-моделям. Для того, щоб сільське господарство та аквакультура відповідали викликам майбутнього, інновації повинні не лише підвищувати ефективність перетворення вхідних ресурсів у вихідну продукцію, а й зберігати дефіцитні природні ресурси та зменшувати відходи. Сільськогосподарські практики, необхідні для збереження та більш ефективного використання природних ресурсів, залежать від місцевих умов та потреб.

В останні роки спостерігається зростаюча тенденція до ґрунтозахисного сільського господарства. Це підхід, спрямований на зменшення деградації ґрунтів шляхом мінімізації механічного обробітку, підтримання органічного покриву на поверхні ґрунту та вирощування різноманітних однорічних і багаторічних видів рослин у поєднанні, послідовності та ротації дерев, чагарників, трав і сільськогосподарських культур. Це сприяє, наприклад, інтеграції бобових і

зернобобових культур у системи землеробства, які допомагають підвищити і підтримувати рівень азоту в ґрунті. У всьому світі ґрунтозахисне сільське господарство практикується на близько 117 мільйонах гектарів, або близько 8 відсотків від загальної площі орних земель у світі. Найбільше воно поширене в Австралії, Канаді та Південній Америці, де ним охоплено понад 50 відсотків оброблюваної площі.

Агроєкологія [16] означає перехід від "готових рішень" до "індивідуалізації".

Відмовляючись від використання хімікатів, фермери досягають вищої якості та кількості завдяки цілісному, інтегрованому підходу, заснованому на управлінні екосистемами. Це досягається шляхом відновлення біологічної складності, включаючи різноманітність рослин, багаторічний трав'яний покрив і наявність дерев. Агроєкологічне сільське господарство підвищує ефективність виробництва продуктів харчування, перериваючи кругообіг азоту, і забезпечує багато додаткових екологічних переваг, таких як зменшення відходів і забруднення на ландшафтному рівні та підвищення економічної ефективності на рівні фермерських господарств.

Кліматично-розумне сільське господарство має на меті стійке підвищення продовольчої безпеки та доходів, підвищення адаптації та стійкості до зміни клімату, одночасно отримуючи потенційні супутні вигоди від пом'якшення наслідків зміни клімату. Воно поєднує ресурсозберігаюче сільське господарство, агроєкологію, агролісомеліорацію та інші інноваційні технології, такі як виведення сортів сільськогосподарських культур, стійкіших до шкідників і хвороб, посухи, перезволоження та засолення ґрунту. Кліматично орієнтоване сільське господарство сприяє сталому тваринництву, яке збалансовує екологічні та виробничі цілі, наприклад, системи змішування сільськогосподарських культур і тваринництва, чергування пасовищних і кормових культур для поліпшення якості ґрунту і зменшення ерозії, а також використання гною для підтримки родючості ґрунту. У кліматично-орієнтованому сільському господарстві системи агролісомеліорації є важливим інструментом для сталого виробництва продуктів

харчування при одночасному захисті екосистем, особливо на маргінальних територіях, схильних до деградації довкілля [17].

1.2 Агропромисловий комплекс в економіці України

Сільське господарство – одна з найважливіших галузей економіки України, яка була прибутковою навіть у кризові періоди. Україна має значний земельний потенціал і дуже високий рівень розвитку сільського господарства. Країна є значним гравцем на світовому аграрному ринку та займає лідируючі позиції у виробництві насіння соняшнику, ріпаку, кукурудзи, ячменю. При цьому ціни на українську сільськогосподарську продукцію демонструють сильну залежність від світових цін на сільськогосподарську продукцію.

Проте агресія Росії проти України впливає на виробничі та експортні можливості України. Крім того, війна призвела до часткового закриття портів і операцій з дроблення олійних культур, що вплинуло на експорт.

Україна послідовно виходить з кризи після нестабільної політичної та економічної ситуації 2014–2015 років. У 2014–2015 роках українські виробники переключилися на європейських партнерів, які замінили російську частку. ВВП демонстрував стабільну тенденцію до зростання до 2022 року. У 2022 році інфляція прискорилося до 30%, а економіка скоротилася майже на 32% через збитки (рис. 1.1), завдані повномасштабною війною Росії проти України. Водночас з наступного року очікується сповільнення інфляції та помірні темпи зростання економіки.

	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Номінальний ВВП, млрд. дол.	176	183	132	91	93	112	130	154	156	200	136
ВВП на душу населення, дол	3 857	4 030	3 014	2 115	2 186	2 640	3 095	3 659	3 726	4 834	3 287
Інфляція, %	(0,20)	0,50	24,90	43,30	12,40	13,70	9,80	4,10	5,00	10,00	26,60

Рисунок 1.1 - Ключові макроекономічні показники

Джерело: розроблено автором за даними [18].

Станом на кінець 2022 року споживчі ціни зросли на 26,6%. Водночас за останні три місяці інфляція в річному вимірі майже не змінилася. Деокупація

територій, збільшення пропозиції продуктів харчування та зниження споживчого попиту через енергетичний терор Росії допомогли стабілізувати інфляційний тиск. Стримували інфляцію також незмінні тарифи на комунальні послуги, фіксований курс гривні та покращення матеріально-технічного забезпечення. Заходи НБУ, зокрема щодо запровадження депозитних інструментів для хеджування валютного ризику, а також обмеження обсягів монетарного фінансування бюджету сприяли певній стабілізації на валютному ринку наприкінці 2022 року (рис. 1.2).

Цьогорічний економічний спад був спричинений зниженням внутрішнього попиту, порушенням логістики та великими втратами робочої сили та виробничого потенціалу, спричинених війною. З іншого боку, відновленню економічної активності сприяє звільнення окупованих територій, швидша адаптація бізнесу до умов війни та запуск «зернового коридору». За оцінками НБУ, наступного року економіка повернеться до зростання завдяки очікуваному зниженню ризиків безпеки та пожвавленню внутрішнього попиту та інвестиційної активності, оскільки ланцюжки виробництва та постачання налагоджені, а фіскальні стимули залишаються значними. Однак наслідки війни стримуватимуть швидке відновлення. Очікується, що у 2023–2024 роках ВВП зростатиме приблизно на 4–5% на рік.



Рисунок 1.2 - Частка агропромислового комплексу в ВВП, %, 2010-2021
Джерело: розроблено автором за даними [19].

Економічне зростання України значною мірою залежить від розвитку аграрного сектору принаймні протягом останніх 15 років. Зрозуміло, що сільське господарство дає значний внесок у ВВП, що робить агропромисловість третім за величиною сектором економіки України. Сприятливі кліматичні умови для великомасштабного сільського господарства, доступ до багатих земельних і водних ресурсів, а також багатий сільськогосподарський ґрунт, також відомий як «чорнозем», один із найродючіших ґрунтів у світі, є чинниками, які забезпечують його подальший розвиток. Поєднання цих факторів разом з історичними аспектами зробили Україну значним гравцем на світовому продовольчому ринку, особливо високоолеїнових культур, ячменю, кукурудзи та пшениці.

До повномасштабного вторгнення Росії в лютому 2022 року на сільськогосподарський сектор України припадало 11 відсотків ВВП країни, майже 20% робочої сили та майже 40 відсотків загального експорту, при цьому Україна була п'ятим найбільшим експортером пшениці у світі, четвертим за величиною експортером кукурудзи та третім за величиною експортером ріпаку.

До початку війни Україна посідала провідне місце в експорті соняшникової олії, соняшникового шроту (1 місце), кукурудзи, ріпаку (3 місце), ячменю, волоського горіха (4 місце), жита, меду, сорго (5 місце) (рис. 1.3).



Рисунок 1.3 - Місце України на світовому ринку їжі в 2021 році

Джерело: розроблено автором за даними [20].

Основні труднощі в аграрному секторі України, пов'язані з повномасштабним вторгненням Росії в лютому 2022 року:

- Блокада морських портів. До вторгнення Росії 95% експорту здійснювалося через морські порти. З лютого 2022 року всі порти (крім портів річки Дунай з низькою пропускнуою здатністю) були заблоковані. Частково проблема була

вирішена після підписання зернової угоди та відкриття 3 портів: Одеса, Чорноморськ, Південний.

- Обмеження наземних шляхів сполучення, що зумовлено низькою пропускнуою здатністю залізничних станцій на кордоні з країнами Європи, різною шириною колії в Україні та ЄС, ускладненим доступом до європейського ринку в частині імпорту та сертифікації зерна.

- Логістичні витрати. Вартість логістики в гривневому еквіваленті зросла в 2-3 рази через підвищення тарифів на залізниці, збільшення кредитного плеча доставки та інфляції. Витрати на логістику становили 2/3 цін на врожай у березні-липні, що робило його виробництво для українських аграріїв менш рентабельним.

- Проблеми з відшкодуванням ПДВ: на початку повномасштабного вторгнення системне відшкодування було зупинено на чотири місяці. З червня відновлено повернення ПДВ, але із затримками.

- Нестача коштів оборотного капіталу через зростання собівартості продукції, інфляцію, здорожчання банківського кредитування через підвищення облікової ставки НБУ, невідшкодування ПДВ та фіксованого курсу валют тощо.

- У прифронтовій зоні або поблизу неї: захоплення підприємств, розкрадання та вивезення врожаю, техніки, обладнання; неможливість забезпечити звичайну посівну через мінування поля та близькість до зони бойових дій; пошкодження та руйнування сільськогосподарської інфраструктури - складів, зерносховищ, а також сільськогосподарської техніки та обладнання (рис. 1.4).

Загалом у 5 областях (Харківській, Луганській, Донецькій, Запорізькій, Херсонській), які найбільше постраждали від війни, вироблено 32% пшениці, 29% соняшнику, 26% ячменю, 25% овочів, 23% меду, 35% овець і кіз, 20% свиней, 16% яєць. Також, на 27% порівняно з 2021 роком скоротилася площа ріллі.

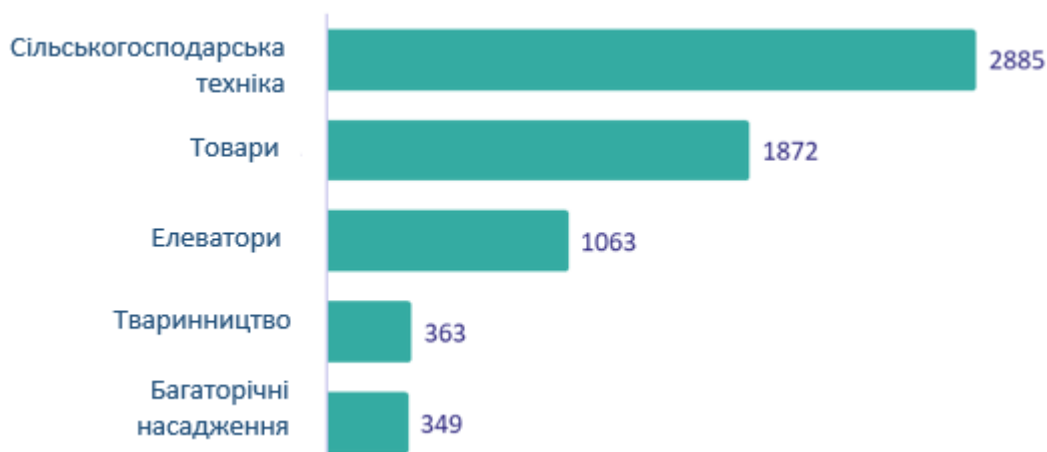


Рисунок 1.4 - Кількість втрат по категоріях, млн. дол.

Джерело: розроблено автором за даними [20].

До закінчення війни важко оцінити розмір збитків і втрат для аграрного сектора України. Згідно з дослідженням Київської школи економіки, сума збитків агросектору становить 6,6 млрд. дол. або 5% від загальної оцінки втрат інфраструктури.

1.3 Структура землі та основні сільськогосподарські товари України

У 2021 році розпочато поступове впровадження земельної реформи. Однак структура землі все ще сильно роздроблена, і підприємствам часто доводиться підтримувати численні договори оренди з окремими власниками. Роздробленість землі також сприяє зниженню ефективності сільськогосподарського виробництва. Найбільша частка ринку землі знаходиться в обробітку малих і середніх компаній (земельні ділянки 1000 – 5000 га). Дрібні фермери часто мають обмежений доступ до фінансових ресурсів і сучасних технологій, що робить обробіток площі менше 500 га майже контрпродуктивним, тоді як великі гравці мають кращий доступ до ресурсів і новітньої техніки. Таким чином, великі фірми можуть досягти вищої продуктивності землі, якості продукції та ефективності операцій.

Україна має значний земельний потенціал і дуже високий рівень розвитку сільського господарства. Сільськогосподарські угіддя становлять 68% загальної

території і 28% світових запасів «чорнозему» розташовано на території України (рис. 1.5 та 1.6). Перевага країни щодо ресурсів перетворюється на її позицію на світовому ринку.

З 2002 року в Україні діє мораторій, метою якого є захист власників землі. На прийняття закону про ринок землі пішло багато років. Проте його впровадження відбуватиметься поступово. З 1 липня 2021 року купувати землі сільськогосподарського призначення можуть лише фізичні особи — громадяни України. Вони можуть купувати не більше 100 га на одну людину. З 1 січня 2024 року купувати землю зможуть юридичні особи, зареєстровані в Україні, засновником яких є громадянин України. При цьому площа землі, що купується, не повинна перевищувати 10 тис. га.

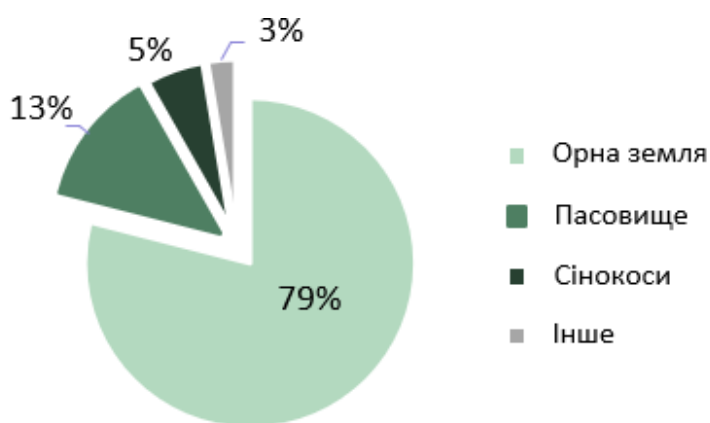


Рисунок 1.5 - Структура землі в агропромисловому комплексі
Джерело: розроблено автором за даними [21].

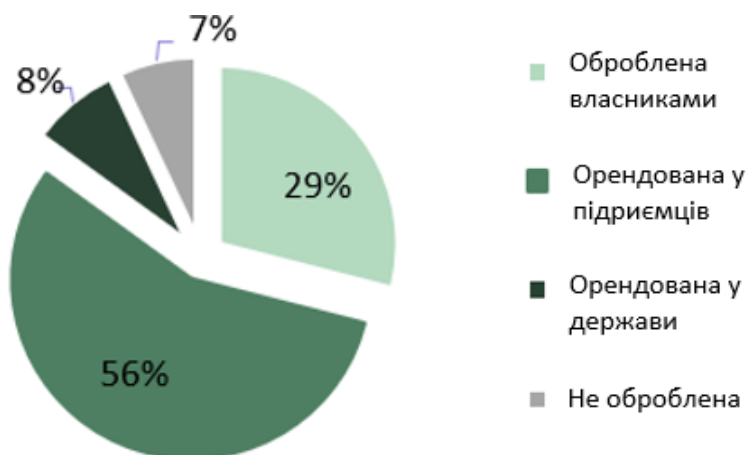


Рисунок 1.6 - Структура використання землі в агропромисловому комплексі
Джерело: розроблено автором за даними [21].

Ринок землі відкрився, але купівля-продаж землі відбувається досить повільно через непідготовлену інфраструктуру (складність юридичного оформлення угод, незаповненість землі в електронному реєстрі) та відсутність охочих продати землю (рис. 1.7).

Потенційні наслідки відкриття ринку землі:

- Приплив внутрішніх і зовнішніх інвестицій (земля стає активом, що викликає довіру інвесторів);
- Розширення переліку банківських послуг (фінансування купівлі земель сільськогосподарського призначення, причому земля може бути заставою кредитів для поточної діяльності підприємства);
- Розширення матеріально-технічного забезпечення господарств;
- Розвиток переробної промисловості;
- Розвиток одноосібних господарств, фермерських господарств та виробників багаторічних насаджень;
- Зменшення неофіційного обробітку землі.

Через повномасштабне вторгнення Російської Федерації в Україну уряд обмежив доступ до Реєстру речових прав на нерухоме майно та Державного зеленого кадастру, що зупинило ринок нерухомості та землі з 24 лютого 2022 року. У травні 2022 року функціонування державних реєстрів було відновлено. Як наслідок, протягом 24 лютого 2022 року – 24 січня 2023 року загальна кількість угод купівлі-продажу склала 44 307 одиниць, загальною площею 82 253,78 га (в середньому 1,8 га).

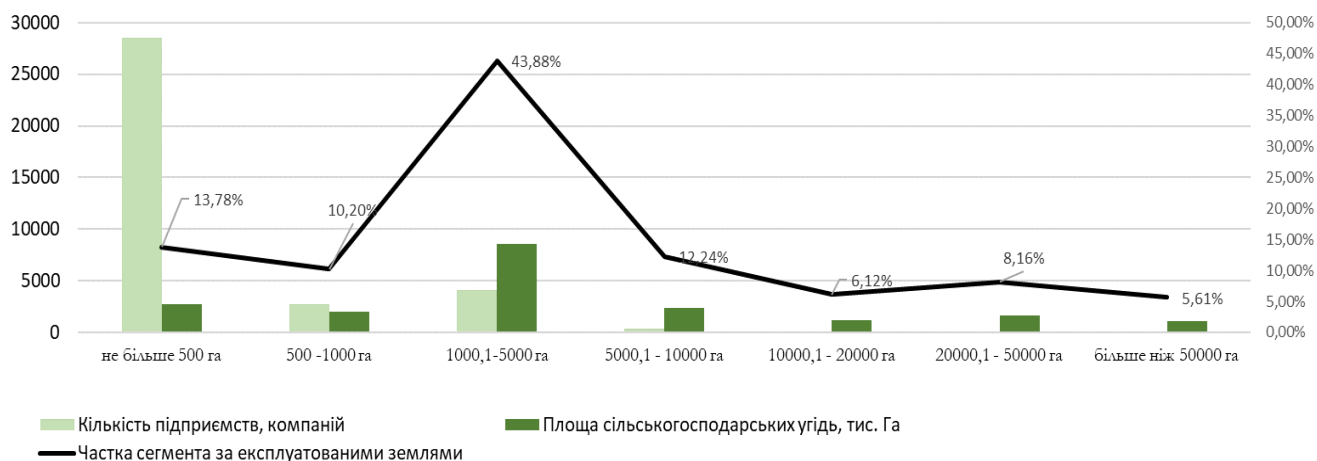


Рисунок 1.7 - Розподіл сільськогосподарських підприємств за масштабами землекористування, 2020

Джерело: розроблено автором за даними [21].

Найбільшу частку землі займають компанії, які обробляють 1000-5000 га (43,88%). Великі фермерські господарства з площею обробітку понад 2000 га мають більше ресурсів і кращий доступ до кредитів чи лізингу для придбання нової імпоротної енергоємної техніки, яка гарантуватиме високу якість обробітку ґрунту, вищу продуктивність землі і, як наслідок, мінімізує можливі ризики для їхнього бізнесу і зробить його більш конкурентоспроможним. Отже, великі фірми можуть досягти вищої продуктивності землі, якості продукції та ефективності операцій (рис. 1.8, 1.9, 1.10).

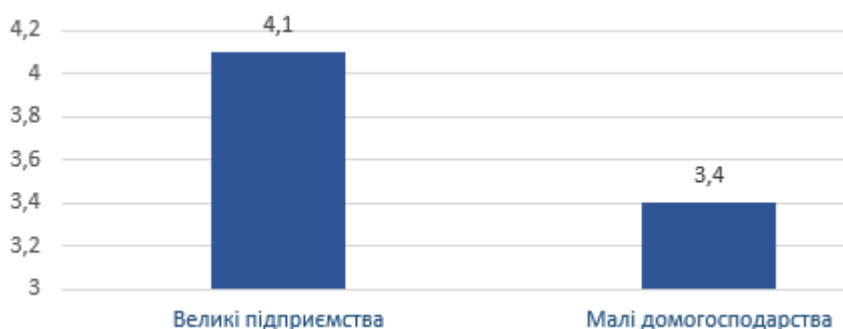


Рисунок 1.8 - Середня врожайність пшениці (2016-2020 рр.), тонн/га

Джерело: розроблено автором за даними [21].

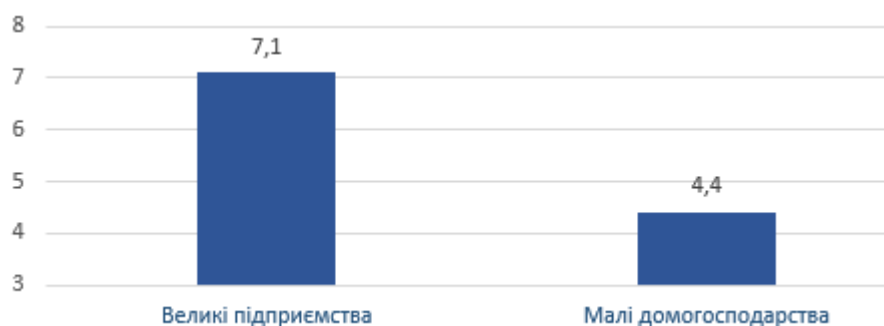


Рисунок 1.9 - Середня врожайність кукурудзи на зерно (2016-2020 рр.), тонн/га

Джерело: розроблено автором за даними [21].



Рисунок 1.10 - Середня врожайність соняшнику (2016-2020 рр.), тонн/га

Джерело: розроблено автором за даними [21].

Зернові та технічні культури становлять найбільшу частку сільськогосподарського виробництва. Продукти переробки включають виробництво олії, цукру та борошна. Протягом останніх 10 років посівні площі під найбільш популярними культурами (пшениця, соняшник, кукурудза, ячмінь, ріпак, соя) збільшувалися (рис. 1.11). Урожайність основних сільськогосподарських культур протягом останніх 10 років стабільно зростає. За останні 10 років урожайність основних сільськогосподарських культур зростала в середньому на 8% щорічно. Основні причини: вдосконалення технологій, вдосконалення технічного парку, тенденція до інтенсифікації виробництва. Країна має значні переробні потужності, а обсяги експорту демонструють тенденцію до зростання. Оскільки споживання скорочується, а експорт зростає, можна припустити, що ціни на сільськогосподарську продукцію значною мірою залежать від світових цін. Проте врожайність агроциклу 2021/2022 нижча в середньому на 15-30% залежно від культури та регіону. Основні причини: погані погодні умови порівняно з 2020/2021;

зменшення посівних площ через війну (деякі поля були заміновані або пошкоджені); зменшення внесення добрив за рахунок збільшення їх вартості в 2-3 рази; відсутність елеваторів, через що деякі фермери залишали кукурудзу до зими в полі.

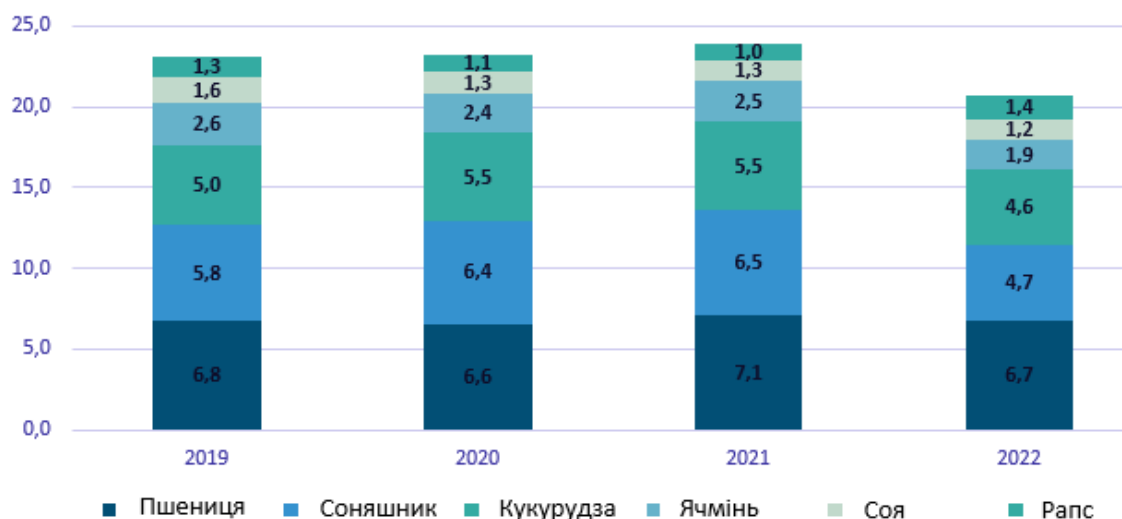


Рисунок 1.11 - Посівна площа під головні культури в Україні, млн. га
Джерело: розроблено автором за даними [21].

ТОП-5 культур за площами зайняли 75% загальних посівних площ в Україні у 2021 році та створили основний дохід сільгоспвиробників України. У 2021 році це були соняшник (23,9%), пшениця (23,2%), кукурудза (18,5%), соя (5,1%), картопля (4,5%).

81% усіх посівних площ у 2022 році зайнято цими культурами. Тому ідентифікація ризиків, пов'язаних із цими культурами, охоплює більшість ризиків, пов'язаних із вирощуванням усіх культур на підприємствах.

У нинішньому сезоні спостерігається значне зменшення порівняно з попереднім роком площ під соняшником, ячменем та кукурудзою – -28%, -24% та -16% відповідно (рис. 1.12). Основною причиною скорочення посівних площ соняшнику та ячменю є регіони зайнятості, які спеціалізувалися на вирощуванні соняшнику та ячменю (Херсонська та Запорізька області). Основними причинами зменшення площ під кукурудзою є високий обсяг урожаю культури та високі

витрати добрив, в умовах дорогої логістики та високої ціни на добрива це робить культуру не такою бажаною, як у попередні сезони.



Рисунок 1.12 - Динаміка посівних площ під найпопулярніші культури в українських агропромислових компаніях

Джерело: розроблено автором за даними [21].

Урожайність основних сільськогосподарських культур протягом останніх років стабільно зростає. Загальна тенденція – підвищення врожайності, але цей показник сильно залежить від погодних умов. Вплив цих умов можна спостерігати у випадку з кукурудзою: посушливе літо 2017 року дало врожайність 5,4 т/га, тоді як достатня кількість вологи у 2018 році забезпечила врожайність 7,4 т/га. 2020 рік характеризувався несприятливими погодними умовами на півдні України. Проте 2021 рік назвали роком з одними з найсприятливіших погодних умов за останні 10 років (рис. 1.13).



Рисунок 1.13 - Врожайність найпопулярніших культур в агропромисловому комплексі України

Джерело: розроблено автором за даними [21].

Проте врожайність агроциклу 2021/2022 нижча в середньому на 15-30% залежно від культури та регіону.

Зменшення врожайності порівняно з чотирма попередніми сезонами на більшій частині України в сезоні 2021/2022 має три основні причини:

- 1) більш сухі умови (детальніша інформація наведена нижче);
- 2) порушення виробничого циклу, пов'язане з військовими діями;
- 3) скорочення фінансування сільськогосподарського циклу через ризики повномасштабної війни.

Розраховано показник, який показує загальне зниження врожайності серед виробників у кожному регіоні, враховуючи структуру посівів у кожному регіоні порівняно з чотирма попередніми сезонами. За показником всі регіони виділили власну зону ризику: низький ($> -5\%$), середній ($-5\% - -10\%$) і високий ($< -10\%$). Найбільший недобір урожаю спостерігався у Вінницькій, Закарпатській, Чернігівській та Житомирській областях.

Крім того, минулий сезон характеризується низькою часткою збирання кукурудзи (на поточний момент 10.11.2022) було зібрано лише 39% від загальної посівної площі. У попередні роки за цей період українські виробники збирали 70-80% посівних площ. Основною причиною низького рівня збирання кукурудзи українськими виробниками стала висока ціна на сушіння зерна. Вони чекали, коли кукурудза втратить вологу природним шляхом на полях, а потім її збирали.

Порівняно з Європейським Союзом є ще багато можливостей для покращення з точки зору врожайності. Проте Україна до війни демонструвала постійне зростання та зуміла отримати вищі результати протягом останніх 10 років. Найвища врожайність за останні 10 років припала на 2021 рік.

Іншим показником, який може вказувати на регіони, де врожайність сільськогосподарських культур може бути низькою, є «Відсоток вологості ґрунту» — доступна для рослини вода поділена на загальну водоутримуючу здатність ґрунтового профілю. Це корисно для визначення того, чи достатньо води в профілі ґрунту для розвитку культур.

Площа сухих зон за роки незалежності збільшилася на 7%, а розміри надмірно зволжених зон зменшилися на 10%. В даний час близько 60% орних земель в Україні відчувають потребу в більшій кількості води. Деякі підприємства використовують зрошувальні системи, щоб зменшити потенційні втрати врожаю через посуху. Отже, впровадження таких систем може зменшити ризик.

В Україні гостро стоїть проблема зміни клімату та незадовільного стану зрошувальної інфраструктури. Наприклад, у 1991 році зрошувалося 2,29 млн. га сільськогосподарських угідь, а зараз лише близько 0,55 млн. га (загальна посівна площа — близько 28 млн. га). Проте останнім часом цьому питанню приділяється все більше уваги, на ринку розширюється пропозиція сучасних технологій та обладнання для інфраструктури зрошення, з'являються програми державної підтримки.

Найбільше зрошувальних систем у регіонах з більшою ймовірністю посухи: Одеській, Херсонській, Запорізькій. Отже, середня врожайність із зрошуваних земель вища.

Згідно з інформацією Всесвітньої метеорологічної організації, 2021 рік був більш вологим порівняно з двома попередніми сезонами. Погодні умови протягом останніх трьох сезонів, дуже коливалися, були моменти, коли для вирощування сільськогосподарських культур наставали ідеальні часи. Але також, можна прослідкувати сезон, коли відчувалася максимальна посуха, що свідчить про несприятливі умови для посіву, вирощування чи збирання урожаю.

За даними Міністерства аграрної політики та продовольства України щодо врожайності, є кілька регіонів, де врожайність у поточному сезоні значно знизилася порівняно з чотирма попередніми роками (2018-2021). Серед них: Київ, Харків, Вінниця, Одеса, Миколаїв, Запоріжжя, Дніпропетровськ, Донецьк. Найбільше падіння пшениці та ячменю -29% та -28% відповідно у Київській області.

Оцінка показала, що погодні ризики залишаються основною загрозою для сільськогосподарських клієнтів. Ризик коливання цін і ризик подальшої роботи морських портів як основного джерела експорту врожаю виявилися найбільш суттєвими макроекономічними ризиками. Залежність від постачальників, логістичні маршрути, складські потужності є іншими критичними факторами, які враховуються під час оцінки ризиків клієнта.

Україна є значним гравцем на світовому аграрному ринку. Країна має значні переробні потужності, а обсяги експорту демонструють тенденцію до зростання. Але через вторгнення Росії до України ці можливості значно скоротилися і тільки завдяки угоді про «Зерновий коридор» [22], Україна продовжила експортувати власну продукцію. Оскільки споживання скорочується, а експорт зростає, можна припустити, що ціни на сільськогосподарську продукцію значною мірою залежать від світових цін. Аналіз експорту також свідчить про значну залежність від коливань світового попиту на сільськогосподарську продукцію. Експортні квоти ЄС не становлять значного ризику для українських виробників, оскільки поточні обсяги експорту перевищують квоти в 5-6 разів. Покупці сільськогосподарської продукції дуже диверсифіковані; тому немає ризику втрати частки експорту через торгові бар'єри чи квоти.

РОЗДІЛ 2

ПОБУДОВА ТА АНАЛІЗ ЕКОНОМЕТРИЧНИХ МОДЕЛЕЙ

2.1 Побудова та перевірка значущості економетричної моделі

Економетричні моделі будуються з економічних даних за допомогою методів статистичного висновку. Ці моделі зазвичай базуються на економічних теоріях, які припускають оптимізацію поведінки економічних агентів [23].

Економетрична модель – це функціональна модель. Вона кількісно описує зв'язок між вхідними показниками (X) та вихідними показниками (Y) економічної системи. Загалом економетричну модель можна записати так:

$$Y = f(X, e), \quad (2.1)$$

де X – вхідні економічні показники;

e – випадкова, або стохастична, складова.

Змінна X часто є детермінованою. Враховуючи, що адитивний член e є випадковою величиною, а залежна змінна Y залежить від нього, отже, вона також є стохастичною. Звідси висновок: економетричні моделі є стохастичними.

Існує кілька особливостей у побудові та дослідженні економетричних моделей. Ці особливості пов'язані з тим, що економетричні моделі є стохастичними [24].

Математична форма парної лінійної регресії:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad (2.2)$$

де β_0 і β_1 константи (параметри моделі), які необхідно визначити;

ε_i – похибка моделі.

З кількома предикторами рівняння регресії має вигляд:

$$\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_{1i} + \hat{b}_2 X_{2i} + \dots + \hat{b}_p X_{pi} + e, \quad (2.3)$$

Оцінки параметрів множинної регресії обчислюють за методом найменших квадратів (МНК, least squares), розрахункова формула якого має вигляд:

$$\beta = (X^T X)^{-1} X^T Y, \quad (2.4)$$

де β — вектор оцінок параметрів моделі;

X — матриця предикторів з одиничним першим стовпчиком;

X^T — транспонована матриця предикторів;

Y — матриця незалежної змінної.

Похибка моделі визначається як різниця між спостереженим (Y_i) та модельним значенням результативного показника ($\hat{Y}_i$)

$$e_i = Y_i - \hat{Y}_i \quad (2.5)$$

Для визначення параметрів моделі використовують метод найменших квадратів (МНК, least squares).

Метод найменших квадратів — це важливий статистичний метод, який використовується для обчислення оцінок параметрів регресії [25].

Даний метод визначає такі значення параметрів моделі, за яких сума квадратів відхилень модельних значень показника від його реальних значень наближатиметься до мінімуму.

$$RRS = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - \hat{b}_0 - \hat{b}_1 X_{i1})^2 \quad (2.6)$$

Щоб коефіцієнти, обчислені за методом найменших квадратів, були якісними, необхідно, щоб виконувалися такі умови:

1. Нормальність — значення залежної змінної нормально розподілені при фіксованих значеннях незалежних змінних;
2. Незалежність — значення залежної змінної мають бути незалежні одна від одної;
3. Лінійність — залежна змінна лінійно пов'язана з незалежними;
4. Гомоскедастичність — дисперсія залежної змінної постійна при різних значеннях незалежних змінних.

Якщо ці вимоги не будуть дотримані, то результати на перевірку значущості та довірчі інтервали можуть показати неточність моделі.

Можна вважати, що одним з дуже важливих показників для визначення ефективності моделі є середньоквадратична похибка цієї моделі (RMSE, root mean squared error) [26], яка саме і оцінює загальну точність моделі:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \quad (2.7)$$

Подібною до середньоквадратичної похибки є стандартна похибка залишків (RSE, residual standard error). На відміну від RMSE у RSE вже використовується показник p – кількість предикторів у моделі.

$$RSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n - p - 1}} \quad (2.8)$$

Одним з найбільш інформативних показників при оцінюванні параметрів моделі є коефіцієнт детермінації – він вимірює пояснювальну в моделі частку варіації в даних [272]. Якщо сказати іншими словами, коефіцієнт детермінації показує, яка частина змінювання результативного показника відображається даною регресійною моделлю:

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = 1 - \frac{\text{Var}(\hat{Y})}{\text{Var}(Y)} = \frac{\text{Var}(\hat{Y})}{\text{Var}(Y)} \quad (2.9)$$

Скоригований коефіцієнт детермінації – аналогічний до звичайного коефіцієнту детермінації, але з врахуванням штрафу на кількість предикторів у моделі.

$$R^2_{adj} = 1 - (1 - R^2) \frac{(n - z)}{(n - k)} \leq R^2 \quad (2.10)$$

Для вимірювання щільності кореляційного зв'язку між показниками застосовують безрозмірну величину – коефіцієнт кореляції, який для парної моделі обчислюється за формулою [28]:

$$r = \frac{\sum (x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum (x_t - \bar{x})^2} \sqrt{\sum (y_t - \bar{y})^2}}, \quad (2.11)$$

де кожен вираз в дужках визначає відхилення спостережених значень показника від його середнього значення.

У разі множинної регресії коефіцієнт кореляції обчислюється як корінь з коефіцієнта детермінації.

Перевірка побудованої регресійної моделі виконується з використанням відомих статистичних критеріїв. Загальна адекватність моделі перевіряється за критерієм Фішера, а значущість кореляційного зв'язку та значущість окремих параметрів моделі – за критерієм Стюдента.

Щоб перевірити на значущість рівняння регресії, використовують F-критерій (інша назва критерій Фішера чи F-статистика [29]):

$$F = \frac{R^2}{1-R^2} \times \frac{n-m-1}{m} \quad (2.12)$$

Отже, F-тестом (чи критерієм Фішера, чи F-критерієм) – називають будь-який статистичний критерій, тестова статистика якого при виконанні нульової гіпотези має розподіл Фішера (F-розподіл).

Статистика тесту у певний спосіб зводиться до відношення дисперсій вибірки (сума квадратів, поділена на «ступені свободи»). Щоб статистика мала розподіл Фішера, чисельник і знаменник мають бути незалежними випадковими величинами, а відповідні суми квадратів мати розподіл χ^2 . Це вимагає, щоб дані були нормально розподіленими. Крім того, припускається, що випадкові величини (залишки моделі) мають однакову дисперсію.

Властивостями F-розподілу є:

- Крива F-розподілу позитивно нахилена вправо з діапазоном від 0 до ∞ ;
- Значення F завжди додатне або дорівнює нулю. Немає негативних значень;
- Форма розподілу залежить від ступенів свободи чисельника ϑ_1 і знаменника ϑ_2 ;
- Ступінь асиметрії зменшується зі збільшенням ступенів свободи чисельника та знаменника;

- Крива F-розподілу ніколи не може бути симетричною; при збільшенні ступенів свободи вона буде більш схожою на асиметричну.

Значення цього показника порівнюють з табличним значенням розподілу Фішера, встановлюючи рівень значущості ($\alpha = 0,05$) і ступені вільності ($m, n-m-1$).

Після розрахунку статистики необхідно знайти табличне значення критерію. Якщо $F_{\text{розрах.}} > F_{\text{таб.}}$, то коефіцієнт детермінації і модель загалом статистично значущі, в іншому разі - незначущі.

t-критерій [30] – визначає статистичну значущість оцінок параметрів моделі (він перевіряє гіпотезу чи рівні параметри моделі «нулю»).

$$t_b = \frac{\hat{b}}{SE(\hat{b})}, \quad (2.13)$$

де у чисельнику дробу – оцінка параметра, а в знаменнику – його стандартна похибка.

Критичні значення F-критерію та t-критерію, визначаються за спеціальними таблицями, або ж програмно за значеннями p-value, для якого критичним прийнято вважати значення, що дорівнює 0.05.

Отже, відповідно, що чим більшим буде значення обох критеріїв і нижчим p-value, то більш значущим буде той чи інший результат.

2.2 Діагностика лінійної регресії та відбір предикторів

Важливе місце в аналізі лінійної регресії посідають викиди. Викиди – це значення, що виникають хаотично і не прогнозовано, а для моделі вони мають дуже великий за абсолютним значенням залишок.

В більшості випадків побудова відповідних графіків допоможе виявити такі викиди. У середовищі R Studio побудова графіків здійснюється за допомогою спеціальних функцій, що представлені на (рис. 2.1):

```
m <- lm(y ~ x1 + x2)
par(mfrow = (c(2, 2)))
plot(m)
```

Рисунок 2.1 – Побудова діаграми розсіювання R Studio

Джерело: розроблено автором

Функція `outlierTest()` допомагає визначити значення з найбільшим відхиленням по моделі.

Функція `influence.measures()` розраховує одразу цілий ряд статистичних даних:

- DFBETAS (standardized difference of the beta);
- DFFITS (difference in fits);
- Коефіцієнт варіації;
- Відстань Кука;
- Значення матриці проєкції на простір регресорів (Projection matrix).

Функція `gvlma:gvlma()` виконує загальну перевірку вимог, що мають виконуватися у лінійних моделях, з окремим оцінюванням асиметрії, ексцеса та гомоскедастичності.

Досить часто зустрічаються випадки, коли результативна змінна залежить не від всіх предикторів. Задачу виявлення впливових предикторів називають відбором інформативних змінних.

В ідеалі потрібно перебрати велику кількість моделей, кожна з яких включає свій набір предикторів.

Наприклад, при двох незалежних змінних можна отримати щонайменше 4 моделі:

- Модель без змінних;
- Модель тільки з X_1 ;
- Модель тільки з X_2 ;
- Модель з X_1 та X_2 .

Можна використовувати такі критерії для вибору моделі:

- Ср Меллоу;
- Інформаційний критерій Акаїке (AIC);

- Баєсовий інформаційний критерій (BIC);
- Скоригований коефіцієнт детермінації (R^2);
- та інші їх модифікації.

Часто на практиці використовується принцип леза Оккама [26]. Його суть полягає в тому, що якщо всі моделі є рівними, то обирається модель, яка є простішою у використанні.

Включення в модель нових предикторів завжди зменшує RMSE та збільшує R^2 . Відповідно, вони не завжди підходять у якості орієнтирів з вибору кращої моделі.

Інформаційний критерій Акаїке (AIC) штрафує модель за додаткові змінні:

$$AIC = 2P + n \log(RSS/n), \quad (2.14)$$

де P – кількість змінних;

n – кількість спостережень.

Серед альтернатив обирається модель з меншим AIC.

Існує декілька підходів до відбору предикторів. Базовий з них – покрокова регресія. Розрізняють три методи її побудови:

- Пряма покрокова регресія (forward stepwise) – передбачає початок з моделі без предикторів з подальшим включенням змінної, що призводить до зменшення RSS. Після цього додається нова змінна до попередньої моделі, що знову призводить до зменшення RSS. Цей процес продовжується до виконання певного правила зупинки;
- Зворотна покрокова регресія (backward stepwise) – передбачає початок з моделі з усіма змінними (повна модель), а потім прибирає змінну/змінні з найбільшим p-value, після чого будується нова модель і процес знову повторюється до виконання певного правила;
- Комбінована покрокова регресія (stepwise stepwise) [31].

Якісна перевірка моделі передбачає розбиття вибірки на тестову та навчальну. На навчальній вибірці будується модель, на тестовій перевіряється її якість. За таким підходом якість моделі буде перевірено значно краще.

Базовий тест ANOVA розраховує, чи мають дані нормальний розподіл, чи хоча б наближаються до нього, близькі до дзвоноподібного. Якщо виявляється, що це не так, то слід використовувати критерій Краскела-Уолліса [32].

Критерій Тьюкі використовують, щоб перевірити нульові гіпотези, які свідчать про однаковість середніх для двох груп даних. Цей критерій можна застосувати для перевірки однорідності вибірових даних, адже змінювання дисперсії спостережених даних може призвести до гетероскедастичності залишків.

Гетероскедастичність залишків – змінювання дисперсії залишків для окремих спостережень чи груп спостережень – означає неефективність оцінок параметрів регресійної моделі, обчислених за методом найменших квадратів.

В разі гетероскедастичності параметри моделі мають великі стандартні похибки, а це в свою чергу може призвести до неправильних висновків при перевірці гіпотез і неточних прогнозів.

За алгоритм розрахунку критерія Тьюкі початкова вибірка впорядковується за зростанням і розбивається на m груп, в кожній з яких обчислюється середнє значення. При цьому найбільші середні порівнюються з найменшими. Порівняння проводиться до того моменту, коли всі пари не будуть перебрані.

За результатами порівняння обчислюється модифікована статистика критерія Стюдента:

$$q = \frac{\bar{x}_B - \bar{x}_A}{SE} \quad (2.15)$$

$$SE = \sqrt{\frac{MS_w}{n}}, \quad (2.16)$$

де MS_w – це розрахункове значення, яке було знайдене в процесі дисперсійного аналізу внутрішньогрупової дисперсії.

Якщо групи складаються з різної кількості спостережень, то використовується формула:

$$SE = \sqrt{\frac{MS_w}{2} \left(\frac{1}{n_A} + \frac{1}{n_B} \right)} \quad (2.17)$$

Отриманий результат порівнюється з критичним значенням розподілу Стьюдента чи розрахунковим значенням p-value.

Критерій Краскела-Уолліса – це непараметричний тест, який передбачає перевірку однорідності вибірки навіть без вимоги нормального розподілу даних. В ньому замість середніх значень розглядаються медіани. Статистика критерію в цьому разі має розподіл Хі-квадрат і обчислюється за формулою

$$H = \frac{12}{N(N+1)} \sum_{i=1}^m \frac{R_i^2}{n_i} - 3(N+1) \quad (2.18)$$

де n_i – кількість спостережень у групі.

$N = \sum n_i$ – загальна кількість спостережень у всіх групах m .

R_i = сума рангів спостережень у групі i . Іншими словами – це порядковий номер конкретного спостереження в рядку впорядкованих спостережень.

Чим більше значення Н-критерія, тим більша ймовірність відхилення нульової гіпотези про відсутність різниці між групами, що порівнюються.

Критичне значення при $m > 5$ даний критерій порівнюється з табличним значенням $\chi^2_{(m-1)}$. При меншій кількості груп, необхідно вносити певні поправки.

2.3 Дослідження прогностичних та аналітичних якостей моделей

Модель з багатьма змінними часто призводить до проблеми мультиколінеарності, яка означає існування тісної лінійної залежності або кореляції між незалежними змінними множинної регресії [33].

Існує два види мультиколінеарності:

- повна чи структурна має місце, коли існує функціональний зв'язок між факторами;
- часткова існує у разі кореляційного зв'язку між факторами.

Це може призвести до:

- зміщених оцінок параметрів, тобто таких, що неправильно відображають вплив кожної факторної змінної на результативний показник;
- великих довірчих інтервалів параметрів моделі (коефіцієнти чутливі до змін);
- низька точність оцінок параметрів моделі, що знижує статистичну потужність моделі.

Можливі індикатори наявності мультиколінеарності:

- високий R^2 при низькому t-тесті;
- нелогічні знаки при коефіцієнтах регресії;
- невеликі зміни в даних суттєво змінюють модель.

Фактор інфляції дисперсії (VIF) – чим він вищий для j -го предиктора, тим сильніший лінійний зв'язок між ним та іншими предикторами [34].

$$VIF = \frac{1}{1 - R_j^2}, \quad (2.19)$$

де R_j^2 - коефіцієнт детермінації j -го предиктора відносно інших.

Порогове значення відрізняється у різних джерелах: $VIF = 2$; або 5; або 10.

Основні кроки для розрахунку і дослідження фактору інфляції дисперсії:

Для кожного фактора X_j будується регресійна модель його залежності від інших предикторів (2.20) та обчислюється значення VIF за (2.19)

$$X_j = a_2 X_2 + a_3 X_3 + \dots + a_k X_k + c_0 + \varepsilon \quad (2.20)$$

Отримане значення порівнюється з пороговим. Предиктор з максимальним VIF, що перевищує порогове значення, забирається з моделі. Попередні кроки повторюються, поки в моделі не залишаться предиктори з низьким VIF.

Алгоритм Фаррара-Глобера включає три види статистичних критеріїв для визначення мультиколінеарності:

- усього масиву незалежних змінних (критерій Пірсона);
- кожної незалежної змінної з усіма іншими (F – критерій (критерій Фішера));
- кожної пари незалежних змінних (t – критерій (критерій Стюдента)).

Алгоритм Фаррара-Глобера складається з таких кроків:

1. Стандартизація незалежних змінних:

Матриця змінних факторів X замінюється стандартизованою (нормалізованою) матрицею X^* , елементи якої обчислюють за формулою:

$$x_{ik}^* = \frac{x_{ik} - \bar{x}_k}{\sqrt{n\sigma_{x_k}^2}}, \quad (2.21)$$

де n – кількість спостережень;

$\bar{x}_k$ – середнє арифметичне k -ї пояснювальної змінної;

$\sigma_{x_k}^2$ – дисперсія k -ї пояснювальної змінної.

2. Визначення кореляційної матриці:

$$r_{xx} = \begin{pmatrix} 1 & r_{12} & r_{13} & \dots & r_{1m} \\ r_{21} & 1 & r_{23} & \dots & r_{2m} \\ r_{31} & r_{32} & 1 & \dots & r_{3m} \\ \dots & \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & r_{m3} & \dots & 1 \end{pmatrix} \quad (2.22)$$

Обчислення кореляційної матриці згідно з двома методами нормалізації змінних:

$$r_{xx} = \frac{1}{n-1} (X^{*'} X^*), \quad (2.23)$$

де X^* – матриця нормалізованих незалежних (пояснювальних) змінних;

$X^{*'}$ – матриця, транспонована до матриці X^* .

3. Визначення критерію χ^2 :

$$\chi^2 = - \left[n - 1 - \frac{1}{6} (2m + 5) \right] \ln |r_{xx}|, \quad (2.24)$$

де $|r_{xx}|$ – визначник кореляційної матриці r_{xx} .

Розрахункове значення порівнюється з критичним (табличним). Якщо $\chi^2 > \chi_{кр}^2 = (\alpha, k)$, це інформує нас про наявність ознаки мультиколінеарності в моделі. У протилежному разі – ця ознака відсутня.

4. Визначення оберненої матриці:

$$C = r_{xx}^{-1} = \frac{1}{n-1} (X^{*'} X^*)^{-1} \quad (2.25)$$

5. Обчислення F-критеріїв:

$$F_j^* = (c_{jj} - 1) \cdot \frac{n - m}{m - 1}, \quad j = \overline{1, m}, \quad (2.26)$$

де c_{jj} – діагональні елементи оберненої матриці C .

За заданим рівнем значущості α і числом ступенів свободи $k_1 = n - m$, $k_2 = m - 1$ за таблицею визначається $F_{кр} = (\alpha, k_1, k_2)$.

Якщо $F_j^* > F_{кр} = (\alpha, k_1, k_2)$, то j -й регресор буде мати суттєвий зв'язок з іншими регресорами, тобто j -та пояснювальна змінна мультиколінеарна з ними. У протилежному разі – зв'язок не суттєвий.

Коефіцієнт детермінації для кожної змінної:

$$R_k^2 = 1 - \frac{1}{c_{kk}}. \quad (2.27)$$

Якщо коефіцієнт детермінації наближається до одиниці, то пояснювальна змінна мультиколінеарна з іншими.

6. Обчислення частинних коефіцієнтів кореляції:

$$r_{kj} = \frac{-c_{kj}}{\sqrt{c_{kk} \cdot c_{jj}}}, \quad (2.28)$$

де c_{kj} – елемент матриці C , що міститься в k -му рядку і j -му стовпці; c_{kk} і c_{jj} – діагональні елементи матриці C .

Частинні коефіцієнти кореляції показують тісноту зв'язку між двома пояснюючими змінними за умови, що інші змінні не впливають на цей зв'язок.

7. Обчислення t -критеріїв:

$$t_{kj} = \frac{r_{kj} \sqrt{n - m}}{\sqrt{1 - r_{kj}^2}}. \quad (2.29)$$

Фактичні значення критеріїв t_{kj} порівнюються з табличними при ступенях свободи $n - m$ і рівні значущості α . Якщо $t_{kj} > t_{кр}$, то між пояснювальними змінними X_k і X_j існує мультиколінеарність [35].

РОЗДІЛ 3

МОДЕЛЮВАННЯ ТА ОЦІНЮВАННЯ ПРИБУТКОВОСТІ КУЛЬТУР В АГРОПРОМИСЛОВОМУ КОМПЛЕКСІ У ВОЄННИЙ ЧАС

3.1 Побудова економетричних моделей за трьома культурами та оцінка їх ефективності

Робота виконана на даних ПроКредит Банку (додатки Б, В, Г), наданих під час проходження виробничої практики. Всі розрахунки проведено в середовищі R-studio. Розглядається вибірка даних, яка складається з показників 70 малих підприємств, що займаються вирощуванням трьох культур: пшениці, кукурудзи та соняшнику.

Визначається залежність прибутковості найпоширеніших сільськогосподарських культур від регіону, засіяної площі, валового збору, виробничої собівартості культури, фактичної реалізації культури, виручки за реалізацію, ціни, вкладення на га. Змінна регіон подана значеннями від 1 до 3, де 1 це – Південь України (у цей регіон увійшли області: Одеська, Миколаївська, Херсонська та Запорізька), 2 – Центр (Дніпровська, Вінницька, Київська, Кіровоградська, Черкаська, Полтавська, Донецька, Луганська, Харківська, Чернігівська та Сумська) та 3 – Захід (Закарпатська, Івано-Франківська, Рівненська, Хмельницька, Чернівецька, Львівська, Тернопільська та Волинська).

Введемо позначення. Нехай:

Y – валовий прибуток від одного сільськогосподарського циклу, млн.грн;

X_1 – регіон, де знаходяться землі;

X_2 – зібрана площа, га;

X_3 – валовий збір, тонн;

X_4 – виробнича собівартість культури, млн. грн;

X_5 – фактична реалізація продукції, тонн;

X_6 – виручка за реалізацію продукції, млн. грн;

X_7 – ціна за 1 тону продукції, грн;

X_8 – врожайність культур, тонн/га;

X_9 – вкладення на 1 гектар, грн.

Для визначення кореляційних зв'язків між факторними та залежною змінною побудовано кореляційну матрицю (рис. 3.1)

	x2	x3	x4	x5	x6	x7	x8	x9	Y
x2	1,000	0,916	0,891	0,793	0,767	0,144	-0,150	-0,183	0,302
x3		1,000	0,954	0,851	0,835	0,161	0,178	0,048	0,335
x4			1,000	0,841	0,823	0,156	0,131	0,154	0,166
x5				1,000	0,958	0,003	0,100	0,025	0,040
x6					1,000	0,199	0,131	0,050	0,195
x7						1,000	0,263	0,145	0,647
x8							1,000	0,737	0,148
x9								1,000	-0,155

Рисунок 3.1 – Кореляційна матриця факторних та залежної змінної
Джерело: розроблено автором на підставі даних ПроКредит Банку

З останнього стовпця матриці випливає, що найбільший вплив на результативний показник має ціна за 1 тону продукції (коефіцієнт кореляції між ними дорівнює 0,647), валовий збір та зібрана площа. Решта інших показників слабо корелюють з валовим прибутком одного циклу. Наявність кореляційних зв'язків між факторними змінними може бути ознакою існування мультиколінеарності змінних.

Однорідність вибірки перевіримо засобами однофакторного дисперсійного аналізу. Використовуючи функцію `oneway.test()` (рис. 3.2), виконаємо порівняння середніх груп вибраних даних та розрахуємо ймовірність того, чи будуть вони статистично ідентичними. Розглядається залежність прибутковості від регіону.

За результатами проведеного аналізу $p\text{-value} = 0,1142$ та $= 0,3896$, та $= 0,6545$ для пшениці, кукурудзи та соняшнику відповідно. Значення $p\text{-value}$, що перевищує 0,05 означає ймовірність отримати такі або ще більш екстремальні результати (тобто різницю між середніми значеннями прибутковості культур у різних

регіонах), якщо нульова гіпотеза є істинною. У цьому випадку нульова гіпотеза стверджує, що середні значення прибутковості однакові для всіх регіонів, отже немає достатніх доказів відхилити дану гіпотезу.

```
> oneway.test(wheat$Gross_profit~wheat$Region)

One-way analysis of means (not assuming equal variances)

data: wheat$Gross_profit and wheat$Region
F = 2.378, num df = 2.000, denom df = 23.996, p-value = 0.1142

> oneway.test(corn$Gross_profit~corn$Region)

One-way analysis of means (not assuming equal variances)

data: corn$Gross_profit and corn$Region
F = 0.97545, num df = 2.000, denom df = 27.614, p-value = 0.3896

> oneway.test(sunflower$Gross_profit~sunflower$Region)

One-way analysis of means (not assuming equal variances)

data: sunflower$Gross_profit and sunflower$Region
F = 0.43109, num df = 2.000, denom df = 25.436, p-value = 0.6545
```

Рисунок 3.2 - Результат функції oneway.test для пшениці, кукурудзи та соняшнику відповідно

Джерело: розроблено автором на підставі даних ПроКредит Банку.

Щоб перевірити нульові гіпотези, що свідчать про однаковість середніх для трьох груп вибраних даних, зазвичай використовують критерій Тьюкі (рис. 3.3). Його реалізовано за допомогою функції TukeyHSD() і такий результат можливих попарних порівнянь було отримано:

```

> T_wheat <- aov(wheat$Gross_profit~as.factor(wheat$Region))
> TukeyHSD(T_wheat)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = wheat$Gross_profit ~ as.factor(wheat$Region))

$`as.factor(wheat$Region)`
      diff       lwr       upr      p adj
2-1 -2.4957735 -5.9369561 0.945409 0.1985100
3-1  0.5825524 -3.6953989 4.860504 0.9429801
3-2  3.0783259 -0.8589207 7.015572 0.1541042

> T_corn <- aov(corn$Gross_profit~as.factor(corn$Region))
> TukeyHSD(T_corn)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = corn$Gross_profit ~ as.factor(corn$Region))

$`as.factor(corn$Region)`
      diff       lwr       upr      p adj
2-1  3.2418084 -4.436027 10.919644 0.5719118
3-1  2.4725657 -7.120811 12.065942 0.8109284
3-2 -0.7692427 -9.679962  8.141476 0.9766816

> T_sunflower <- aov(sunflower$Gross_profit~as.factor(sunflower$Region))
> TukeyHSD(T_sunflower)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = sunflower$Gross_profit ~ as.factor(sunflower$Region))

$`as.factor(sunflower$Region)`
      diff       lwr       upr      p adj
2-1 -2.1845571 -8.612512  4.243398 0.6953505
3-1 -0.6484744 -8.680138  7.383189 0.9795716
3-2  1.5360827 -5.924053  8.996219 0.8746339

```

Рисунок 3.3 - Результат функції TukeyHSD для пшениці, кукурудзи та соняшнику відповідно

Джерело: розроблено автором на підставі даних ПроКредит Банку.

de diff – різниця груп;

lwr та upr – довірчі інтервали.

За результатами застосування критерію Тьюки, отримали попарні порівняння в залежності від регіону для трьох різних культур. Для кожної групи попарних порівнянь $p\text{-value} > 0.05$, що свідчить про те, що немає статистично значущих різниць між середніми значеннями валового прибутку для різних регіонів.

Це означає, що наявні дані не надають достатніх доказів для того, щоб стверджувати, що середні значення валового прибутку різняться статистично значущим чином між різними регіонами.

Побудуємо графік критерію Тьюкі (рис. 3.4) для наглядного розуміння попарних порівнянь. Якщо лінії не перетинають нуль, то це означає, що існує найбільша різниця:

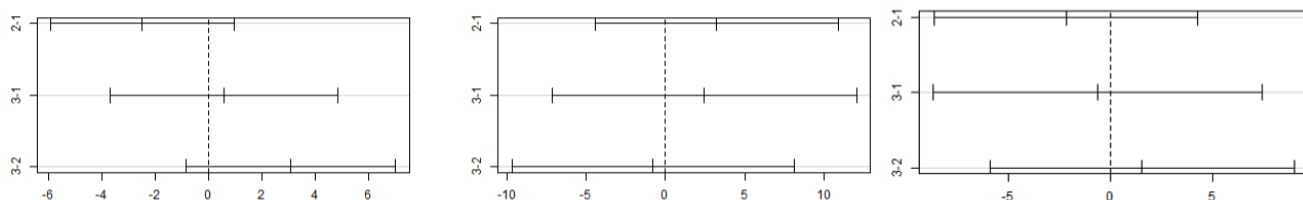


Рисунок 3.4 - Графік використання функції TukeyHSD для пшениці, кукурудзи та соняшнику відповідно

Джерело: розроблено автором на підставі даних ПроКредит Банку.

Отже, як видно з графіку, всі три моделі перетнули значення «0», що говорить про те, що найбільшої різниці не існує.

На підставі проведеного дисперсійного аналізу побудуємо моделі множинної регресії прибутковості культур для всієї сукупності спостережень без виокремлення регіонів. У моделі, що залежала від факторів X_1 - X_9 , два останніх параметри виявилися статистично незначущими, тому виключаємо їх із розгляду і будуємо модель, що містить параметри X_1 - X_7 .

Для побудови економетричної моделі прибутковості пшениці використовується функція `lm()`. Результати її застосування представлені на рис. 3.5:

```
Call:
lm(formula = Gross_profit ~ Region + Cultivated_area + Gross_out +
    Crop_cost + Sales + Gross_revenue + Price + Crop_yield +
    Investments_per_1_ha, data = wheat)
```

Coefficients:					
(Intercept)	Region	Cultivated_area	Gross_out	Crop_cost	Sales
-11.235392	1.238336	0.004128	0.006405	-1.146747	-0.003191
Gross_revenue	Price	Crop_yield	Investments_per_1_ha		
0.350095	0.001653	-0.267216	0.037574		

Рисунок 3.5 - Результат використання функції `Lm()` для пшениці

Джерело: розроблено автором на підставі даних ПроКредит Банку.

Завдяки функції `summary()`, можна спостерігати, що для цієї моделі коефіцієнт детермінації $R^2 = 0,8848$, тобто значення прибутковості на 88.48% визначається значеннями факторних змінних, що увійшли до моделі. Співвідношення між модельними та спостереженими значення подано на рис. 3.6:



Рисунок 3.6 - Модельні та спостережені значення валового прибутку від вирощування пшениці для моделі з усіма факторними змінними (X_1 - X_9) для всіх регіонів

Джерело: розроблено автором на підставі даних ПроКредит Банку

Як відомо, в багатофакторних моделях значення коефіцієнта детермінації автоматично зростає при збільшенні кількості факторних змінних. Тому у цьому разі його значення коригується з урахуванням степенів свободи і дорівнює 0,86.

Аналогічні моделі побудуємо для кукурудзи (рис. 3.7) та соняшнику (рис. 3.8):

```
Call:
lm(formula = Gross_profit ~ Region + Cultivated_area + Gross_out +
    Crop_cost + Sales + Gross_revenue + Price, data = corn)
```

Coefficients:	(Intercept)	Region	Cultivated_area	Gross_out	Crop_cost	Sales	Gross_revenue	Price
	-7.0831602	0.2337881	0.0021653	0.0065827	-1.1162834	-0.0029614	0.5329586	0.0009894

Рисунок 3.7 - Результат використання функції Lm() для кукурудзи

Джерело: розроблено автором на підставі даних ПроКредит Банку.

```
Call:
lm(formula = Gross_profit ~ Region + Cultivated_area + Gross_out +
    Crop_cost + Sales + Gross_revenue + Price, data = sunflower)
```

Coefficients:	(Intercept)	Region	Cultivated_area	Gross_out	Crop_cost	Sales	Gross_revenue	Price
	-3.090e-01	-4.824e-02	4.222e-04	1.317e-02	-1.016e+00	-1.395e-02	1.037e+00	4.541e-05

Рисунок 3.8 - Результат використання функції Lm() для соняшнику

Джерело: розроблено автором на підставі даних ПроКредит Банку.

В результаті застосування МНК побудовано такі економетричні моделі:

- Для пшениці:

$$Y = -11.24 + 1.24X_1 + 0.0041X_2 + 0.0064X_3 - 1.15X_4 - 0.0032X_5 + 0.35X_6 + 0.0017X_7; \quad (3.1)$$

- Для кукурудзи:

$$Y = -7.083 + 0.2338X_1 + 0.0022X_2 + 0.0066X_3 - 1.116X_4 - 0.003X_5 + 0.533X_6 + 0.00099X_7; \quad (3.2)$$

- Для соняшнику:

$$Y = -0.309 - 0.0482X_1 + 0.00042X_2 + 0.0132X_3 - 1.016X_4 - 0.014X_5 + 1.037X_6 + 0.000045X_7. \quad (3.3)$$

Завдяки функції summary(), можна спостерігати, що:

1) Коефіцієнт детермінації для пшениці $R^2 = 0,884$, для кукурудзи $R^2 = 0,9816$ та соняшнику $R^2 = 0,9925$, тобто значення прибутковості на 88.4%, 98.16% та 99,25% відповідно визначається за значеннями факторних змінних, що увійшли до моделі. Тобто, модельні значення досить близькі до реальних.

2) Коефіцієнт кореляції (R) становить для пшениці – 0,9402, для кукурудзи – 0,9908, для соняшнику – 0,9962, це показує, що між залежною та незалежними змінними існує досить сильний лінійний зв'язок.

3) Моделі є статистично значущими та адекватними, оскільки $F > F_t$, а $p\text{-value} < 0.05$.

4) для трьох різних культур не всі коефіцієнти є значущими. А саме: для пшениці значущі коефіцієнти мають всі змінні, але площа збору та виручка мають меншу значущість порівняно з іншими; для кукурудзи: вільний член, валовий збір, виробнича собівартість, реалізація, виручка від реалізації та ціна; та для соняшнику: валовий збір, виробнича собівартість, реалізація, виручка за реалізацію та ціна (ціна є найменш значущою для перелічених змінних).

3.2 Дослідження аналітичних якостей моделей прибутковості трьох культур

Отримані в попередньому параграфі результати засвідчують досить високі показники їх статистичної значущості. Але аналіз моделі буде неповним без аналізу

залишків. Тому залишки моделей перевіримо на нормальність та гомоскедастичність.

Побудовані гістограми залишків для моделі прибутковості пшениці (рис. 3.9), кукурудзи (рис. 3.10) та соняшнику (рис. 3.11) мають розподіл похибок, подібний до нормального розподілу.

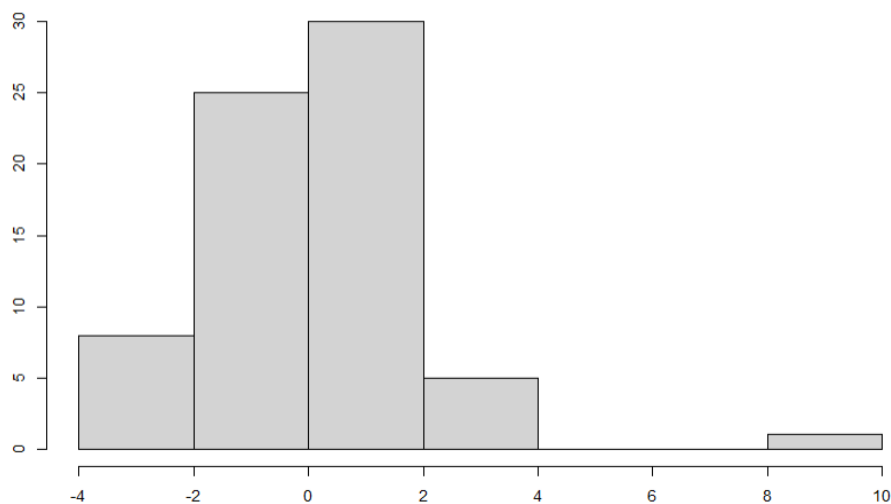


Рисунок 3.9 - Гістограма похибок для пшениці

Джерело: розроблено автором на підставі даних ПроКредит Банку.

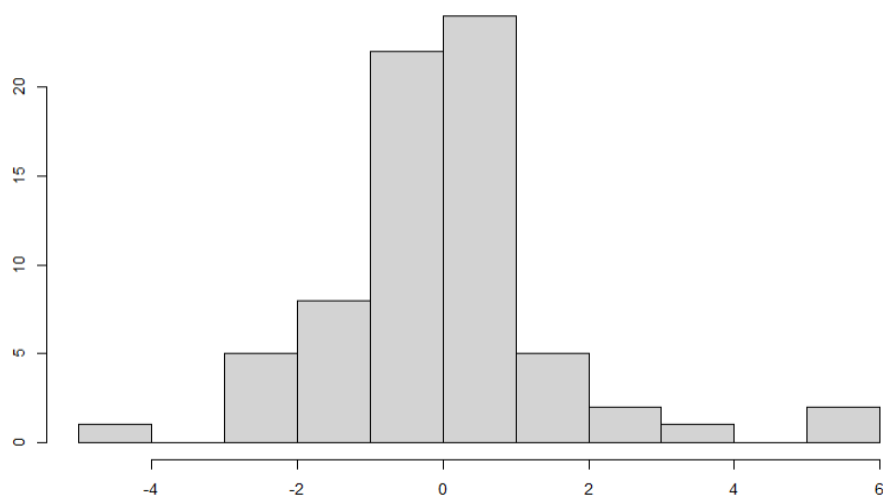


Рисунок 3.10 - Гістограма похибок для кукурудзи

Джерело: розроблено автором на підставі даних ПроКредит Банку.

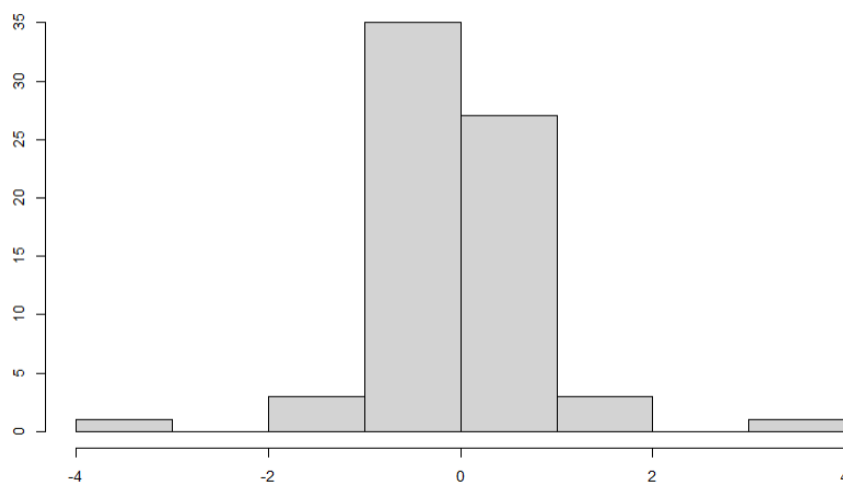


Рисунок 3.11 - Гістограма похибок для соняшнику
Джерело: розроблено автором на підставі даних ПроКредит Банку.

На ймовірнісних графіках Normal Q-Q (рис. 3.12) видно, що дані переважно лежать на прямій, яка має кут нахилу менший, ніж 45 градусів. Однак по краях графіка спостерігаються досить значні відхилення від прямої, що може свідчити про наявність аномальних даних (викидів). Це означає, що набір даних не підпорядкований нормальному закону.

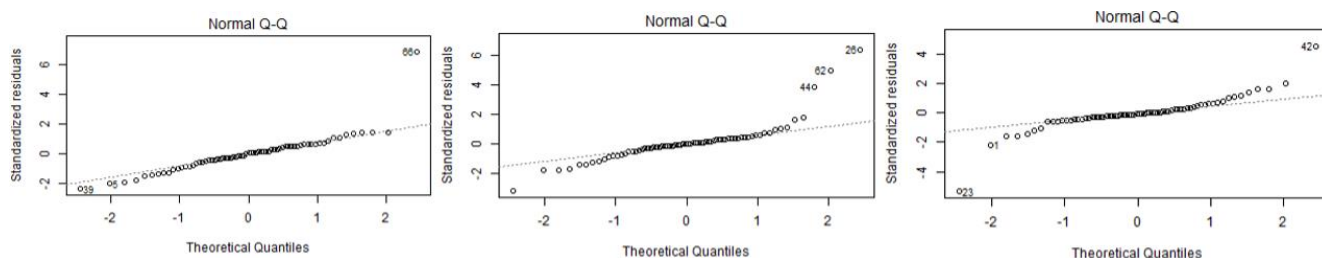


Рисунок 3.12 - Normal Q-Q для трьох культур
Джерело: розроблено автором на підставі даних ПроКредит Банку.

Діаграми розсіювання залишків засвідчують присутність досить помітної нелінійної тенденції в залишках. На графіках Residuals vs Fitted (рис. 3.13) видно, що червона лінія є не зовсім рівною. Це говорить про те, що залишки неоднаково розподілені у всьому діапазоні встановлених значень. Вони більше скупчені в одному місці, і на графіках присутні викиди з аномальними значеннями. Графіки подані за такою послідовністю: пшениця, кукурудза та соняшник.

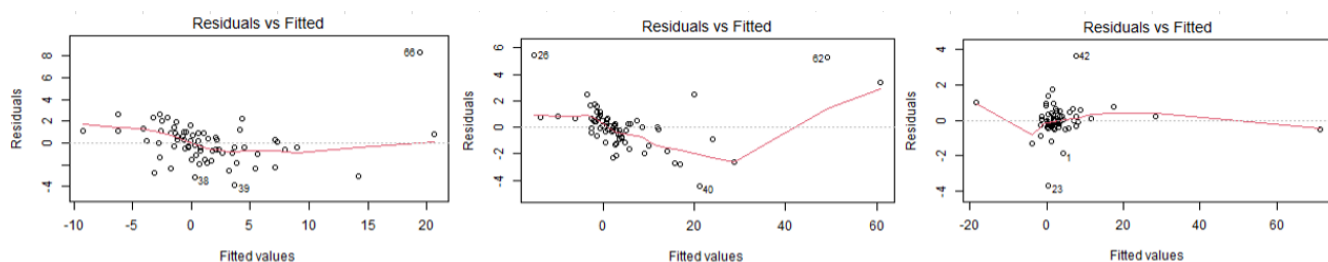


Рисунок 3.13 - Residuals vs Fitted для трьох культур

Джерело: розроблено автором на підставі даних ПроКредит Банку.

На графіках Scale–Location та Residuals vs Leverage (рис. 3.14) бачимо, що всі дані знаходяться в купі, що свідчить про низьку напруженість.

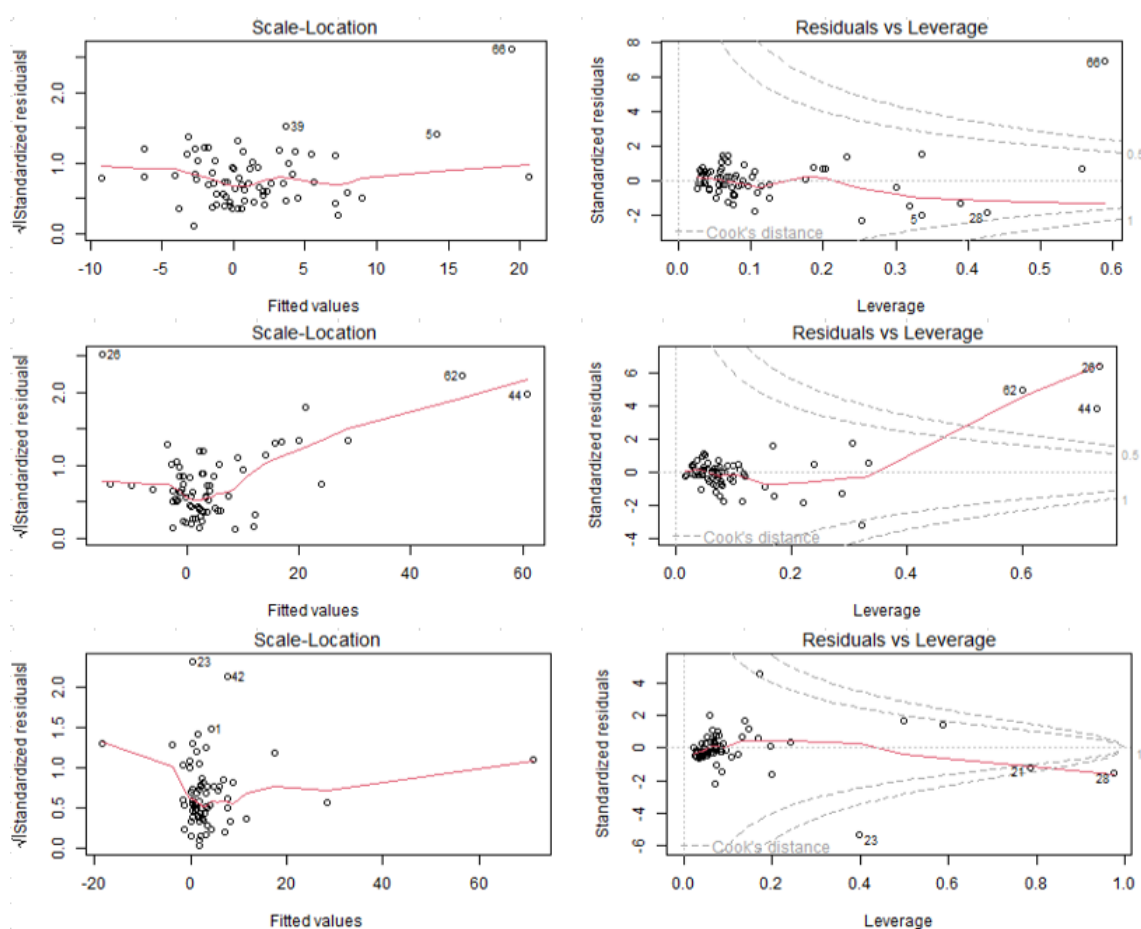


Рисунок 3.14 - Scale-Location and Residuals vs Leverage для трьох культур

Джерело: розроблено автором на підставі даних ПроКредит Банку.

Результати функції `gvlma()` (рис. 3.15) показують, що статистика, асиметрія, ексцес, лінійна залежність та гетероскедастичність не є присутніми в моделі прибутковості пшениці. Стосовно моделі кукурудзи, можна побачити, що функція `gvlma()` вказує на те, що залишки моделі не задовільняють умови

гомоскедастичності, тобто є гетероскедастичними. Для моделі соняшнику, то результат є інакшим, а саме ця модель є лінійно залежною та має асиметрію.

```
Call:
gvlma(x = lm_W)
```

	Value	p-value	Decision
Global Stat	123.850	0.000e+00	Assumptions NOT satisfied!
Skewness	16.148	5.858e-05	Assumptions NOT satisfied!
Kurtosis	80.938	0.000e+00	Assumptions NOT satisfied!
Link Function	20.774	5.167e-06	Assumptions NOT satisfied!
Heteroscedasticity	5.989	1.439e-02	Assumptions NOT satisfied!

```
Call:
gvlma(x = lm_C)
```

	Value	p-value	Decision
Global Stat	78.733	3.331e-16	Assumptions NOT satisfied!
Skewness	6.458	1.104e-02	Assumptions NOT satisfied!
Kurtosis	26.283	2.949e-07	Assumptions NOT satisfied!
Link Function	44.090	3.136e-11	Assumptions NOT satisfied!
Heteroscedasticity	1.901	1.679e-01	Assumptions acceptable.

```
Call:
gvlma(x = lm_S)
```

	Value	p-value	Decision
Global Stat	2.251e+02	0.00000	Assumptions NOT satisfied!
Skewness	1.216e-03	0.97218	Assumptions acceptable.
Kurtosis	2.203e+02	0.00000	Assumptions NOT satisfied!
Link Function	7.368e-01	0.39069	Assumptions acceptable.
Heteroscedasticity	4.057e+00	0.04398	Assumptions NOT satisfied!

Рисунок 3.15 - Результат функції gvlma для пшениці, кукурудзи та соняшнику відповідно

Джерело: розроблено автором на підставі даних ПроКредит Банку.

Повертаючись до кореляційної матриці, які містить високі значення парних коефіцієнтів, маємо перевірити мультиколінеарність факторних змінних.

Результат застосування функції vif() (рис. 3.16) показує, що фактор інфляції дисперсії по трьох культурах приблизно знаходиться в межах від 1,2 до 47.3. За різними джерелами порогове значення може бути: 2, 5, 10. Отже, маємо такий результат, що між j - предиктором та іншими змінними предиктора існує кореляції, а отже, є серйозні ознаки мультиколінеарності.

```
> vif(lm_W)
```

Region	Cultivated_area	Gross_out	Crop_cost	Sales	Gross_revenue	Price
1.255893	7.725999	15.662722	12.039316	33.104166	28.083789	2.464852

```
> vif(lm_C)
```

Region	Cultivated_area	Gross_out	Crop_cost	Sales	Gross_revenue	Price
1.112849	4.201297	4.504763	4.487624	13.326855	15.584639	1.875364

```
> vif(lm_S)
```

Region	Cultivated_area	Gross_out	Crop_cost	Sales	Gross_revenue	Price
1.227265	8.409672	17.618218	8.355314	47.300289	37.222558	2.780052

Рисунок 3.16 - Результат функції vif() для пшениці, кукурудзи та соняшнику відповідно

Джерело: розроблено автором на підставі даних ПроКредит Банку.

Це спонукає перейти до тесту Фаррара - Глаубера (F - тест) щодо більш детального дослідження мультиколінеарності.

Стовпці VIF, TOL та W_i забезпечують діагностичний вихід коефіцієнта дисперсії, допуску та F-тесту Фаррара-Глаубера відповідно.

Аналізуючи модель прибутковості пшениці (рис. 3.17) на виявлення, у яких саме змінних присутня мультиколінеарність, було отримано такий результат:

```
> imcdiag(lm_W)

Call:
imcdiag(mod = lm_W)

All Individual Multicollinearity Diagnostics Result
```

	VIF	TOL	W_i	F_i	Leamer	CVIF	Klein	IND1	IND2
Region	1.2559	0.7962	2.6442	3.2243	0.8923	0.5686	0	0.0771	0.2614
Cultivated_area	7.7260	0.1294	69.5020	84.7476	0.3598	3.4980	0	0.0125	1.1170
Gross_out	15.6627	0.0638	151.5148	184.7503	0.2527	7.0915	1	0.0062	1.2011
Crop_cost	12.0393	0.0831	114.0729	139.0954	0.2882	5.4509	1	0.0080	1.1764
Sales	33.1042	0.0302	331.7431	404.5125	0.1738	14.9883	1	0.0029	1.2443
Gross_revenue	28.0838	0.0356	279.8658	341.2557	0.1887	12.7152	1	0.0034	1.2373
Price	2.4649	0.4057	15.1368	18.4571	0.6369	1.1160	0	0.0393	0.7625

```

1 --> COLLINEARITY is detected by the test
0 --> COLLINEARITY is not detected by the test

* all coefficients have significant t-ratios

R-square of y on all x: 0.884

```

Рисунок 3.17 - Результат функції `imcdiag()` для пшениці
Джерело: розроблено автором на підставі даних ПроКредит Банку.

Отже, F-статистика для змінних "валовий збір", "виробнича собівартість", "реалізація" та "виручка за реалізацію" є досить високою. Отже, F-тест показує, що ці змінні будуть основною причиною мультиколінеарності.

Щодо моделі прибутковості кукурудзи (рис. 3.18), то тут функція `imcdiag()` не виявила змінні, де присутня мультиколінеарність. Аналогічна ситуація і для моделі соняшнику (рис. 3.19).

```
> imcdiag(lm_C)
```

```
Call:
imcdiag(mod = lm_C)
```

All Individual Multicollinearity Diagnostics Result

	VIF	TOL	Wi	Fi	Leamer	CVIF	Klein	IND1	IND2
Region	1.1128	0.8986	1.1849	1.4445	0.9479	-0.0230	0	0.0856	0.1496
Cultivated_area	4.2013	0.2380	33.6136	40.9766	0.4879	-0.0868	0	0.0227	1.1238
Gross_out	4.5048	0.2220	36.8000	44.8610	0.4712	-0.0930	0	0.0211	1.1475
Crop_cost	4.4876	0.2228	36.6201	44.6416	0.4721	-0.0927	0	0.0212	1.1462
Sales	13.3269	0.0750	129.4320	157.7837	0.2739	-0.2752	0	0.0071	1.3642
Gross_revenue	15.5846	0.0642	153.1387	186.6834	0.2533	-0.3218	0	0.0061	1.3802
Price	1.8754	0.5332	9.1913	11.2047	0.7302	-0.0387	0	0.0508	0.6884

```
1 --> COLLINEARITY is detected by the test
0 --> COLLINEARITY is not detected by the test
```

Region , Cultivated_area , coefficient(s) are non-significant may be due to multicollinearity

R-square of y on all x: 0.9816

Рисунок 3.18 - Результат функції `imcdiag()` для кукурудзи
Джерело: розроблено автором на підставі даних ПроКредит Банку.

```
> imcdiag(lm_S)
```

```
Call:
imcdiag(mod = lm_S)
```

All Individual Multicollinearity Diagnostics Result

	VIF	TOL	Wi	Fi	Leamer	CVIF	Klein	IND1	IND2
Region	1.2273	0.8148	2.3863	2.9090	0.9027	-0.0110	0	0.0776	0.2365
Cultivated_area	8.4097	0.1189	77.8016	94.8438	0.3448	-0.0755	0	0.0113	1.1250
Gross_out	17.6182	0.0568	174.4913	212.7132	0.2382	-0.1581	0	0.0054	1.2044
Crop_cost	8.3553	0.1197	77.2308	94.1480	0.3460	-0.0750	0	0.0114	1.1241
Sales	47.3003	0.0211	486.1530	592.6437	0.1454	-0.4245	0	0.0020	1.2499
Gross_revenue	37.2226	0.0269	380.3369	463.6487	0.1639	-0.3341	0	0.0026	1.2426
Price	2.7801	0.3597	18.6905	22.7847	0.5998	-0.0249	0	0.0343	0.8176

```
1 --> COLLINEARITY is detected by the test
0 --> COLLINEARITY is not detected by the test
```

Region , Cultivated_area , coefficient(s) are non-significant may be due to multicollinearity

R-square of y on all x: 0.9925

Рисунок 3.19 - Результат функції `imcdiag()` для соняшнику
Джерело: розроблено автором на підставі даних ПроКредит Банку.

Наступний крок – дослідження мультиколінеарності в групах даних для окремих регіонів.

Після групування даних щодо вирощування пшениці по регіонах (рис. 3.20), були отримані такі результати. А саме, що мультиколінеарність не виявлена ні в жодній з трьох груп даних по регіонах.

```

> imcdiag(lm_W_1)

Call:
imcdiag(mod = lm_W_1)

All Individual Multicollinearity Diagnostics Result

      VIF      TOL      Wi      Fi Leamer      CVIF Klein      IND1      IND2
Cultivated_area  9.0101 0.1110 24.0304 32.0405 0.3331 -0.1075      0 0.0370 0.9916
Gross_out       12.3535 0.0809 34.0604 45.4138 0.2845 -0.1474      0 0.0270 1.0251
Crop_cost       13.5919 0.0736 37.7757 50.3676 0.2712 -0.1622      0 0.0245 1.0333
Sales           58.5042 0.0171 172.5125 230.0167 0.1307 -0.6980      0 0.0057 1.0963
Gross_revenue   42.9730 0.0233 125.9190 167.8920 0.1525 -0.5127      0 0.0078 1.0894
Price           3.1772 0.3147  6.5316  8.7088 0.5610 -0.0379      0 0.1049 0.7643

1 --> COLLINEARITY is detected by the test
0 --> COLLINEARITY is not detected by the test

* all coefficients have significant t-ratios

R-square of y on all x: 0.9965

```

Рисунок 3.20 - Результат функції `imcdiag()` для пшениці за Південним регіоном

Джерело: розроблено автором на підставі даних ПроКредит Банку.

На відсортованих за регіонами даних побудовано модель прибутковості пшениці з урахуванням змінних $X_2 - X_9$. Отриманий при цьому коефіцієнт детермінації значно перевищує його значення для попередньої моделі, а саме $R^2 = 0,9966$, (скоригований $R^2 = 0,993$). У цій моделі параметри при змінних X_8 та X_9 виявилися статистично незначущими, тому розглядається модель без цих змінних.

Модель залежності прибутковості пшениці від змінних $X_2 - X_7$ для південного регіону показала високе значення $R^2 = 0,9965$ (скоригований $R^2 = 0,9945$), та значущість всіх параметрів моделі (рис. 3.21).

```
> summary(lm_W_1)

Call:
lm(formula = Gross_profit ~ Cultivated_area + Gross_out + Crop_cost +
    Sales + Gross_revenue + Price, data = wheat_1)

Residuals:
    Min       1Q   Median       3Q      Max
-0.64871 -0.27935 -0.01998  0.37966  0.53301

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.204e+00  5.903e-01  -5.427 8.91e-05 ***
Cultivated_area -1.734e-03  5.214e-04  -3.326  0.005 **
Gross_out       7.357e-03  1.726e-04  42.619 3.22e-16 ***
Crop_cost      -1.030e+00  4.126e-02 -24.958 5.25e-13 ***
Sales          -4.591e-03  4.197e-04 -10.940 3.04e-08 ***
Gross_revenue   6.201e-01  5.875e-02  10.556 4.76e-08 ***
Price          5.835e-04  9.474e-05   6.160 2.48e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4304 on 14 degrees of freedom
Multiple R-squared:  0.9965, Adjusted R-squared:  0.9951
F-statistic: 672.9 on 6 and 14 DF, p-value: < 2.2e-16
```

Рисунок 3.21 – Множинна регресія прибутковості пшениці для першого регіону

Джерело: розроблено автором на підставі даних ПроКредит Банку

На рис. 3.22 представлено спостережені та модельні значення результативного показника для першого регіону, обчислені за всіма розглянутими моделями:

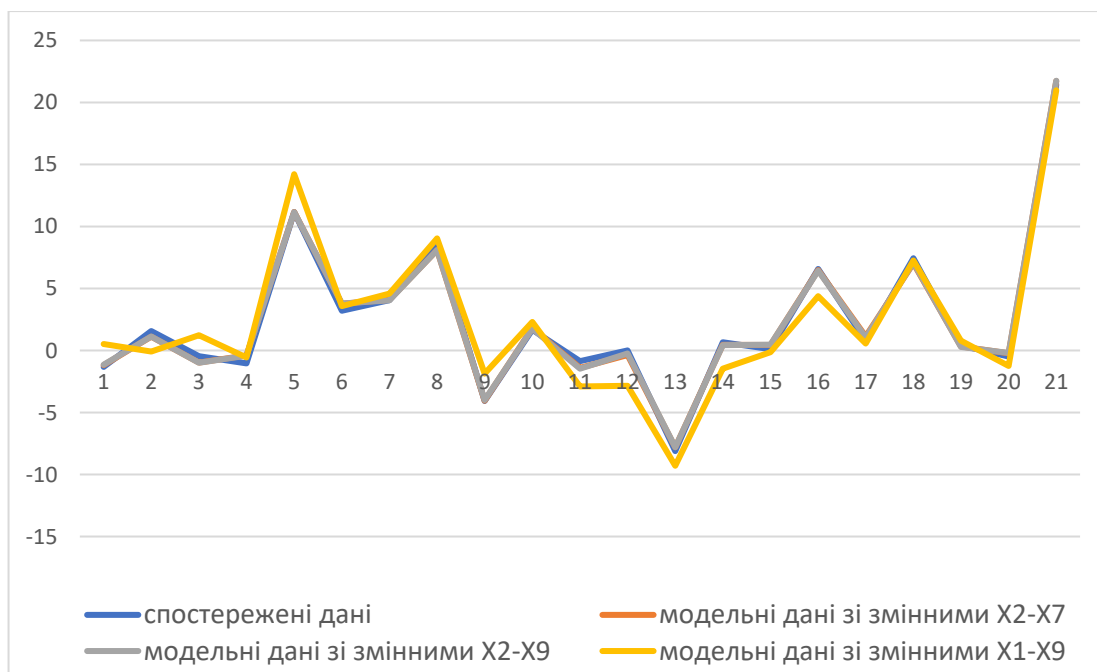


Рисунок 3.22 – Спостережені та модельні значення валового прибутку від вирощування пшениці

Джерело: розроблено автором на підставі даних ПроКредит Банку

Аналогічно побудовані моделі прибутковості пшениці зі змінними X_2 - X_7 для двох інших регіонів показали значення коефіцієнта детермінації 0,961 та 0,9845 відповідно і достатньо близькі до спостережених модельні значення результативного показника.

Аналіз залишків побудованих моделей показав, що в залишках моделі кукурудзи було виявлено ознаки гетероскедастичності за тестом Бройша-Пагана (рис. 3.23).

```
Breusch-Pagan test

data:  lm_C
BP = 131.01, df = 7, p-value < 2.2e-16
```

Рисунок 3.23 - Результат функції `bptest()` для кукурудзи
Джерело: розроблено автором на підставі даних ПроКредит Банку.

Результати тесту показали значущу гетероскедастичність залишків моделі зі значенням статистики $BP = 131,01$ і кількістю ступенів свободи $df = 7$. При цьому отриманий p -value менше за 0,05, що свідчить про статистичну значущість гетероскедастичності.

Щоб усунути наявність гетероскедастичності в моделі прибутковості кукурудзи було застосовано метод групування моделі по регіонах. Отже, отримали три моделі, які мають такі результати за тестом Бройша-Пагана (рис. 3.24):

```
> # Тест Бройша-Пагана
> bptest(lm_C_1, studentize = FALSE)

Breusch-Pagan test

data:  lm_C_1
BP = 13.806, df = 6, p-value = 0.03189

> bptest(lm_C_2, studentize = FALSE)

Breusch-Pagan test

data:  lm_C_2
BP = 21.137, df = 6, p-value = 0.001733

> bptest(lm_C_3, studentize = FALSE)

Breusch-Pagan test

data:  lm_C_3
BP = 2.0865, df = 6, p-value = 0.9116
```

Рисунок 3.24 - Результат функції `bptest()` для кукурудзи по трьох регіонах
Джерело: розроблено автором на підставі даних ПроКредит Банку

Як видно з проведеного аналізу, для регіону першого та третього гетероскедастичність не було виявлено, оскільки p-value більше за 0,05. Для другого регіону ситуація протилежна.

Для усунення гетероскедастичності в моделі прибутковості кукурудзи по другому регіону було застосовано метод вагової регресії, де ваги будуть враховувати нерівномірність дисперсії залишків. Один з підходів полягає у використанні методу робастної стандартної помилки (robust standard errors).

У RStudio, для використання робастних стандартних помилок можна використовувати функцію `coefest()`. Підсумок по застосуванню функції `coefest()` (рис. 3.25):

```
> summary(lm_C_robust)
```

	Estimate	Std. Error	t value	Pr(> t)
Min.	:-5.279190	Min. :0.0002869	Min. : -15.5747	Min. :0.000e+00
1st Qu.	:-0.561389	1st Qu.:0.0005040	1st Qu.: -4.1149	1st Qu.:6.000e-08
Median	: 0.000823	Median :0.0030063	Median : 1.6890	Median :3.181e-05
Mean	:-0.817803	Mean :0.2566360	Mean : 0.4621	Mean :1.625e-02
3rd Qu.	: 0.005265	3rd Qu.:0.0828866	3rd Qu.: 4.9812	3rd Qu.:5.507e-03
Max.	: 0.665997	Max. :1.6263771	Max. : 15.3878	Max. :1.027e-01

Рисунок 3.25 - Результат функції `coefest()` для кукурудзи по другому регіону
Джерело: розроблено автором на підставі даних ПроКредит Банку.

Результати аналізу застосування робастних стандартних помилок на модель можна описати так:

- **Estimate:** це стовпці містять оцінки коефіцієнтів для кожного незалежного змінного. Наприклад, оцінка для першого незалежного змінного становить -5,28 для другого -0,56, для третього 0 і т.д. Вони представляють зсув або зміну в залежній змінній (прибутковість), яку викликає зміна на одиницю відповідного незалежного змінного;
- **Std. Error:** це стовпці містять робастні стандартні помилки для кожного коефіцієнта. Вони вказують на точність оцінки кожного коефіцієнта;
- **t value:** це стовпці містять значення t-статистики для кожного коефіцієнта. Вони вказують на значущість кожного коефіцієнта. Чим більше (або менше в разі від'ємного значення) абсолютне значення t-статистики, тим більш значущим вважається відповідний коефіцієнт;

- $\Pr(>|t|)$: це стовпці містять p-value для кожного коефіцієнта. Вони вказують на статистичну значущість кожного коефіцієнта. Чим менше значення p-value, тим більш значущим вважається відповідний коефіцієнт. В даному випадку, для першого незалежного змінного p-value дорівнює 0, для другого 0,00000006, для третього 0,00003 і т.д.

На основі цих результатів можна зробити висновок про значущість кожного коефіцієнта. Коефіцієнти, які мають низькі p-value (менше 0.05), вважаються статистично значущими, тоді як коефіцієнти з високими p-value (більше 0.05) не є статистично значущими. Будь-який коефіцієнт з високим p-value може бути видалений з моделі, якщо він не вносить значного внеску до пояснення варіації в залежній змінній.

Для подальшого проведення аналізу було прибрано змінну виручка за реалізацію, оскільки вона досить тісно корелює з продажами. Після повторного проведення на тест Бройша-Пагана (рис. 3.26), отримали такі результати, що p-value більше за 0,05, отже гетероскедастичності в моделі не виявлено. Коефіцієнт детермінації становить 0,9364. Отже, значення прибутковості на 93,64% визначається за значеннями факторних змінних, що увійшли до моделі.

```
> lm_C_robust <- lm(data=corn_2, Gross_profit ~ Cultivated_area + Gross_out + Crop_cost + Sales + Price)
> lmtest::bptest(lm_C_robust)

studentized Breusch-Pagan test

data: lm_C_robust
BP = 15.955, df = 5, p-value = 0.006973
```

Рисунок 3.26 - Результат функції `coefTest()` для кукурудзи по другому регіону
Джерело: розроблено автором на підставі даних ПроКредит Банку.

Після усунення гетероскедастичності за допомогою робастних стандартних помилок та розбиттю моделі по регіонах дозволяє вважати її більш надійною і точною при оцінюванні впливу факторів на прибутковість пшениці.

ВИСНОВКИ

Задачі підвищення рівня врожайності та прибутковості сільськогосподарських культур є актуальними в усі часи, а в умовах війни, коли скорочуються посівні площі, ускладнюються умови вирощування та реалізації продукції, вони набувають надзвичайної актуальності. Втрати, які мали сільгосппідприємства в цей період, змушують їх звертатися за позиками у фінансові установи, які в свою чергу мають перевіряти надійність клієнта зокрема з огляду на його заплановану діяльність. У зв'язку з чим прогнозування майбутніх прибутків, які можуть отримувати сільгоспвиробники від вирощування певних культур, викликають інтерес у їхніх можливих кредиторів.

Дослідження проведено на основі статистичної інформації, наданої ПроКредит банком автору під час проходження виробничої практики.

Для дослідження були обрані найбільш популярні для вирощування культури в Україні, а саме: пшениця, кукурудза та соняшник. Ці культури показували найбільшу прибутковість в умовах українського ринку та ресурсних обмежень.

За результативний показник обрано прибуток від вирощування сільськогосподарських культур. В якості незалежних змінних розглядалися такі показники як: регіон, зібрана площа, валовий збір, виробнича собівартість культури, фактична реалізація продукції, виручка за реалізацію продукції, ціна за 1 тону продукції, врожайність культур, та вкладення на 1 гектар. Саме вони найбільше впливають на прибутковість культур.

У даному дослідженні було проведено моделювання прибутковості культур в агропромисловому комплексі за допомогою лінійних багатофакторних моделей. Побудові моделей передували кореляційний та дисперсійний аналіз даних. Кореляційна матриця, побудова для залежної та всіх незалежних змінних, дала змогу виявити щільність зв'язків між змінними, а однофакторний дисперсійний

аналіз дав змогу перевірити однорідність даних всієї вибірки не залежно від регіону вирощування культур.

Побудовані регресійні моделі для трьох найпопулярніших в Україні культур виявилися статистично значущими, але в даних була виявлена мультиколінеарність, а в залишках гетероскедастичність. Перехід до даних, групованих за регіонами, дав змогу позбутися мультиколінеарності, але також підвищити коефіцієнт детермінації. Параметри моделей, що враховують шість із дев'яти факторних змінних, виявилися статистично значущими. В деяких моделях була виявлена гетероскедастичність залишків, яка була подолана зміною специфікації моделі.

Моделі з високим значенням коефіцієнта детермінації навіть у разі мультиколінеарності даних можуть бути використані для прогнозування результативного показника, що, безумовно, є цікавим для кредиторів, а моделі без мультиколінеарності даних зі статистично значущими параметрами можуть бути використані для аналізу впливу незалежних змінних на результативний показник і є цікавими для виробників продукції, адже дають змогу цілеспрямовано змінювати його значення за рахунок керованих незалежних змінних.

Актуальність цієї роботи для фермерів та їхніх кредиторів полягає у можливості зниження ризиків та підвищення ефективності вирощування культур. Знання про те, які фактори впливають на прибутковість культур, дозволяє фермерам зосередитися на оптимальних варіантах вирощування, уникати неправильних висновків і робити рішення більш зваженими і знижувати витрати.

Загалом, дане дослідження підтверджує важливість моделювання прибутковості культур в агропромисловому комплексі. Лінійні багатofакторні моделі виявились корисними інструментами для аналізу та прогнозування прибутковості. Подальше застосування результатів цієї роботи може сприяти ефективнішому використанню ресурсів, підвищенню прибутковості та зниженню ризиків у сільському господарстві.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Klima K., Kliszcz A., Puła J., Lepiarczyk A. Yield and Profitability of Crop Production in Mountain Less Favoured Areas: Article, 2020, 1-10 p. URL: <https://www.mdpi.com/2073-4395/10/5/700>. (accessed: 05.04.2023).
2. Chand R., Saxena R., Simmi S. Estimates and Analysis of Farm Income in India: Article, 2015, 139-147 p.
3. An analysis of net farm income to guide agricultural policies in the Ethiopian highlands: Article / Okoth P., Stoorvogel J., Heesmans H. and others, 2022, 1-13 p. URL: <https://agricultureandfoodsecurity.biomedcentral.com/articles/10.1186/s40066-022-00404-2>. (accessed: 05.04.2023).
4. Finger R., El Benni N. Farm income in European agriculture: new perspectives on measurement and implications for policy evaluation: Article, 2021, 1-13 p. URL: <https://academic.oup.com/erae/article/48/2/253/6134529>. (accessed: 05.04.2023).
5. Li Y., Wang Z. Analysis on the effect of farmer income of policy-based agricultural insurance: Article, 2022, 1-15 p. URL: <https://www.tandfonline.com/doi/full/10.1080/09064710.2021.2005820>. (accessed: 05.04.2023).
6. Highlights from the Farm Income Forecast /U.S. Department of Agriculture: Article, 2023. URL: <https://www.ers.usda.gov/topics/farm-economy/farm-sector-income-finances/highlights-from-the-farm-income-forecast/>. (accessed: 05.04.2023).
7. Kader J., Delseny M. Advances in Botanical Research: Book, 2010, 50-75p.
8. Грицюк П., Бачишина Л. Моделювання динаміки урожайності зернових культур для областей України: Київ: КНЕУ, 2015, 189-199 с.
9. Кобець С., Тесьолкін О. Підхід до прогнозування врожайності озимої пшениці з урахуванням впливу основних гідрометеорологічних факторів: Миколаїв: університет ім. Сухомлинського, 2018, 701-705 с

10. Сотніков В.С., Львовичкіна О.О. Регресійна лінійна модель та прогнозування урожайності озимої пшениці. *Наукові праці Кіровоградського національного технічного університету. Економічні науки*: зб. наук. пр. Вип. 6. Кіровоград: КНТУ, 2004, 356–359 с.
11. Климко О. Моделювання рівня врожайності сільськогосподарських культур.: Полтава: НТУ, 2016, 1-3 с.
12. Лещенко Л. Економіко-математичне моделювання як фактор підвищення прибутковості у сільськогосподарських підприємствах. *Інвестиції: практика та досвід*. 2016. № 8. 77–80 с.
13. Спортивний туризм та активна рекреація: географія, систематизація, практика / Закон Енгеля. URL: <https://geohub.org.ua/node/1222> (accessed: 08.04.2023).
14. Дядик Т.В. Ринок праці в АПК: стан, проблеми, шляхи розвитку: *Наукові праці Полтавської державної аграрної академії*, 2013, 111-118 с.
15. Agricultural economics / established by Britannica: Article. URL: <https://www.britannica.com/topic/agricultural-economics/Government-intervention>. (accessed: 10.04.2023).
16. Косовська М. Агроекологія: стійкість, яку формує кожен та кожна з нас, 2023. URL: <https://ecoaction.org.ua/ahroekolohiia-stijkist.html>. (accessed: 10.04.2023).
17. The future of food and agriculture. Trends and challenges / established by Food and Agriculture Organization of the United Nations: books, 2017, 46-56 p.
18. Показники Мінфін: ВВП України 2021: станом на 10.04.2023 / Офіційний сайт Міністерства фінансів України, 2023. URL: <https://index.minfin.com.ua/ua/economy/gdp/2021/>. (accessed: 10.04.2023)
19. Data of Statista: Share of value added by the agricultural sector to the gross domestic product (GDP) in Ukraine from 2010 to 2021 / published by Statista Research Department, 2022. URL: [https://www.statista.com/statistics/1326724/ukraine-agriculture-share-gdp/#:~:text=In%202021%2C%20the%20agricultural%20sector,gross%20domestic%20product%20\(GDP\)](https://www.statista.com/statistics/1326724/ukraine-agriculture-share-gdp/#:~:text=In%202021%2C%20the%20agricultural%20sector,gross%20domestic%20product%20(GDP)). (accessed: 10.04.2023).

20. Top Lead Report from ProCredit Bank. Ukraine, 2022, 3-7 p.
21. Показники Державної служби статистики України: станом на 2023 / Офіційний сайт Державної служби статистики України. 2023. URL: <https://www.ukrstat.gov.ua/>. (accessed: 10.04.2023).
22. Навіть якщо «зернову угоду» продовжать, ціна фрахту зростатиме – експерт / опубліковано Укрінформ: Стаття, 2023. URL: <https://www.ukrinform.ua/rubric-economy/3708279-navit-akso-zernovu-ugodu-prodovzat-cina-frahtu-zrostatime-ekspert.html>. (accessed: 12.04.2023)
23. Econometric models are used to analyze time series data for the purpose of predicting future sales and market share levels / established by Encyclopedia of Information Systems: Article, 2003. URL: <https://www.sciencedirect.com/topics/economics-econometrics-and-finance/econometric-model>. (accessed: 20.04.2023).
24. Пічуров С., Калашнікова Н. Економетрія: підручник, 2004, 3-10 с. URL: https://lib.dsau.dp.ua/pub/ekonometriya_met.pdf. (accessed: 25.04.2023).
25. Least Square Method / established by BYJU'S: Article. URL: <https://byjus.com/maths/least-square-method/#:~:text=The%20least%20square%20method%20is,the%20points%20from%20the%20curve>. (accessed: 25.04.2023).
26. RMSE: Root Mean Square Error / established by Statistics How to: Article. URL: <https://www.statisticshowto.com/probability-and-statistics/regression-analysis/rmse-root-mean-square-error/>. (accessed: 25.04.2023)
27. Обчислення коефіцієнта детермінації. URL: <https://kc.pnu.edu.ua/wp-content/uploads/sites/11/2021/02/Computer-technology-in-chemistryLecture4.pdf>. (accessed: 25.04.2023).
28. Економічна енциклопедія: У трьох томах. Т. 1. / Редкол.: С. В. Мочерний (відп. ред.) та ін. К.: Видавничий центр “Академія”, 2000, 864 с.
29. Hessian T. F-Distribution, F-Statistic, F-Test: Article. URL: <https://sixsigmastudyguide.com/f-distributionf-statistic-f-test/>. (accessed: 25.04.2023).

30. T Test (Student's T-Test): Definition and Examples / established by Statistics How to: Article. URL: <https://www.statisticshowto.com/probability-and-statistics/t-test/>. (accessed: 25.04.2023).
31. Stepwise Regression / established by PennState Eberly College of Science: Article. URL: <https://online.stat.psu.edu/stat501/lesson/10/10.2>. (accessed: 28.04.2023).
32. What is Analysis of Variance (ANOVA) / established by TIBC: Article. URL: [https://www.tibco.com/reference-center/what-is-analysis-of-variance-anova#:~:text=Analysis%20of%20Variance%20\(ANOVA\)%20is,the%20means%20of%20different%20groups](https://www.tibco.com/reference-center/what-is-analysis-of-variance-anova#:~:text=Analysis%20of%20Variance%20(ANOVA)%20is,the%20means%20of%20different%20groups). (accessed: 28.04.2023).
33. Multicolinealidad / established by economipedia: Article. URL: <https://uk.economy-pedia.com/11038613-multicollinearity>. (accessed: 28.04.2023).
34. Variance Inflation Factor (VIF) / established by THE INVESTOPEDIA TEAM, 2023. URL: [https://www.investopedia.com/terms/v/variance-inflation-factor.asp#:~:text=A%20variance%20inflation%20factor%20\(VIF\)%20is%20a%20measure%20of%20the,adversely%20affect%20the%20regression%20results](https://www.investopedia.com/terms/v/variance-inflation-factor.asp#:~:text=A%20variance%20inflation%20factor%20(VIF)%20is%20a%20measure%20of%20the,adversely%20affect%20the%20regression%20results). (accessed: 28.04.2023).
35. Багатокритеріальний алгоритм крокової регресії / Настенко Є., Павлов В., Бойко Г., Носовець О.: Київ: КПІ, 2020, 1-6 с.

ДОДАТКИ

Додаток А

```
install.packages("readxl")  
install.packages("car")  
install.packages("gvlma")  
install.packages("mctest")  
install.packages("ggplot2")  
install.packages("dplyr")  
install.packages("lmtest")  
install.packages("sandwich")
```

```
library(dplyr)  
library(readxl)  
library(car)  
library(gvlma)  
library(mctest)  
library(ggplot2)  
library(lmtest)  
library(sandwich)
```

```
wheat <- read_excel("C:/Users/zador/OneDrive/Рабочий  
стол/пари/диплом/w.xlsx")
```

```
lm_W <- lm(data=wheat,Gross_profit~Region+Cultivated_area+Gross_out  
+Crop_cost+Sales+Gross_revenue+Price)
```

```
lm_W  
summary(lm_W)
```

```
wheat_1 <- wheat[wheat$Region == '1', ]
```

```
lm_W_1 <- lm(data=wheat_1,Gross_profit~Cultivated_area+Gross_out  
+Crop_cost+Sales+Gross_revenue+Price)
```

```
summary(lm_W_1)
```

```
wheat_2 <- wheat[wheat$Region == '2', ]
```

```
lm_W_2 <- lm(data=wheat_2,Gross_profit~Cultivated_area+Gross_out  
+Crop_cost+Sales+Gross_revenue+Price)
```

```
summary(lm_W_2)
```

```
wheat_3 <- wheat[wheat$Region == '3', ]
```

```
lm_W_3 <- lm(data=wheat_3,Gross_profit~Cultivated_area+Gross_out  
+Crop_cost+Sales+Gross_revenue+Price)
```

```
summary(lm_W_3)
```

```
corn <- read_excel("C:/Users/zador/OneDrive/Рабочий  
стол/пари/диплом/c.xlsx")
```

```
lm_C <- lm(data=corn,Gross_profit~Region+Cultivated_area+Gross_out  
+Crop_cost+Sales+Gross_revenue+Price)
```

```
lm_C
```

```
summary(lm_C)
```

```
sunflower <- read_excel("C:/Users/zador/OneDrive/Рабочий  
стол/пари/диплом/s.xlsx")
```

```
lm_S <- lm(data=sunflower,Gross_profit~Region+Cultivated_area+Gross_out  
+Crop_cost+Sales+Gross_revenue+Price)
```

```
lm_S
```

```
summary(lm_S)
```

```
error_W <- resid(lm_W)
```

```
hist (error_W)
error_C <- resid(lm_C)
hist (error_C)
error_S <- resid(lm_S)
hist (error_S)
```

```
par(mfrow=(c(2,2)))
plot(lm_W)
plot(lm_C)
plot(lm_S)
```

```
gvmodel_W <- gvlma(lm_W)
summary(gvmodel_W)
gvmodel_C <- gvlma(lm_C)
summary(gvmodel_C)
gvmodel_S <- gvlma(lm_S)
summary(gvmodel_S)
```

```
oneway.test(wheat$Gross_profit~wheat$Region)
oneway.test(corn$Gross_profit~corn$Region)
oneway.test(sunflower$Gross_profit~sunflower$Region)
```

```
T_wheat <- aov(wheat$Gross_profit~as.factor(wheat$Region))
TukeyHSD(T_wheat)
plot(TukeyHSD(T_wheat))
```

```
T_corn <- aov(corn$Gross_profit~as.factor(corn$Region))
TukeyHSD(T_corn)
plot(TukeyHSD(T_corn))
```

```
T_sunflower <- aov(sunflower$Gross_profit~as.factor(sunflower$Region))
TukeyHSD(T_sunflower)
plot(TukeyHSD(T_sunflower))
```

```
vif(lm_W)
vif(lm_C)
vif(lm_S)
```

```
imcdiag(lm_W)
imcdiag(lm_C)
imcdiag(lm_S)
```

```
imcdiag(lm_W_1)
```

```
corn_1 <- corn[corn$Region == '1', ]
lm_C_1 <- lm(data=corn_1,Gross_profit~Cultivated_area+Gross_out+
Crop_cost+Sales+Gross_revenue+Price)
```

```
corn_2 <- corn[corn$Region == '2', ]
lm_C_2 <- lm(data=corn_2,Gross_profit~Cultivated_area+Gross_out+
Crop_cost+Sales+Gross_revenue+Price)
```

```
corn_3 <- corn[corn$Region == '3', ]
lm_C_3 <- lm(data=corn_3,Gross_profit~Cultivated_area+Gross_out+
Crop_cost+Sales+Gross_revenue+Price)
```

```
# Тест Бройша-Пагана
bptest(lm_C_1, studentize = FALSE)
bptest(lm_C_2, studentize = FALSE)
bptest(lm_C_3, studentize = FALSE)
```

```
# Отримання робастних стандартних помилок
```

```
lm_C_robust <- coeftest(lm_C_2, vcov = vcovHC(lm_C_2, type = "HC1"))
```

```
# Виведення результатів аналізу
```

```
summary(lm_C_robust)
```

```
lm_C_robust <- lm(data=corn_2, Gross_profit ~ Cultivated_area + Gross_out +  
Crop_cost + Sales + Price)
```

```
lmtest::bptest(lm_C_robust)
```

```
summary(lm_C_robust)
```

Датасет пшениці

Region	Cultivated_area	Gross_out	Crop_cost	Sales	Gross_revenue	Price	Crop_yield	Investments_per_1_ha	Gross_profit
1	793.20	1292.18	11.8166000	2967.60	24.096985	8120.02	1.630	14.89738	-1.32406661
1	840.00	3034.80	17.5103000	3157.90	19.852033	6286.47	3.610	20.84560	1.56786896
1	865.00	1539.40	10.6034000	1042.20	6.868281	6590.18	1.780	12.25827	-0.45848370
1	577.00	1270.80	7.4665000	549.30	2.775934	5053.58	2.200	12.94021	-1.04440474
1	257.00	1329.50	4.6691000	1039.35	12.371500	11903.11	5.170	18.16770	11.15608810
1	230.00	984.00	4.1240000	1348.00	10.036500	7445.47	4.280	17.93043	3.20234718
1	450.00	1561.00	7.5807000	1042.70	7.774475	7456.10	3.470	16.84600	4.05827210
1	977.00	2755.00	9.6425000	2395.00	15.802333	6598.05	2.820	9.86950	8.53513149
1	1450.00	2960.30	13.6973000	3065.80	9.965803	3250.64	2.040	9.44641	-4.07443921
1	660.00	1243.50	4.5370000	824.32	4.113700	4990.42	1.880	6.87424	1.66858272
1	100.00	200.00	1.7883665	200.00	0.923784	4618.92	2.000	17.88366	-0.86458247
1	55.51	257.10	1.1470000	219.90	0.984905	4478.88	4.630	20.66294	0.00451922
1	768.00	2434.60	19.3590000	2285.00	10.554187	4618.90	3.170	25.20703	-8.11382606
1	122.00	431.06	1.4397000	419.40	2.040549	4865.40	3.530	11.80082	0.65757932
1	16.00	64.00	0.2889453	148.30	0.919578	6200.80	4.000	18.05908	0.10790560
1	1700.00	6639.00	31.7394000	6805.90	39.276168	5770.90	3.910	18.67024	6.57360510
1	200.00	541.40	2.3379000	225.50	1.296000	5747.23	2.710	11.68950	0.77364945
1	305.00	1794.00	7.0069130	198.80	1.601216	8054.41	5.880	22.97349	7.44269215
1	657.00	1450.00	7.6601000	1177.30	6.543916	5558.41	2.210	11.65921	0.39959439
1	170.00	595.00	3.7844370	393.30	2.182000	5547.93	3.500	22.26139	-0.48342001
1	2034.50	7682.10	27.9990000	4970.80	31.993739	6436.34	3.780	13.76210	21.44557880
2	1150.00	4884.80	28.2286000	5688.70	26.844315	4718.88	4.250	24.54661	-5.17779579
2	520.00	2547.00	10.6166000	1509.40	7.776700	5152.18	4.900	20.41654	2.50600163
2	815.00	2377.20	15.6802000	1542.30	9.133077	5921.73	2.920	19.23951	-1.60307451
2	425.07	2311.40	11.4819000	722.00	3.465936	4800.47	5.440	27.01179	-0.38610433
2	410.00	1204.80	6.6670676	1068.50	3.374304	3157.98	2.940	16.26114	-2.86233063
2	315.00	1152.80	7.8335000	759.90	4.036474	5311.85	3.660	24.86825	-1.71000056
2	2202.00	8497.00	45.4530000	8497.00	39.502553	4649.00	3.860	20.64169	-5.95044700
2	100.00	563.00	2.0831000	469.46	2.197015	4679.88	5.630	20.83100	0.55167118
2	30.43	170.00	1.5069639	0.00	0.000005	5000.00	5.590	49.52231	-0.65696389
2	117.00	352.20	1.7869000	112.70	0.406958	3610.98	3.010	15.27265	-0.51511111
2	210.50	727.00	4.7855220	698.00	3.123327	4474.68	3.450	22.73407	-1.53242927
2	37.00	132.20	0.8302000	105.50	0.535000	5071.09	3.570	22.43784	-0.15980190
2	27.00	84.50	0.3999241	222.80	1.276000	5727.11	3.130	14.81200	0.08401670
2	400.00	2430.00	8.9485000	1945.50	13.164783	6766.79	6.080	22.37125	7.49479103
2	353.01	1815.79	9.0077229	1047.82	4.855438	4633.85	5.140	25.51691	-0.59362913

Датасет кукурудзи

Region	Cultivated_area	Gross_out	Crop_cost	Sales	Gross_revenue	Price	Crop_yield	Investments_per_1_ha	Gross_profit
1	253.7726	1719.808	7.2730876	1681.968	10.4042513	6185.762	6.7769638	28.6598650	3.36523159
1	272.0000	2000.000	9.2000000	1785.230	15.8000000	8850.400	7.3529412	33.8235294	8.50080046
1	231.0000	1523.000	6.3100000	1325.360	9.6000000	7243.315	6.5930736	27.3160173	4.72156878
1	175.0000	1895.000	7.2900000	2000.320	10.0000000	4999.200	10.8285714	41.6571429	2.18348424
1	401.0000	2300.000	9.8100000	1126.600	7.8000000	6923.487	5.7356608	24.4638404	6.11401917
1	253.7726	1987.000	15.3000000	1681.968	11.3000000	6718.322	7.8298454	60.2902036	-1.95069512
1	301.0000	1881.250	7.9000460	2916.630	23.3330400	8000.000	6.2500000	26.2460000	7.14995400
1	253.7726	1719.808	10.0300000	1719.810	10.9000000	6337.909	6.7769638	39.5235779	0.86998433
1	188.0000	1200.000	5.3100000	528.360	3.6000000	6813.536	6.3829787	28.2446809	2.86624347
1	200.0000	1500.000	7.8100000	987.960	8.9000000	9008.462	7.5000000	39.0500000	5.70269282
1	120.0000	103.500	2.3869470	310.000	1.9496942	6289.336	0.8625000	19.8912250	-1.73600071
1	217.1000	1367.131	6.0624000	2945.600	18.9611020	6437.093	6.2972409	27.9244588	2.73794979
1	111.0000	988.000	11.2300000	1230.360	10.2000000	8290.257	8.9009009	101.1711712	-3.03922657
1	60.0000	327.320	1.6143000	126.900	0.7346876	5789.500	5.4553333	26.9050000	0.28071914
1	707.0000	5416.000	27.0638000	6119.300	40.4310290	6607.133	7.6605375	38.2797737	8.72043236
1	216.2600	1081.100	5.4662971	1351.500	7.3231570	5418.540	4.9990752	25.2765057	0.39168662
1	660.0000	1052.000	9.8964000	1052.000	4.4478560	4228.000	1.5939394	14.9945455	-5.44854400
1	368.0000	2364.000	16.9800000	2145.690	16.9000000	7876.254	6.4239130	46.1413043	1.63946507
1	253.7726	1719.808	9.8700000	1500.000	9.8700000	6580.000	6.7769638	38.8930921	1.44633354
1	198.0000	891.000	4.3051442	662.500	3.9790000	6006.038	4.5000000	21.7431524	1.04623544
1	183.0000	1705.500	3.9438000	112.800	0.7776740	6894.273	9.3196721	21.5508197	7.81438269
1	145.0000	1425.000	6.3600000	1269.980	7.9600000	6267.815	9.8275862	43.8620690	2.57163672
2	643.5600	5703.300	9.1775000	1681.968	10.4042513	6185.762	8.8621108	14.2605196	26.10175368
2	65.0000	135.000	1.5252000	1681.968	10.4042513	6185.762	2.0769231	23.4646154	-0.69012219
2	251.0000	1396.000	6.5221000	1356.250	8.1729700	6026.153	5.5617530	25.9844622	1.89040958
2	798.0000	3290.000	22.2017800	11806.000	45.2564900	3833.347	4.1228070	27.8217794	-9.59006967
2	179.0000	1789.000	8.5900000	1452.030	7.7400000	5330.468	9.9944134	47.9888268	0.94620793
2	437.0000	1394.100	10.6689000	869.700	5.0760430	5836.545	3.1901602	24.4139588	-2.53217291
2	681.0000	5363.000	17.1564000	5363.000	40.4053783	7534.100	7.8751836	25.1929515	23.24897830
2	100.0000	300.000	2.6157533	168.150	0.7654704	4552.307	3.0000000	26.1575330	-1.25006120
2	45.7400	450.260	2.1272965	90.260	0.2414978	2675.579	9.8439003	46.5084502	-0.92259011
2	887.7300	3338.740	17.8608140	3689.200	20.5134810	5560.414	3.7609859	20.1196467	0.70396415
2	6.0000	19.400	0.1568250	791.000	4.0185070	5080.287	3.2333333	26.1375000	-0.05826743
2	90.4900	436.890	2.9541000	909.700	5.2099000	5727.053	4.8280473	32.6455962	-0.45200787
2	67.1400	269.400	0.9968000	674.300	3.5270000	5230.610	4.0125112	14.8465892	0.41232620
2	1199.0000	6000.000	26.6956000	3889.600	25.1996510	6478.726	5.0041701	22.2648874	12.17675346

Датасет соняшнику

Region	Cultivated_area	Gross_out	Crop_cost	Sales	Gross_revenue	Price	Crop_yield	Investments_per_1_ha	Gross_profit
1	238.8300	159.3300	0.355100	478.1000	8.387298	17542.978	0.6671272	1.486832	2.44002276
1	637.0000	1005.0000	13.468000	521.9000	7.028239	13466.639	1.5777080	21.142857	0.06597240
1	532.0000	965.8000	11.258000	449.7000	6.704851	14909.609	1.8154135	21.161654	3.14170001
1	492.7000	195.6000	2.759100	335.3000	4.655160	13883.567	0.3969961	5.599959	-0.04347430
1	373.1593	809.7495	8.054361	765.7418	10.741389	14027.429	2.1699836	21.584244	3.30434311
1	375.8500	985.3600	11.000000	800.0000	14.960000	18700.000	2.6216842	29.266995	7.42623200
1	304.8700	914.6100	9.650050	677.0100	9.860448	14564.700	3.0000000	31.653000	3.67097016
1	566.0000	1100.0000	8.800000	1008.3000	14.521918	14402.378	1.9434629	15.547703	7.04261609
1	1395.7300	2295.1000	22.462400	2204.9000	32.932000	14935.825	1.6443725	16.093657	11.81681139
1	419.6000	325.5000	2.325000	564.6900	5.476670	9698.543	0.7757388	5.540991	0.83187560
1	113.9200	85.0000	3.829119	257.7000	4.237940	16445.246	0.7461376	33.612351	-2.43127306
1	337.2400	724.5600	5.014600	713.3000	12.151076	17035.015	2.1484996	14.869529	7.32829027
1	712.0000	976.0000	10.525000	1170.1000	14.894788	12729.500	1.3707865	14.782303	1.89899200
1	65.0000	154.7000	1.071100	154.7000	2.164253	13990.000	2.3800000	16.478462	1.09315300
1	346.0000	479.0000	7.174800	198.0000	4.292965	21681.641	1.3843931	20.736416	3.21070624
1	216.2600	599.0000	5.315961	836.1000	10.715547	12816.107	2.7698141	24.581342	2.36088705
1	1385.8000	1723.3000	24.485380	1723.3000	7.286112	4228.000	1.2435416	17.668769	-17.19926800
1	201.0000	432.0000	6.980000	264.0000	6.000000	22727.273	2.1492537	34.726368	2.83818182
1	677.0000	459.7000	8.182300	459.8000	6.299746	13701.057	0.6790251	12.086115	-1.88392411
1	190.0000	386.1400	4.314106	386.1000	5.472252	14173.147	2.0323158	22.705824	1.15871246
1	1367.9000	5141.0000	18.335000	5261.5000	91.232470	17339.631	3.7583157	13.403758	70.80804443
1	12.3500	25.4000	0.202900	25.4000	0.343800	13535.433	2.0566802	16.429150	0.14090000
2	1145.3000	3070.3000	33.809400	1694.9000	16.926208	9986.553	2.6807823	29.520126	-3.14768756
2	458.0000	1232.0000	10.776000	1384.9000	16.547419	11948.458	2.6899563	23.528384	3.94449983
2	382.0000	678.2000	11.546600	917.3000	15.840531	17268.648	1.7753927	30.226702	0.16499721
2	556.3800	1768.0000	14.252930	1160.8000	14.763651	12718.514	3.1776843	25.617258	8.23340267
2	645.0400	1152.9200	12.934744	1510.1000	19.207267	12719.202	1.7873620	20.052623	1.72947823
2	333.0000	201.7000	19.137800	153.0000	16.793636	109762.327	0.6057057	57.470871	3.00126132
2	2117.8000	4873.1000	55.084400	4873.1000	73.273881	15036.400	2.3010199	26.010199	18.18948084
2	100.0000	150.0000	1.950000	185.5000	2.332318	12573.143	1.5000000	19.500000	-0.06402853
2	67.3800	164.1800	1.844589	132.5600	2.016055	15208.621	2.4366281	27.375917	0.65236214
2	500.0000	1000.0000	15.000000	989.0000	16.500000	16683.519	2.0000000	30.000000	1.68351871
2	221.5100	672.0000	5.152005	470.0000	4.601472	9790.366	3.0337231	23.258566	1.42712092
2	100.0000	295.0000	2.537500	75.0000	1.002200	13362.667	2.9500000	25.375000	1.40448667
2	100.0000	150.2000	1.718600	84.6000	1.092000	12907.801	1.5020000	17.186005	0.22015130
2	1259.0000	1800.0000	24.261100	1205.3000	12.775209	10599.194	1.4297061	19.270135	-5.18255010