

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний університет «Запорізька політехніка»

Факультет комп'ютерних наук і технологій

(повне найменування факультету)

Кафедра системного аналізу та обчислювальної математики

(повне найменування кафедри)

Пояснювальна записка

до дипломного проєкту (роботи)

магістр

(ступінь вищої освіти)

на тему: Аналіз гендерних відмінностей результатів ЗНО/НМТ

Виконав(ла): студент(ка) 2 курсу, групи КНТ-812м
Спеціальності 124 Системний аналіз

(код і найменування спеціальності)

Освітня програма (спеціалізація)

Інтелектуальні технології та прийняття рішень в складних системах

МАНЬКОВ В.В.

(ПРІЗВИЩЕ та ініціали)

Керівник БАХРУШИН В.Є.

(ПРІЗВИЩЕ та ініціали)

Рецензент ІГНАХІНА М.О.

(ПРІЗВИЩЕ та ініціали)

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний університет «Запорізька політехніка»

Факультет _____ комп'ютерних наук і технологій
Кафедра _____ системного аналізу та обчислювальної математики
Ступінь вищої освіти _____ магістр
Спеціальність _____ 124 – Системний аналіз
Освітня програма (спеціалізація) _____ Інтелектуальні технології та прийняття рішень в складних системах

ЗАТВЕРДЖУЮ

В.о. завідувача кафедри системного аналізу
та обчислювальної математики
канд.фіз-мат. наук
доц. Терещенко Еліна Валентинівна
« _____ » _____ 2023 року

З А В Д А Н Н Я
НА ДИПЛОМНИЙ ПРОЄКТ (РОБОТУ) СТУДЕНТА(КИ)

МАНЬКОВА Владислава Володимировича _____ i

(прізвище, ім'я, по батькові)

1. Тема проєкту (роботи) Аналіз гендерних відмінностей результатів ЗНО/НМТ,
керівник проєкту (роботи) БАХРУШИН Володимир Євгенович, д.ф.-м.н., проф.

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом закладу вищої освіти від «06» грудня 2023 року № 492

2. Строк подання студентом проєкту (роботи) 21 грудня 2023 року.

3. Вихідні дані до проєкту (роботи) Матеріали українського центру оцінювання якості освіти.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити): Проблема гендерних відмінностей результатів навчання вступників, методика виконання роботи, аналіз результатів ЗНО/НМТ.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, кількість слайдів, плакатів)

6. Консультанти розділів проєкту (роботи) Бахрушин В.Є. д.ф.-м.н., професор

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	прийняв виконане завдання
1	Бахрушин В.Є. д.ф.-м.н., професор		
2	Бахрушин В.Є. д.ф.-м.н., професор		
3	Бахрушин В.Є. д.ф.-м.н., професор		
Нормоконтроль	Широкорад Д.В., к.ф.-м.н., доцент		

7. Дата видачі завдання «04» вересня 2023 року.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного проєкту (роботи)	Строк виконання етапів проєкту (роботи)	Примітка
1	Аналіз та огляд літератури за темою дослідження	04.09-18.09	
2	Підготовка даних до дослідження	19.09-03.10	
3	Відпрацювання методики виконання роботи	04.10-25.10	
4	Виконання розрахунків та аналіз результатів	26.10-23.11	
5	Написання тексту пояснювальної записки	24.11-01.12	

Студент

(підпис)

Владислав МАНЬКОВ

Керівник проєкту (роботи)

(підпис)

Володимир БАХРУШИН

РЕФЕРАТ

Дипломна робота: 130 с., 58 табл, 103 рисунки, 1 додаток.

Об'єктом дослідження є результати Зовнішнього незалежного оцінювання (ЗНО) та Національного мультипредметного тесту (НМТ) за 2021-2023 роки.

Мета дослідження: визначення наявності або відсутності істотних відмінностей результатів ЗНО та НМТ між чоловіками та жінками.

Предмет дослідження: вплив статі на результати складання ЗНО/НМТ.

Методи дослідження: статистичний аналіз результатів ЗНО/НМТ.

Ключові слова: ЗОВНІШНЄ НЕЗАЛЕЖНЕ ОЦІНЮВАННЯ, НАЦІОНАЛЬНИЙ МУЛЬТИПРЕДМЕТНИЙ ТЕСТ, СТАТИСТИЧНИЙ АНАЛІЗ, ПАРАМЕТРИЧНІ КРИТЕРІЇ, НЕПАРАМЕТРИЧНІ КРИТЕРІЇ, ГЕНДЕРНІ ВІДМІННОСТІ, РЕЗУЛЬТАТИ НАВЧАННЯ.

ЗМІСТ

ВСТУП.....	6
1. ПРОБЛЕМА ГЕНДЕРНИХ ВІДМІННОСТЕЙ РЕЗУЛЬТАТІВ НАВЧАННЯ	7
2. МЕТОДИКА ВИКОНАННЯ РОБОТИ	12
3. АНАЛІЗ РЕЗУЛЬТАТІВ ЗНО/НМТ	18
3.1 Аналіз результатів тестувань з української мови	18
3.2 Аналіз результатів тестувань з математики	29
3.3 Аналіз результатів тестувань з історії України.....	37
3.4 Аналіз результатів тестувань з фізики	45
3.5 Аналіз результатів тестувань з хімії.....	51
3.6 Аналіз результатів тестувань з біології	57
3.7 Аналіз результатів тестувань з географії.....	63
3.8 Аналіз результатів тестувань з англійської мови	67
3.9 Аналіз результатів тестувань з французької мови.....	73
3.10 Аналіз результатів тестувань з німецької мови	79
3.11 Аналіз результатів тестувань з іспанської мови	85
3.12 Аналіз різниці середніх балів між чоловіками та жінками.....	91
ВИСНОВКИ.....	93
ПЕРЕЛІК ПОСИЛАНЬ	95
ДОДАТОК А. Код програми.....	97

ВСТУП

Через воєнні дії в Україні, вже другий рік поспіль замість традиційного зовнішнього незалежного оцінювання (ЗНО) проводився національний мультипредметний тест (НМТ), метою якого є забезпечення реалізації конституційних прав громадян на рівний доступ до фахової передвищої та вищої освіти [1].

Якщо у 2022 році НМТ передбачав оцінювання результатів навчання з трьох предметів – української мови, математики та історії України, то у 2023 році перелік предметів був розширений. Тестування проводилося з української мови, математики, біології, іноземної мови (на вибір: англійська, німецька, французька, іспанська), історії України, фізики та хімії. Кожен учасник виконував три предметні субтести, а саме субтести з української мови та математики як обов’язкові для всіх учасників та один субтест із переліку предметів за вибором [1].

Після завершення всіх завдань НМТ учасник отримує кількість тестових балів за кожний субтест. Так, аби отримати 100 балів (за шкалою 100-200) достатньо набрати 4 бали з математики, іноземної мови, фізики та хімії; 5 балів – з української мови та біології; 6 балів – з історії України. Учасник НМТ, який не набрав мінімально необхідної кількості тестових балів, отримував результат 0 [1].

Для досягнення мети дослідження в цій дипломній роботі були опрацьовані методики дослідження, розраховано основні показники описової статистики, перевірена однорідність вибірок результатів чоловіків та жінок за предметами та проаналізовано отримані результати про наявність або відсутність істотних відмінностей між чоловіками та жінками.

1 ПРОБЛЕМА ГЕНДЕРНИХ ВІДМІННОСТЕЙ РЕЗУЛЬТАТІВ НАВЧАННЯ

Згідно даних ООН, досягнення гендерної рівності входить до однієї з цілей у сфері сталого розвитку країн до 2030 року [2]. Ключову роль у вирішенні цього завдання грає етап здобуття освіти, на якому необхідно надати однаковий доступ до якісного навчання представникам різної статі.

Незважаючи на загальні позитивні тенденції, гендерна нерівність у школі та вишах продовжує існувати і особливо проявляється в окремих країнах.

За даними ЮНЕСКО, гендерна нерівність в освіті у всьому світі в основному визначається «бідністю, географічною ізоляцією, статусом меншості, інвалідністю, ранніми шлюбами та вагітністю, а також гендерним насильством» [3]. Тільки в Північній Америці, Латинській Америці та країнах Карибського басейну дівчатка відвідують школу так само або частіше, ніж хлопчики. Різниця між рівнем зарахування дівчат і хлопчиків є найбільшою на Близькому Сході та в Африці на південь від Сахари.

Таблиця 1.1 – Чистий коефіцієнт зарахування в початкову школу на основі даних Світового банку [4].

Регіон	Чоловіки (%)	Жінки (%)
Арабські країни	85,86	82,81
Східна Азія та Тихий океан	95,61	95,53
Європа та Центральна Азія	96,00	95,86
Латинська Америка та Карибський басейн	91,73	91,76
Близький схід та Північна Африка	95,23	91,80
Північна Америка	93,38	93,87

Продовження таблиці 1.1

Регіон	Чоловіки (%)	Жінки (%)
Південна Азія	89,72	88,55
Африка на південь від Сахари	80,02	75,83

В 1970 році різниця в роках отриманого навчання між чоловіками та жінками складала майже 2 роки у всьому світі, а до 2018 року ця цифра скоротилась до 0,3 років. В ряді країн дана тенденція призвела до того, що жінки перевершили чоловіків за роками навчання. Довше чоловіків жінки стали навчатися в країнах з найбільшим рівнем доходу – в Європі та Центральній Азії, але протилежна ситуація досі спостерігається в частині Південної Азії та в Африці на південь від Сахари. Так до країн з найбільшою різницею в роках отриманого навчання входять Екваторіальна Гвінея, Республіка Того, Камерун.

Таблиця 1.2 – Топ-10 країн з найбільшою різницею в роках отриманого навчання між чоловіками та жінками [5].

Країна	Чоловіки	Жінки	Різниця
Екваторіальна Гвінея	7,3	4	3,3
Республіка Того	6,5	3,3	3,2
Камерун	7,6	4,7	2,9
Центрально-Африканська республіка	5,40	2,62	2,78
Афганістан	3,91	1,17	2,74
Ліберія	6,1	3,5	2,6
Непал	5,6	3,24	2,36
Індія	7,62	5,52	2,1

Продовження таблиці 1.2

Країна	Чоловіки	Жінки	Різниця
Демократична республіка Конго	7,45	5,5	1,95
Пакистан	5,03	3,89	1,14

В цих країнах, дівчатка в класі можуть частіше стикатися зі стереотипним мисленням вчителів, піддаватися насиллю та ранньому виключенню з процесу навчання через вагітність або заміжжя. Так, регулярно відвідуючи школу, дівчатка не завжди мають можливість повноцінно навчатися та отримувати високі оцінки.

У переважній більшості країнах за останні кілька десятиліть доступ жінок до освіти збільшився. За даними ОЕСР в 43 розвинених країнах, 15-річні дівчатка випереджають хлопців у грамотності і є більш впевненими. В більшості західних країн дівчатка навчаються в школі краще, ніж хлопчики. Через нижчі оцінки у хлопчиків менше шансів отримати додаткову освіту, що в кінцевому підсумку може обмежити їхні шанси на успіх на ринку праці. Згідно з дослідженнями Бйорна Тірефорса, Еріка Хегліна та Магнуса Йоханнесона, було доведено, що дівчатка успішніші в школі, ніж хлопчики [8]. Дуже важливо продовжувати досліджувати причини цієї гендерної різниці. Найбільш імовірним поясненням є те, що дівчатка докладають більше зусиль у школі, оскільки ця різниця не є результатом дискримінації та напевно чи залежить від вроджених відмінностей у здібностях. Розробка заходів для усунення гендерного розриву в освіті вимагає дослідження. Школа повинна враховувати використовувати різноманітні стилі навчання для хлопців та дівчаток. Хлопчиків можуть вважати неспокійними, ледачими та менш ініціативними, якщо навчання не відповідає їхнім потребам. Враховуючи уявлення про те, що в класі повинні бути створені рівні умови, важливо досліджувати гендерну нерівність і нижчу успішність хлопців без упереджень. Замість того, щоб зосереджуватися на поведінці та успішності своїх учнів, вчителі також повинні бути готові до критичного аналізу та проблематизації власного викладання.

Гендерна дискримінація в освіті також існує через різне ставлення до учнів з боку вчителів-чоловіків чи жінок. В Ньюфаундленді, в дослідженні Джима Даффі, було виявлено, що вчителі можуть мати вищі очікування від хлопчиків з математики та природничих наук, а також від дівчаток є вищі очікування щодо мови. Було виявлено, що вчителі також мають тенденцію хвалити учнів, які відповідають гендерним очікуванням нормам. Учні частіше хвалили вчителі математики, ніж вчителі літератури. Також було виявлено, що на уроках літератури та математики критика адресована учням чоловічої статі значно частіша, ніж ученицям жіночої статі незалежно від статі вчителя [9].

Існує мало доказів того, що хлопчики почуваються гірше, ніж раніше. Скоріше, високі досягнення дівчаток в останні роки показують, що від їх однолітків-хлопчиків можна очікувати набагато більшого.

Міжнародне дослідження TIMSS-2019, спрямоване на оцінку знань з математики та природознавства, показує, що з математики в 27 країнах хлопчики-четверокласники випереджали дівчаток, а дівчатка – лише в 4 країнах. До 8-го класу, тим не менш, ці відмінності повністю вирівнюються: дівчатка випереджають хлопчиків у 7 країнах, а хлопчики в 6 країнах. Відомо, що саме за математичними дисциплінами (поряд з технічними) представники чоловічої статі демонструють більш високі досягнення.

За природничими предметами загальним трендом в світі є перевага дівчаток як в четвертому, так й у восьмому класі [10].

Також саму ситуацію ми можемо спостерігати, якщо звернемося до іншої предметної області – читання. Ще одне міжнародне дослідження – PIRLS – дозволяє оцінити навички четверокласників у читацькій грамотності. У переважній більшості країн-учасниць дослідження дівчатка випередили хлопців в навичках читання [11].

У PISA, яка вимірює навички застосування знань на практиці у всіх трьох предметних областях серед 15-річних підлітків, моделі гендерної нерівності приблизно такі самі [12].

Дані демонструють, що проблема гендерної нерівності виявляється складнішою, ніж прийнято вважати.

Крім того, читацькі стратегії хлопчиків відрізняються меншою уважністю та скрупульозністю порівняно з дівчатками. Для підвищення їх результатів та залучення до освітнього процесу пропонується працювати над дисципліною у класі, розвитком відносин із учителем та поширенням культури тестувань. З іншого боку, дівчатка відстають з математики і рідше обирають кар'єру в STEM-областях (пов'язаних із точними науками та технікою), що пояснюється насамперед відсутністю у них інтересу до подібних предметів та впевненості у своїх силах. Це вимагає цілеспрямованої роботи з установками та стереотипами у процесі навчання [13].

У вищій освіті ситуація з гендерною рівністю виглядає досить обнадійливо. Повсюдно спостерігається загальне збільшення чисельності вступників до вищих навчальних закладів. При цьому кількість жінок у вищій освіті зростає швидше порівняно з чоловіками. У розвинених країнах це вже призвело до того, що жінки не просто наздогнали, а й випередили чоловіків за вступом до університетів. Проте зберігаються деякі тривожні відмінності. У всьому світі жінки набагато менше навчаються на STEM-спеціальностях у порівнянні з чоловіками. У більшості країн їх чисельність на відповідних факультетах не сягає 25% [13].

Також проблема гендерної нерівності торкається ринку праці. На робочих місцях по всьому світу чоловіки, як і раніше, користуються величезними перевагами в порівнянні з жінками, у тому числі вищою оплатою праці та більшою ймовірністю просування по службі [14]. Дійсно, жінки рідше потрапляють до високотехнологічних галузей економіки, проте це не пояснює повною мірою відмінності в оплаті праці. Відсутність прогресу у встановленні гендерної рівності на ринку праці стає пріоритетним питанням для досліджень.

2 МЕТОДИКА ВИКОНАННЯ РОБОТИ

Для проведення дослідження використовувались програмні інструменти для статистичної обробки даних, зокрема Microsoft Excel та мови програмування R та Python.

Для аналізу даних з сайту Українського центру оцінювання якості освіти були завантажені таблиця результатів Зовнішнього незалежного оцінювання за 2021 рік та таблиці результатів Національного мультипредметного тесту за 2022-2023 рік[1].

	A	B	R	S	T	U	V	W	X	Y	Z
1	OUTID	Birth	Block1	Block1f	Block1f-T	Block2	Block2f	Block2f	Block3	Block3f	Block3f
2	d60381f3-	2005	Українськ	142	14	Історія Ук	147	15	Математи	128	5
3	eb25a9fc-	2005	Українськ	137	11	Історія Ук	149	17	Математи	136	8
4	1cb161bd-	2005	Українськ	169	25	Історія Ук	158	24	Математи	185	28
5	0311b8d8-	2005	Українськ	154	20	Історія Ук	146	14	Математи	148	14
6	a8b35a53-	2005	Українськ	166	24	Історія Ук	144	13	Математи	144	12
7	c6655f55-	2005	Українськ	184	31	Історія Ук	150	18	Математи	150	16
8	8c37ba54-	2005	Українськ	145	16	Історія Ук	142	12	Математи	134	7
9	e07ca3df-	2005	Українськ	180	29	Історія Ук	153	21	Математи	128	5
10	1be9a6e8-	2005	Українськ	140	13	Історія Ук	152	20	Математи	128	5
12	4a67e525-	2005	Українськ	190	33	Історія Ук	200	35	Математи	128	5
13	26f85642-	2004	Українськ	149	18	Історія Ук	156	23	Математи	128	5
14	0bab3e81-	2004	Українськ	137	11	Історія Ук	146	14	Математи	142	11
15	efe57475-	2004	Українськ	142	14	Історія Ук	134	8	Математи	140	10
17	1671eccf-	2005	Українськ	140	13	Історія Ук	140	11	Математи	134	7
18	d322287f-	2005	Українськ	194	34	Історія Ук	200	35	Математи	192	29
19	e72fdf96-	2005	Українськ	154	20	Історія Ук	154	22	Математи	162	22

Рисунок 2.1 – Електронна таблиця результатів тестувань

Таблиця містить стовпці OutID (ID), Birth (Рік народження), SexTypeName (Стать), Regname (Регіон реєстрації/проживання учасника), AreaName (Район/Місто реєстрації учасника), TerName (Населений пункт реєстрації/проживання учасника), RegTypeName (Статус учасника), TerTypeName (Тип території), EOName (Заклад освіти учасника), EOTypeName (Тип закладу освіти), EORegName (Регіон, де розташований заклад освіти), EOAreaName (Район /місто, де розташований заклад освіти), EOTerName (Населений пункт, де розташований заклад освіти), EOParent (Орган, якому підпорядковується заклад освіти), Test (Назва тесту), TestDate (Дата проведення тесту), TestStatus (Результат складання тесту), Block1, Block2, Block3 (Назва блоків тесту), Block1Ball100,

Block2Ball100, Block3Ball100 (Оцінка за шкалою 100-200 з відповідного блоку), Block1Ball, Block2Ball, Block3Ball (Тестовий бал з відповідного блоку), PTRegName (Регіон, де розташований тимчасовий екзаменаційний центр), PTArenaName (Район/місто, де розташований тимчасовий екзаменаційний центр), PTTerName (Населений пункт, де розташований тимчасовий екзаменаційний центр).

Перш ніж переходити до аналізу результатів Національного мультипредметного тесту було розраховано загальні статистичні показники та підраховано кількість чоловіків та жінок, що взяли участь у тестуванні.

На основі даних Українського центру оцінювання освіти було створено таблиці з первинними (до переведення у шкалу 100-200) результатами НМТ з усіх предметів.

За допомогою спеціальних функцій MS Excel розраховано основні показники елементів описової статистики результатів НМТ з кожного предмету (окремо для чоловіків та жінок):

- середнє значення;
- медіана;
- стандартне відхилення;
- мода вибірки;
- мінімум;
- максимум.

Для знаходження показника середнього балу застосовувалась функція AVERAGE(число1:число2), для розрахунку медіани функція MEDIAN(число1:число2).

Наступною по важливості характеристикою вибірки є міра розсіювання елементів вибірки від середнього значення. Такою мірою є стандартне відхилення. Для визначення стандартного відхилення була застосована функція STDEV.

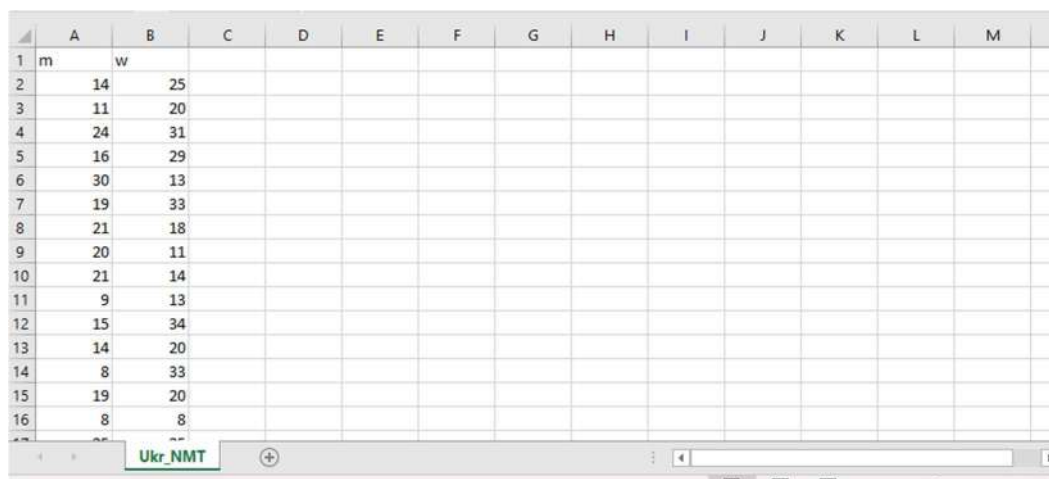
Також елементи описової статистики було розраховано за допомогою вбудованого в MS Excel пакету Аналіз даних.

	A	B	C	D	E	F	G
1	SexTypeNBlock1Ball	ID					
2	чоловіча	14	1				
3	чоловіча	11	2			Показники	Значення
4	чоловіча	24	3			Середнє значення (до переведення в шкалу 100-200)	17,50821
5	чоловіча	16	4			Медіана	16
6	чоловіча	30	5			Стандартне відхилення	7,286737
7	чоловіча	19	6			Мода	13
8	чоловіча	21	7			Кількість учасників	101784
9	чоловіча	20	8				
10	чоловіча	21	9				
11	чоловіча	9	10				
12	чоловіча	15	11			Середнє	17,50821
13	чоловіча	14	12			Стандартна помилка	0,02284
14	чоловіча	8	13			Медіана	16
15	чоловіча	19	14			Мода	13
16	чоловіча	8	15			Стандартне відхилення	7,286737
17	чоловіча	25	16			Дисперсія вибірки	53,09654
18	чоловіча	10	17			Екссес	-0,54961
19	чоловіча	12	18			Асиметричність	0,504704
20	чоловіча	9	19			Інтервал	35
21	чоловіча	16	20			Мінімум	0
22	чоловіча	11	21			Максимум	35
23	чоловіча	13	22			Сумма	1782056
24	чоловіча	9	23			Кількість	101784

Рисунок 2.2 – Застосування вбудованого пакету Аналіз даних

Далі для продовження аналізу вибірок та порівняння результатів було застосовано мови програмування R та Python. В Python був використаний ряд бібліотек, таких як pandas, numpy, matplotlib, seaborn та scipy.stats.

Для початку аналізу результатів тестувань були створені нові таблиці даних з балами чоловіків та жінок та збережено їх у форматі csv.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	m	w											
2	14	25											
3	11	20											
4	24	31											
5	16	29											
6	30	13											
7	19	33											
8	21	18											
9	20	11											
10	21	14											
11	9	13											
12	15	34											
13	14	20											
14	8	33											
15	19	20											
16	8	8											

Рисунок 2.3 – Сформована електронна таблиця з балами чоловіків та жінок

Виконаємо зчитування за допомогою наступних команд:

```
file_path = '/content/drive/MyDrive/Geography_2021.csv'
```

```
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=58548)
```

```
datm = np.array(dat.iloc[0:58548, 0])
```

```
datw = np.array(dat.iloc[0:54568, 1])
```

Перед застосуванням статистичних критеріїв для оцінки однорідності вибірок, проведемо їх порівняння за допомогою графічних методів. На початку проведемо побудову емпіричних функцій розподілу вибірок чоловіків та жінок.

Для цього використаємо такий набір команд:

```
erfm = np.linspace(1/len(datm), 1, len(datm))
```

```
erfw = np.linspace(1/len(datw), 1, len(datw))
```

```
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
```

```
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--")
```

```
plt.xlabel("Бал ЗНО")
```

```
plt.ylabel("ЕФР")
```

```
plt.legend()
```

```
plt.show()
```

Далі побудуємо діаграми типу “boxplot”:

```
plt.figure(figsize=(8, 6))
```

```
sns.boxplot(x=["Чоловіки"] * len(datm) + ["Жінки"] * len(datw),
```

```
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
```

```
plt.ylabel("Кількість балів")
```

```
plt.title("Географія")
```

```
plt.show()
```

Для більш докладного кількісного порівняння характеристик розподілів визначимо їх квантилі за допомогою наступних команд:

```
print("Квантилі для чоловіків:", np.percentile(datm, [10, 25, 75, 90]))
```

```
print("Квантилі для жінок:", np.percentile(datw, [10, 25, 75, 90]))
```

Для побудови Q-Q діаграм та гістограм розподілу показників використаємо мову програмування R з наступним набором команд:

```
par(mfrow=c(1,2), mar=c(4,4,2,2))
```

```

hist(datm, col= 3, main="Гістограма: Чоловіки", xlab="Кількість балів",
ylab="Frequency", breaks=8)
hist(datw, col= 4, main="Гістограма: Жінки", xlab="Кількість балів",
ylab="Frequency", breaks=8)
old.par <- par(mfrow=c(1,2))
qqnorm(datm, main = "Географія, чоловіки", col="#61d04f", col.main =
"brown",
      cex.main = 1.2)
qqline(datm, col = 2)
qqnorm(datw, main = "Географія, жінки", col = "#2297e6", col.main = "brown",
      cex.main = 1.2)
qqline(datw, col = 2)
par(old.par)

```

На основі побудованої Q-Q діаграми можна буде зробити висновки про те, чи спостерігається відхилення даних від нормального розподілу.

Для вибору відповідних критеріїв перевірки однорідності вибірок, важливим є попереднє визначення їх відповідності нормальному розподілу. Для цього використовувались статистичні критерії, такі як критерій Андерсона-Дарлінга та критерій Ліллієфорса:

```

# Тести Андерсона – Дарлінга
result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)

```

```

print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)
# Тест Ліллієфорса
p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Ліллієфорса для чоловіків:", p_value_lilliefors_m)
print("Тест Ліллієфорса для жінок:", p_value_lilliefors_w)

```

Якщо порівнювані вибірки мають нормальний розподіл, можна провести перевірку рівності їхніх середніх арифметичних за допомогою параметричного статистичного критерію Уелча, і перевірку рівності дисперсій за допомогою критерію Фішера. У випадку, якщо вибірки не відповідають нормальному розподілу даних необхідно застосовувати статистичні непараметричні критерії, такі як критерій Колмогорова-Смирнова, який перевіряє близькість емпіричних функцій розподілу вибірок, та критерій Уїлкоксона, який є чутливим до зсуву центрів розподілу:

```

# Тест Колмогорова-Смирнова
print("KS-тест:", ks_2samp(datm, datw))
# Тест Уїлкоксона
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона:", wilcoxon(datm[:common_length],
datw[:common_length], alternative='two-sided'))

```

3 АНАЛІЗ РЕЗУЛЬТАТІВ ЗНО/НМТ

3.1 Аналіз результатів тестувань з української мови

На основі завантажених даних з сайту Українського центру оцінювання якості освіти підрахуємо кількість учасників тестування з української мови за 2021-2023 роки. Для цього сформуємо таблицю з відповідними вибірками з первинними (до переведення в шкалу 100-200) результатами:

	2021			2022			2023		
	SexTypeName	UkrBlockBall	ID	SexTypeName	UkrBlockBall	ID	SexTypeName	UkrBlockBall	ID
3	чоловіча		5	1 чоловіча		14	1 жіноча		21
4	чоловіча		7	2 чоловіча		11	2 жіноча		28
5	жіноча		8	3 жіноча		25	3 жіноча		44
6	жіноча		11	4 жіноча		20	4 жіноча		24
7	жіноча		11	5 чоловіча		24	5 чоловіча		17
8	жіноча		8	6 жіноча		31	6 жіноча		37
9	жіноча		6	7 чоловіча		16	7 чоловіча		26
10	жіноча		4	8 жіноча		29	8 жіноча		28
11	жіноча		4	9 жіноча		13	9 жіноча		31
12	жіноча		10	10 жіноча		33	10 чоловіча		17
13	чоловіча		4	11 жіноча		18	11 чоловіча		44
14	чоловіча		6	12 жіноча		11	12 чоловіча		19
15	жіноча		8	13 жіноча		14	13 жіноча		15
16	жіноча		11	14 жіноча		13	14 чоловіча		31
17	чоловіча		4	15 жіноча		34	15 жіноча		19
18	чоловіча		2	16 жіноча		20	16 чоловіча		18
19	чоловіча		7	17 жіноча		33	17 чоловіча		14
20	жіноча		5	18 жіноча		20	18 чоловіча		8
21	жіноча		5	19 чоловіча		30	19 жіноча		30
22	жіноча		6	20 жіноча		8	20 жіноча		9
23	жіноча		5	21 жіноча		25	21 жіноча		17
24	жіноча		7	22 чоловіча		19	22 чоловіча		25
25	чоловіча		7	23 жіноча		10	23 чоловіча		26
26	чоловіча		3	24 жіноча		14	24 чоловіча		19
27	чоловіча		3	25 жіноча		17	25 жіноча		30
28	жіноча		5	26 чоловіча		21	26 жіноча		40
29	жіноча		11	27 чоловіча		20	27 чоловіча		24
30	жіноча		4	28 чоловіча		21	28 жіноча		23
31	чоловіча		8	29 чоловіча		9	29 чоловіча		11
32	чоловіча		12	30 жіноча		10	30 чоловіча		25
33	жіноча		4	31 жіноча		12	31 чоловіча		14
34	жіноча		4	32 чоловіча		15	32 чоловіча		30
35	чоловіча		3	33 жіноча		13	33 жіноча		41

Рисунок 3.1 – Електронна таблиця даних результатів тестувань за 2021-2023 роки

Кількість учасників будемо визначати за допомогою функції COUNT(C3:C222250); COUNT(F3:F214076); COUNT(I3:268130). Отримані дані внесемо до табл. 3.1.

Таблиця 3.1 – Кількість учасників тестування з української мови (за 3 роки)

2021 р.		2022 р.		2023 р.	
Кількість учасників	222248	Кількість учасників	214074	Кількість учасників	268128

Аналогічно підрахуємо кількість чоловіків та жінок, що взяли участь у тестуванні та занесемо результати обчислень в табл. 3.2.

Таблиця 3.2 – Процентне співвідношення чоловіків та жінок, які беруть участь у тестуванні

Стать	2021 р. (ЗНО)		2022 р. (НМТ)		2023 р. (НМТ)	
	кількість	% від загальної кількості учасників	кількість	% від загальної кількості учасників	кількість	% від загальної кількості учасників
Чоловіча	91668	41,25%	101784	47,55%	128405	47,89%
Жіноча	130580	58,75%	112290	52,45%	139723	52,11%

Виходячи з розрахункових даних, в 2022 році спостерігаються зміни гендерного складу учасників – зниження частки жінок (2021 р.: 58,75% жінок, 41,25% чоловіків; 2022 р.: 52,45% жінок, 47,55% чоловіків; 2023 р.: 52,11% жінок, 47,89% чоловіків).

На основі даних сформованої таблиці з первинними (до переведення у шкалу 100-200) результатами ЗНО та НМТ розрахуємо основні показники описової статистики, окремо для чоловіків та жінок. Для цього будемо застосовувати вбудований пакет Аналіз даних в MS Excel:

2021 р. (ЗНО) Українська мова			
Чоловіки		Жінки	
Середнє	49,774403	Середнє	62,899449
Стандартна помилка	0,0714718	Стандартна помилка	0,0652931
Медіана	46	Медіана	61
Мода	34	Мода	50
Стандартне відхилення	21,639334	Стандартне відхилення	23,594214
Дисперсія вибірки	468,26078	Дисперсія вибірки	556,68694
Ексцес	-0,263423	Ексцес	-0,943354
Асиметричність	0,6567351	Асиметричність	0,1601251
Інтервал	116	Інтервал	116
Мінімум	0	Мінімум	0
Максимум	116	Максимум	116
Сума	4562720	Сума	8213410
Кількість	91668	Кількість	130580

Рисунок 3.2 – Показники описової статистики за 2021 р.

2022 р. (НМТ) Українська мова			
Чоловіки		Жінки	
Середнє	17,508213	Середнє	20,498255
Стандартна помилка	0,0228399	Стандартна помилка	0,0227695
Медіана	16	Медіана	20
Мода	13	Мода	15
Стандартне відхилення	7,2867373	Стандартне відхилення	7,6300027
Дисперсія вибірки	53,09654	Дисперсія вибірки	58,216942
Ексцес	-0,549612	Ексцес	-0,979077
Асиметричність	0,504704	Асиметричність	0,1247003
Інтервал	35	Інтервал	35
Мінімум	0	Мінімум	0
Максимум	35	Максимум	35
Сума	1782056	Сума	2301749
Кількість	101784	Кількість	112290

Рисунок 3.3 – Показники описової статистики за 2022 р.

2023 р. (НМТ) Українська мова			
Чоловіки		Жінки	
Середнє	21,527916	Середнє	25,490356
Стандартна помилка	0,0243417	Стандартна помилка	0,0251514
Медіана	20	Медіана	24
Мода	17	Мода	19
Стандартне відхилення	8,7225015	Стандартне відхилення	9,4014602
Дисперсія вибірки	76,082032	Дисперсія вибірки	88,387453
Ексцес	-0,313825	Ексцес	-0,941118
Асиметричність	0,6260764	Асиметричність	0,21696
Інтервал	45	Інтервал	45
Мінімум	0	Мінімум	0
Максимум	45	Максимум	45
Сума	2764292	Сума	3561589
Кількість	128405	Кількість	139723

Рисунок 3.4 – Показники описової статистики за 2023 р.

Для продовження аналізу вибірок було застосовано мову програмування R для статистичної обробки даних.

Перш ніж порівнювати результати тестувань з української мови між чоловіками та жінками була створена нова таблиця даних з балами чоловіків та жінок у форматі csv та побудовано емпіричні функції розподілу (ЕФР) сформованих вибірок.

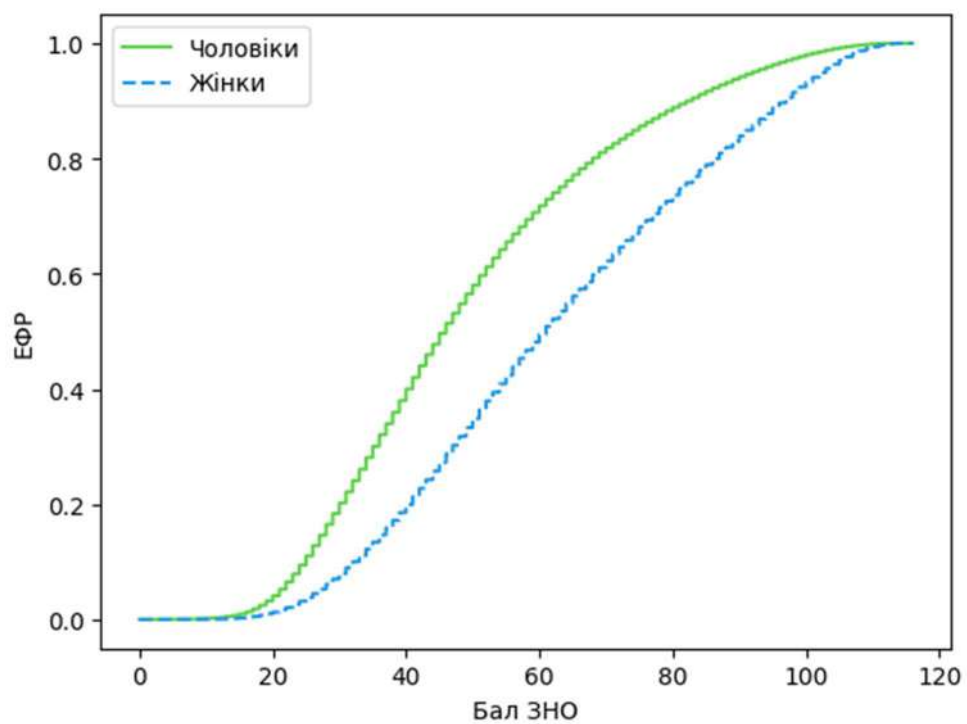


Рисунок 3.5 – Емпірична функція розподілу (Українська мова 2021 р.)

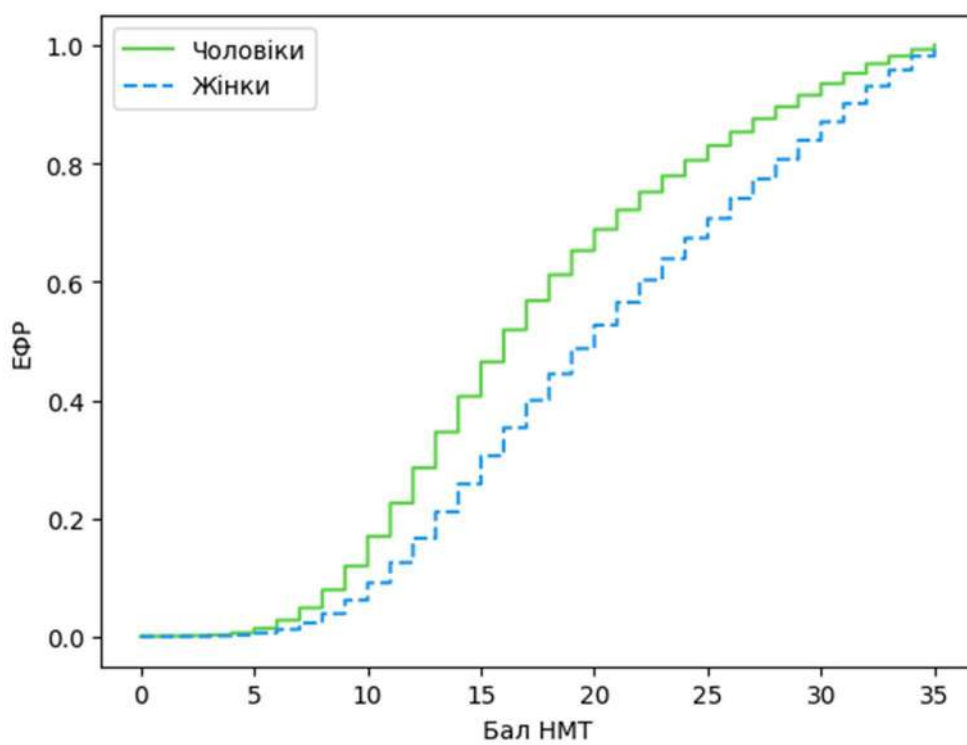


Рисунок 3.6 – Емпірична функція розподілу (Українська мова 2022 р.)

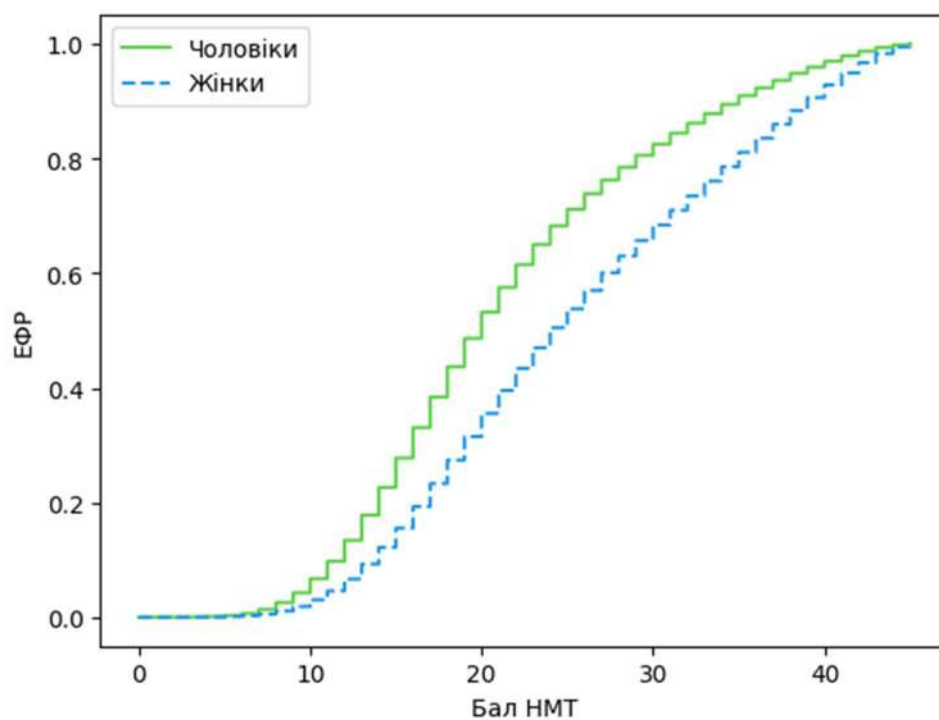


Рисунок 3.7 – Емпірична функція розподілу (Українська мова 2023 р.)

Також для візуалізації даних побудуємо гістограми розподілу:

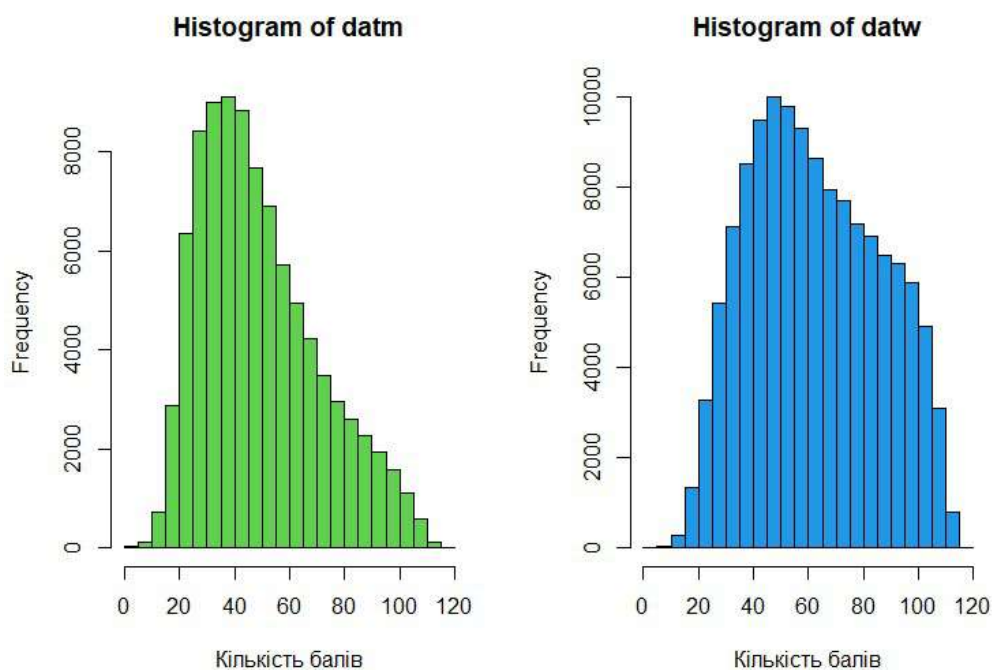


Рисунок 3.8 – Гістограма розподілу показників (Українська мова 2021р.)

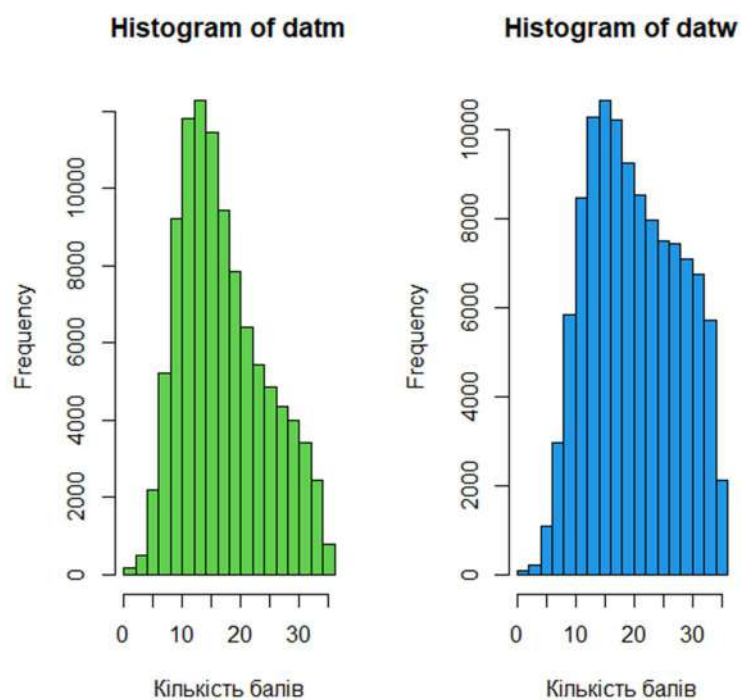


Рисунок 3.9 – Гістограма розподілу показників (Українська мова 2022р.)

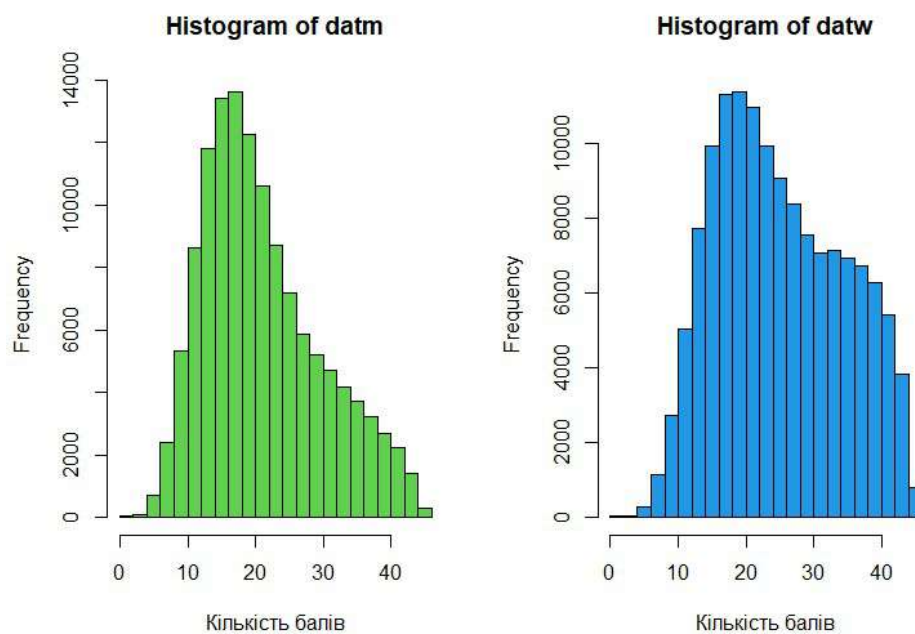


Рисунок 3.10 – Гістограма розподілу показників (Українська мова 2023 р.)

Побудуємо boxplot діаграми (діаграма типу “ящик з вусами”):

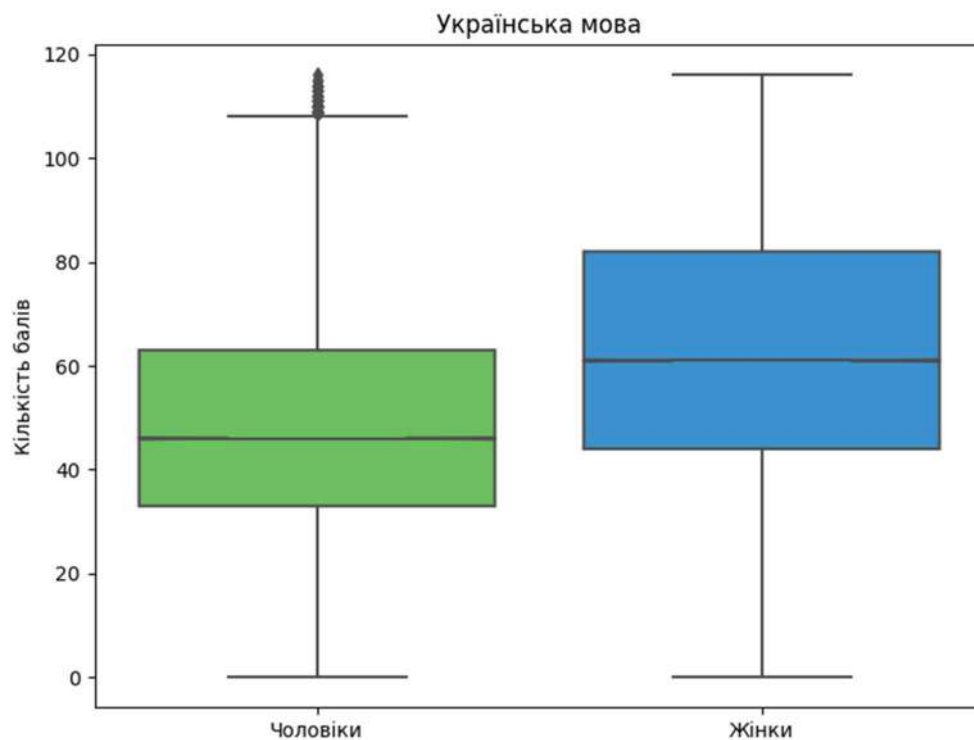


Рисунок 3.11 – Boxplot діаграма розподілу результатів чоловіків та жінок (Українська мова 2021 р.)

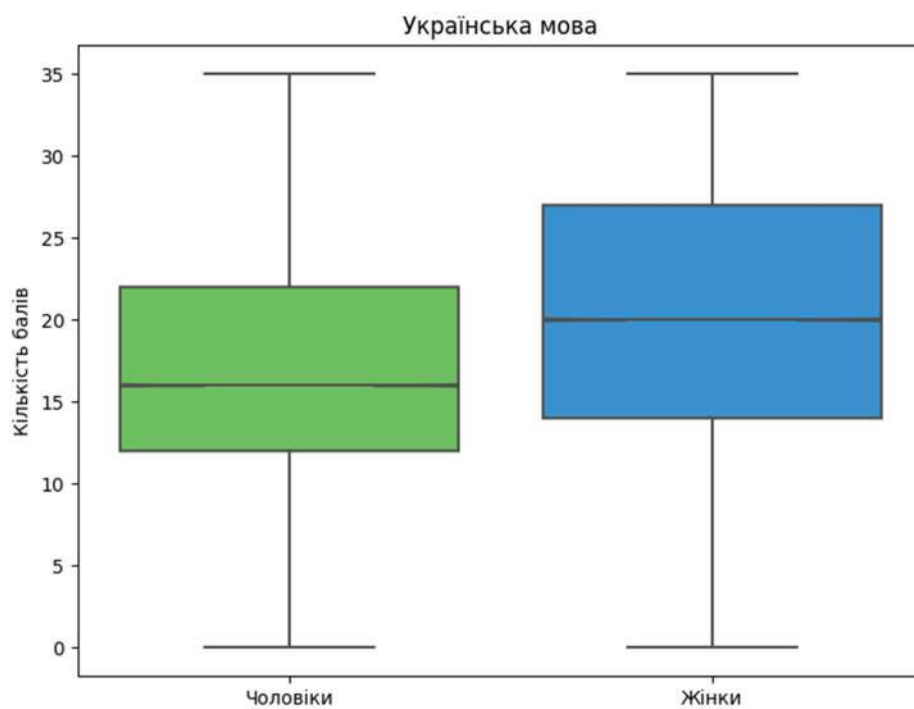


Рисунок 3.12 – Boxplot діаграма розподілу результатів чоловіків та жінок (Українська мова 2022 р.)

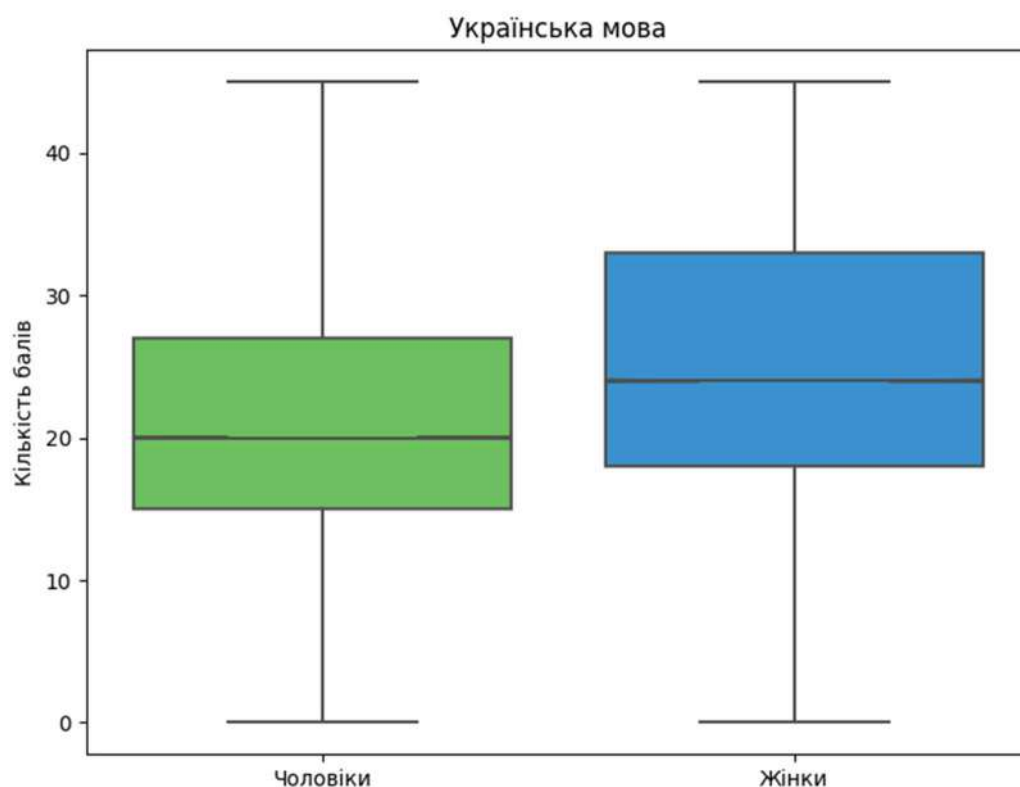


Рисунок 3.13 – Boxplot діаграма розподілу результатів чоловіків та жінок
(Українська мова 2023р.)

Для подальшого докладного кількісного порівняння характеристик розподілів було визначено їх квантили.

Отримані результати наведено у таблиці 3.3.

Таблиця 3.3 – Квантили розподілу результатів з української мови

2021 р.	10%	25%	75%	90%	Макс
Чоловіки	25	33	63	83	116
Жінки	32	44	82	97	116
2022 р.	10%	25%	75%	90%	Макс
Чоловіки	9	12	22	29	35
Жінки	11	14	27	31	35
2023 р.	10%	25%	75%	90%	Макс
Чоловіки	12	15	27	35	45
Жінки	14	18	33	39	45

З розрахунків видно, що всі квантилі для вибірки чоловіків є істотно нижчими, ніж для жінок.

Для вибору статистичних критеріїв перевірки однорідності вибірок необхідно перевірити підпорядкованість даних нормальному розподілу. Для цього було побудовано Q-Q діаграми.

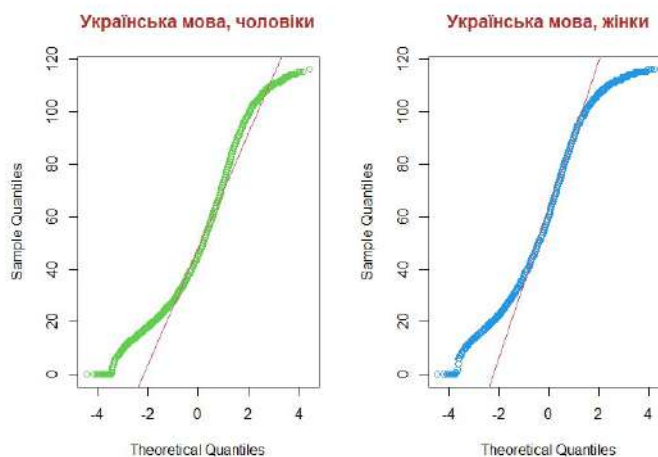


Рисунок 3.14 – Результат побудови Q-Q діаграми (Українська мова 2021р.)

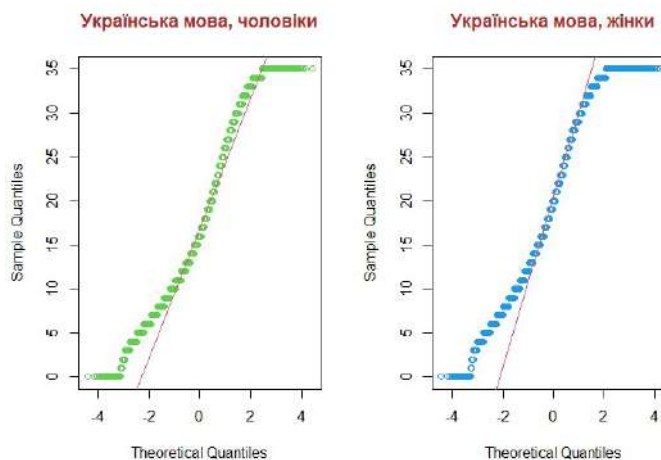


Рисунок 3.15 – Результат побудови Q-Q діаграми (Українська мова 2022р.)

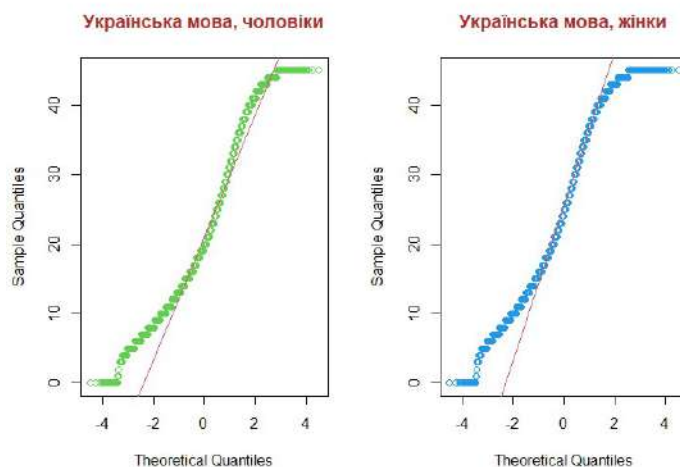


Рисунок 3.16 – Результат побудови Q-Q діаграми (Українська мова 2023р.)

З результату побудови Q-Q діаграм видно, що у всіх випадках є істотні відхилення даних від нормального розподілу.

Перевіримо відповідність даних нормальному розподілу за допомогою статистичних критеріїв та занесемо отримані дані до табл. 3.4.

Таблиця 3.4 – Результати перевірки нормальності

Вибірка	Критерії			
	Ліллєфорса		Андерсона-Дарлінга	
	D	p-value	A	p-value
Укр.мова, чоловіки 2021р.	0.083712	$< 2.2 \cdot 10^{-16}$	1197.5	$< 2.2 \cdot 10^{-16}$
Укр.мова, жінки 2021р.	0.056998	$< 2.2 \cdot 10^{-16}$	859.99	$< 2.2 \cdot 10^{-16}$
Укр.мова, чоловіки 2022р.	0.10062	$< 2.2 \cdot 10^{-16}$	1265.2	$< 2.2 \cdot 10^{-16}$
Укр.мова, жінки 2022р.	0.076235	$< 2.2 \cdot 10^{-16}$	919.17	$< 2.2 \cdot 10^{-16}$

Продовження табл.3.4.

Укр.мова, чоловіки 2023р	0.10062	$< 2.2 \cdot 10^{-16}$	1265.2	$< 2.2 \cdot 10^{-16}$
Укр.мова, жінки 2023р.	0.076235	$< 2.2 \cdot 10^{-16}$	919.17	$< 2.2 \cdot 10^{-16}$

Критерії Ліллієфорса та Андерсона-Дарлінга показують, що розподіли результатів тестувань у чоловіків та жінок не відповідають нормальному розподілу. P-value у обох критеріях менше за 0.05, що свідчить про статистичну значущість відхилень від нормального розподілу. Тому для перевірки гіпотези про однорідність вибірок чоловіків та жінок треба використовувати непараметричні критерії, для цього застосуємо критерій Колмогорова – Смирнова та критерій Уїлкоксона.

Таблиця 3.5 – Результат перевірки однорідності показників з української мови

Вибірка	Критерії			
	Колмогорова-Смирнова		Уїлкоксона	
	D	p-value	W	p-value
2021 р.	0.23318	$< 2.2 \cdot 10^{-16}$	4056221866	$< 2.2 \cdot 10^{-16}$
2022 р.	0.16838	$< 2.2 \cdot 10^{-16}$	4414890601	$< 2.2 \cdot 10^{-16}$
2023 р.	0.18095	$< 2.2 \cdot 10^{-16}$	6762127866	$< 2.2 \cdot 10^{-16}$

Критерій Колмогорова-Смирнова та Уїлкоксона також показують, що є статистично значуща різниця між результатами чоловіків та жінок. P-value менше за 0.05, що означає, що нульова гіпотеза про відсутність різниці у розподілах має бути відхилена.

3.2 Аналіз результатів тестувань з математики

Таблиця 3.6 – Кількість учасників тестування з математики (за 3 роки)

2021 р.		2022 р.		2023 р.	
Кількість учасників	244202	Кількість учасників	214074	Кількість учасників	268128

Таблиця 3.7 – Процентне співвідношення чоловіків та жінок, які беруть участь у тестуванні

Стать	2021 р. (ЗНО)		2022 р. (НМТ)		2023 р. (НМТ)	
	кількість	% від загальної кількості учасників	кількість	% від загальної кількості учасників	кількість	% від загальної кількості учасників
Чоловіча	125319	51,32%	101784	47,55%	128405	47,89%
Жіноча	118883	48,68%	112290	52,45%	139723	52,11%

2021 р. (ЗНО) Математика			
Чоловіки		Жінки	
Середнє	19,641459	Середнє	19,566809
Стандартна помилка	0,0416999	Стандартна помилка	0,0411661
Медіана	14	Медіана	14
Мода	8	Мода	8
Стандартне відхилення	14,761946	Стандартне відхилення	14,193822
Дисперсія вибірки	217,91506	Дисперсія вибірки	201,46458
Ексцес	0,4397838	Ексцес	0,2277122
Асиметричність	1,151225	Асиметричність	1,0544567
Інтервал	67	Інтервал	67
Мінімум	0	Мінімум	0
Максимум	67	Максимум	67
Сума	2461448	Сума	2326161
Кількість	125319	Кількість	118883

Рисунок 3.17 – Показники описової статистики за 2021 р.

2022 р. (НМТ) Математика			
Чоловіки		Жінки	
Середнє	14,147685	Середнє	13,934687
Стандартна помилка	0,0244946	Стандартна помилка	0,0228579
Медіана	12	Медіана	12
Мода	8	Мода	7
Стандартне відхилення	7,8146594	Стандартне відхилення	7,6596015
Дисперсія вибірки	61,068901	Дисперсія вибірки	58,669496
Ексцес	-0,888516	Ексцес	-0,849728
Асиметричність	0,5259764	Асиметричність	0,531564
Інтервал	30	Інтервал	30
Мінімум	0	Мінімум	0
Максимум	30	Максимум	30
Сума	1440008	Сума	1564726
Кількість	101784	Кількість	112290

Рисунок 3.18 – Показники описової статистики за 2022 р.

2023 р. (НМТ) Математика			
Чоловіки		Жінки	
Середнє	11,990117	Середнє	11,923012
Стандартна помилка	0,0197312	Стандартна помилка	0,0187582
Медіана	10	Медіана	10
Мода	7	Мода	7
Стандартне відхилення	7,0703984	Стандартне відхилення	7,0117272
Дисперсія вибірки	49,990533	Дисперсія вибірки	49,164318
Ексцес	0,2143546	Ексцес	0,1155028
Асиметричність	0,9851076	Асиметричність	0,9341041
Інтервал	32	Інтервал	32
Мінімум	0	Мінімум	0
Максимум	32	Максимум	32
Сума	1539591	Сума	1665919
Кількість	128405	Кількість	139723

Рисунок 3.19 – Показники описової статистики за 2023 р.

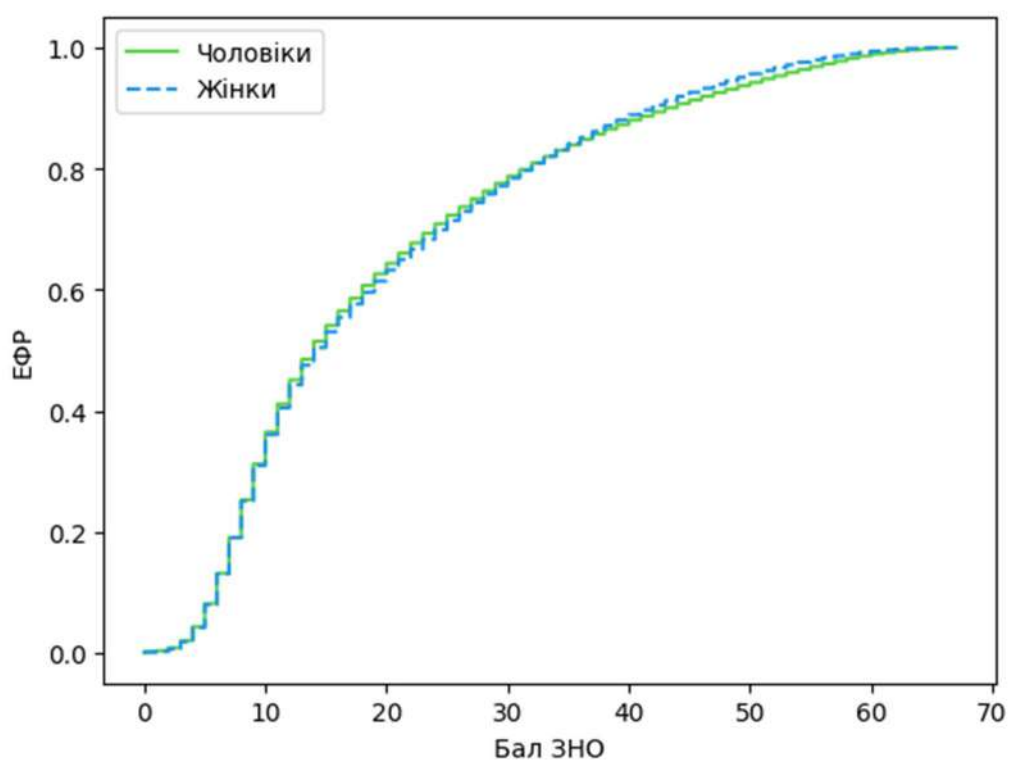


Рисунок 3.20 – Емпірична функція розподілу (Математика 2021 р.)

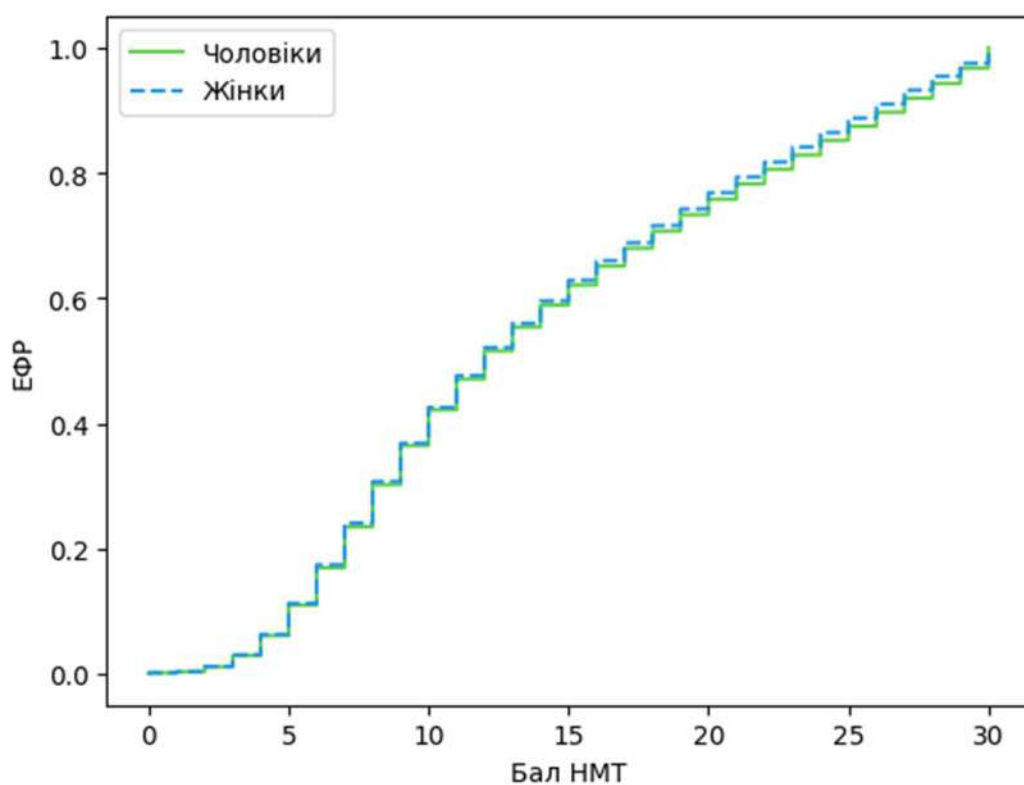


Рисунок 3.21 – Емпірична функція розподілу (Математика 2022 р.)

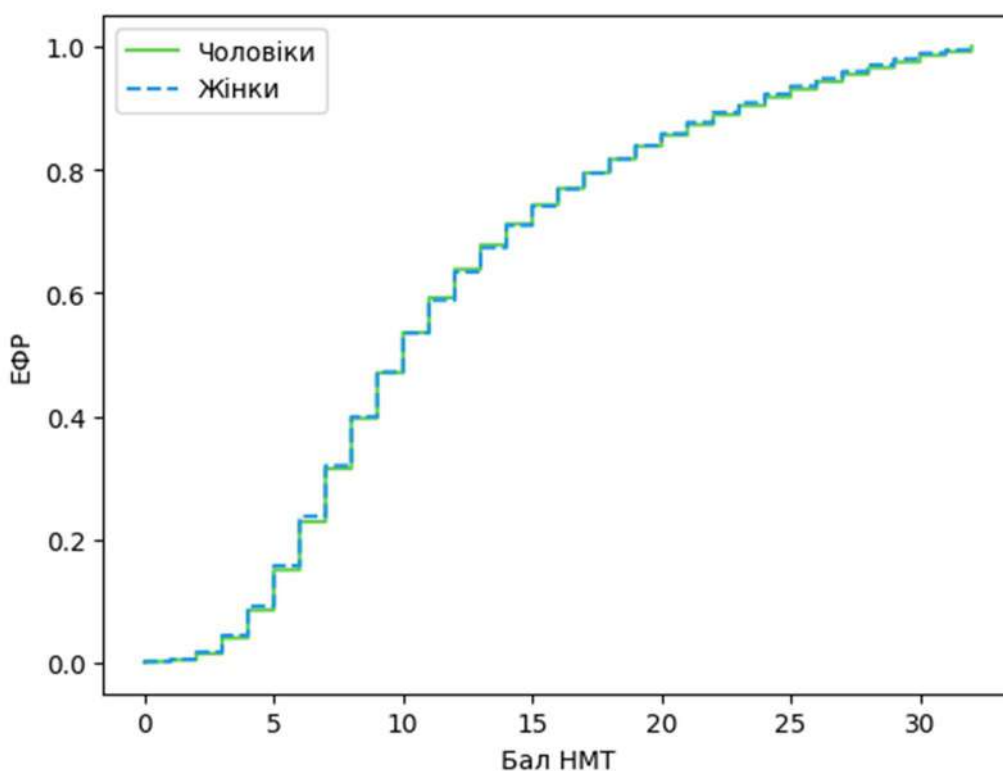


Рисунок 3.22 – Емпірична функція розподілу (Математика 2023 р.)

Також для візуалізації даних побудуємо гістограми розподілу:

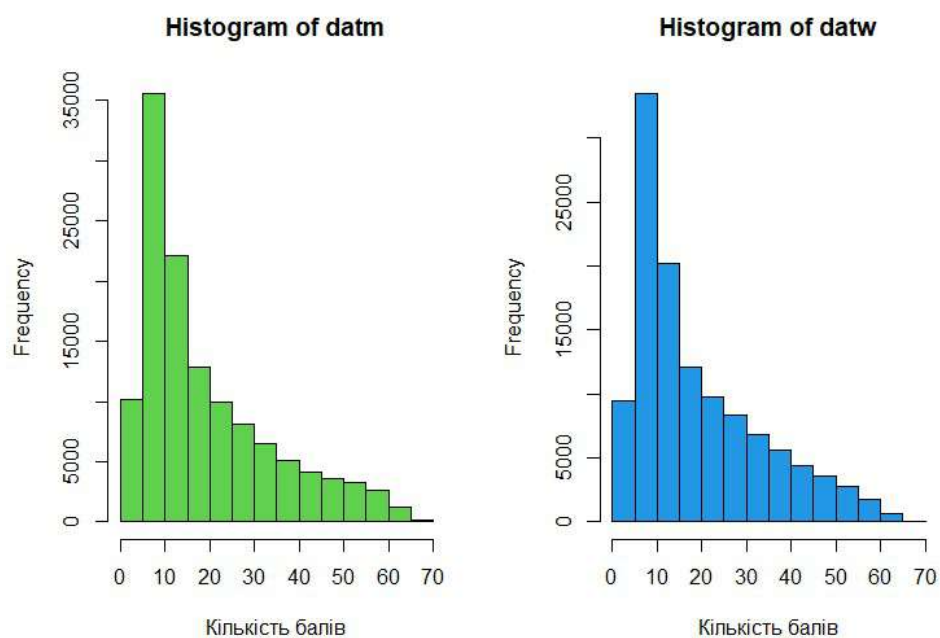


Рисунок 3.23 – Гістограма розподілу показників (Математика 2021 р.)

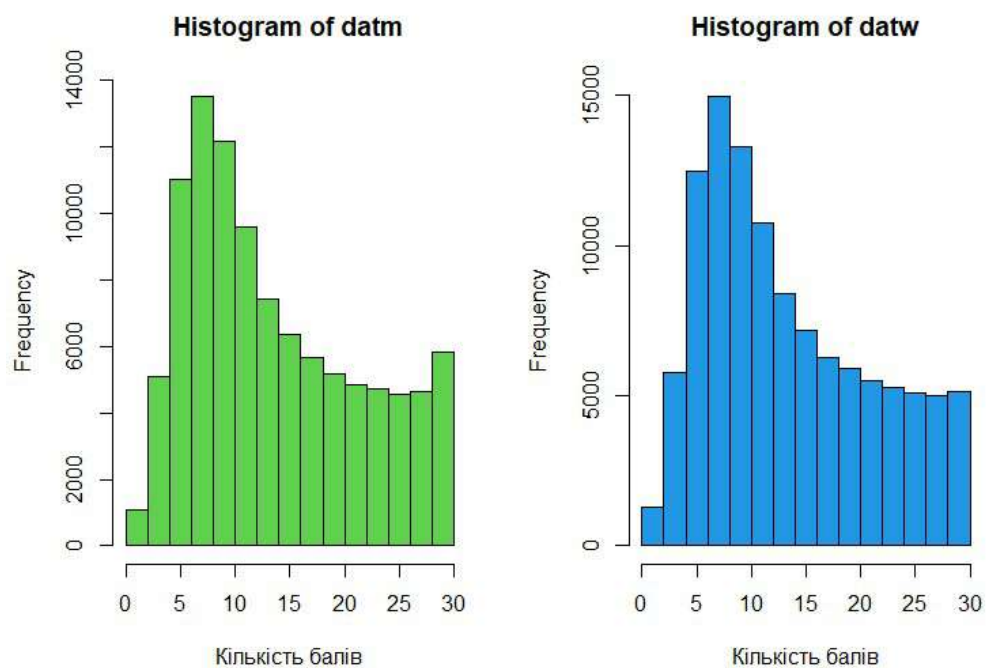


Рисунок 3.24 – Гістограма розподілу показників (Математика 2022р.)

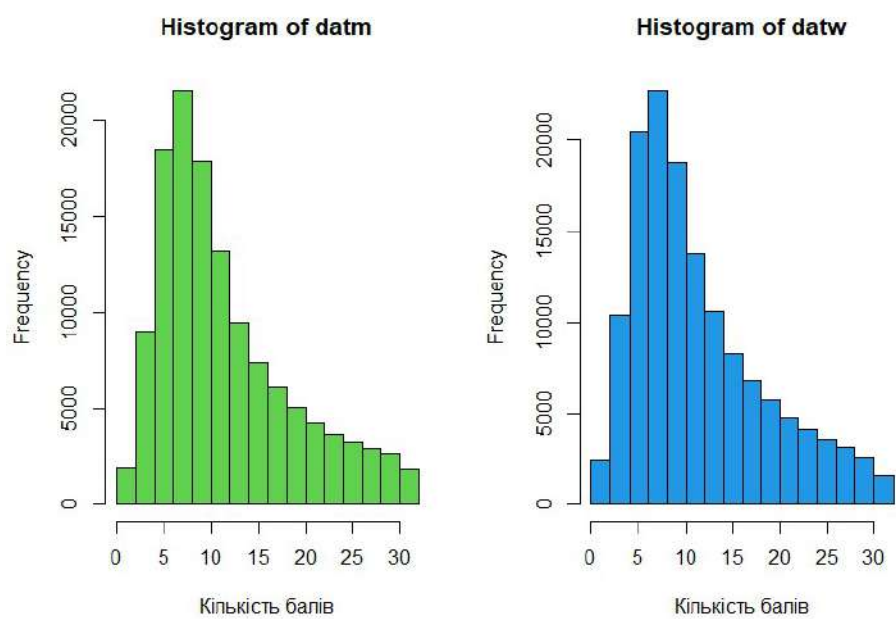


Рисунок 3.25 – Гістограма розподілу показників (Математика 2023 р.)

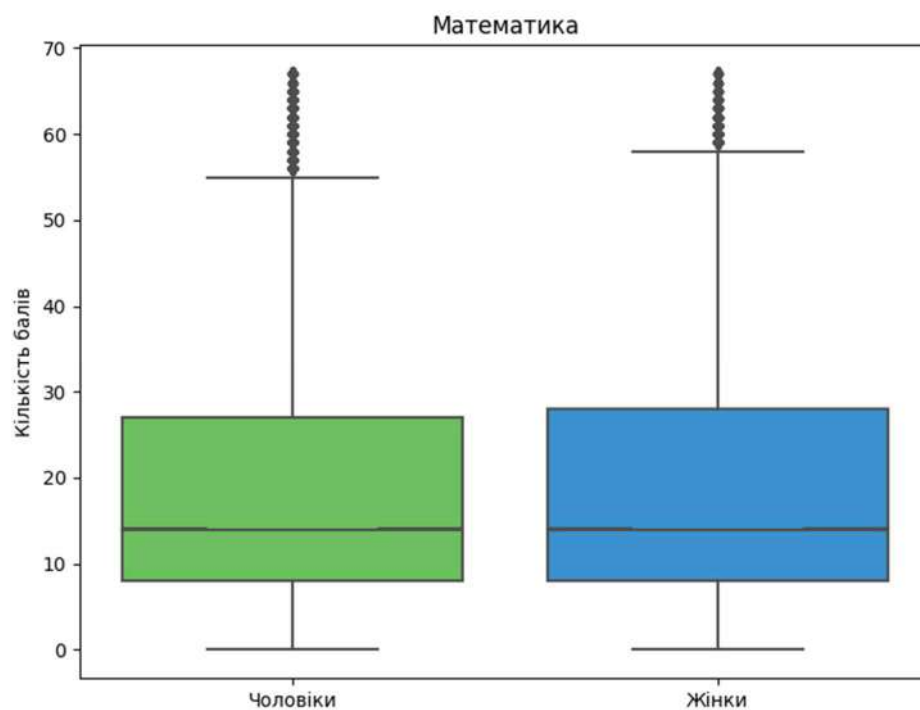


Рисунок 3.26 – Boxplot діаграма розподілу результатів чоловіків та жінок (Математика 2021р.)

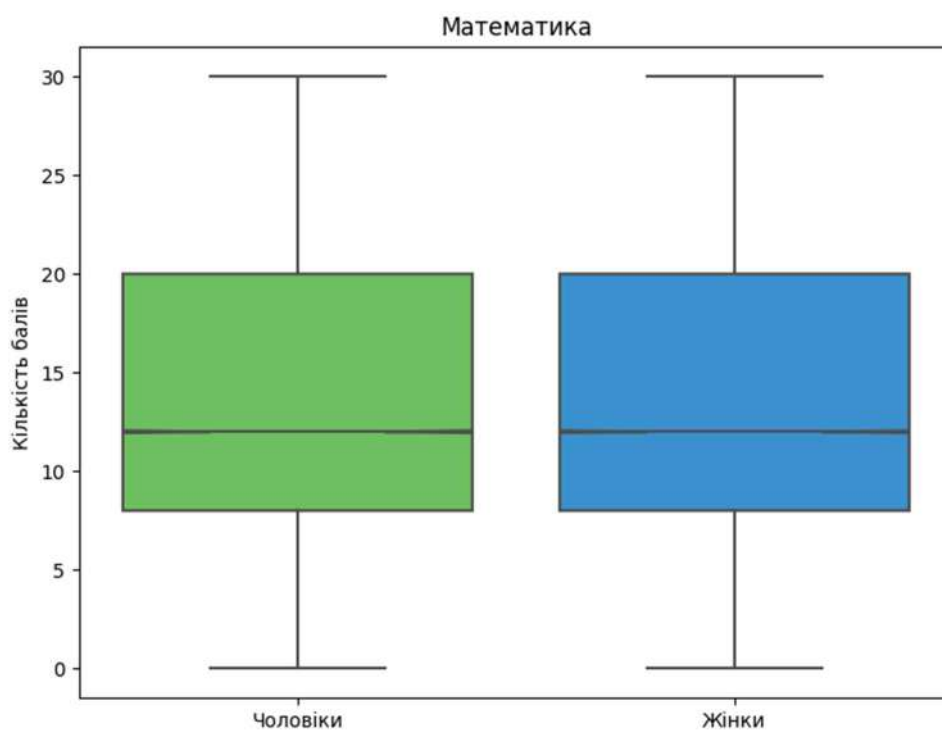


Рисунок 3.27 – Boxplot діаграма розподілу результатів чоловіків та жінок (Математика 2022р.)

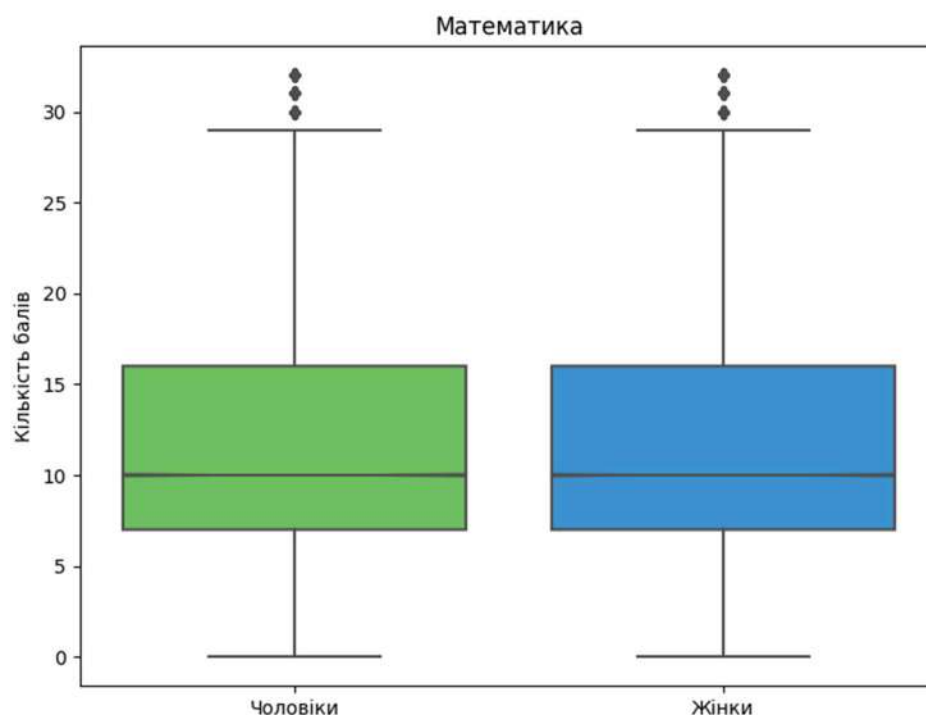


Рисунок 3.28 – Boxplot діаграма розподілу результатів чоловіків та жінок
(Математика 2023р.)

Для подальшого докладного кількісного порівняння характеристик розподілів було визначено їх квантілі.

Отримані результати наведено у таблиці 3.8.

Таблиця 3.8 – Квантілі розподілу результатів з математики

2021 р.	10%	25%	75%	90%	Макс
Чоловіки	6	8	27	43	67
Жінки	6	8	28	42	67
2022 р.	10%	25%	75%	90%	Макс
Чоловіки	5	8	20	27	30
Жінки	5	8	20	26	30
2023 р.	10%	25%	75%	90%	Макс
Чоловіки	5	7	16	23	32
Жінки	5	7	16	23	32

Перевіримо відповідність даних нормальному розподілу за допомогою статистичних критеріїв та занесемо отримані дані до табл. 3.9.

Таблиця 3.9 – Результати перевірки нормальності

Вибірка	Критерії			
	Лілієфорса		Андерсона-Дарлінга	
	D	p-value	A	p-value
Математика, чоловіки 2021р.	0.16412	$< 2.2 \cdot 10^{-1}$	5862.5	$< 2.2 \cdot 10^{-16}$
Математика, жінки 2021р.	0.15673	$< 2.2 \cdot 10^{-16}$	4890.2	$< 2.2 \cdot 10^{-16}$
Математика, чоловіки 2022р.	0.12708	$< 2.2 \cdot 10^{-16}$	2301.5	$< 2.2 \cdot 10^{-16}$
Математика, жінки 2022р.	0.12543	$< 2.2 \cdot 10^{-16}$	2405.4	$< 2.2 \cdot 10^{-16}$
Математика, чоловіки 2023р	0.14793	$< 2.2 \cdot 10^{-1}$	3940.8	$< 2.2 \cdot 10^{-16}$
Математика, жінки 2023р.	0.14237	$< 2.2 \cdot 10^{-16}$	3908.7	$< 2.2 \cdot 10^{-16}$

Критерії Лілієфорса та Андерсона-Дарлінга показують, що розподіли результатів тестувань у чоловіків та жінок не відповідають нормальному розподілу. P-value у обох критеріях менше за 0.05, що свідчить про статистичну значущість відхилень від нормального розподілу. Тому для перевірки гіпотези про однорідність вибірок чоловіків та жінок треба використовувати непараметричні критерії, для цього застосуємо критерій Колмогорова – Смирнова та критерій Уїлкоксона.

Таблиця 3.10 – Результат перевірки однорідності показників з математики

Вибірка	Критерії			
	Колмогорова-Смирнова		Уїлкоксона	
	D	p-value	W	p-value
2021 р.	0.014279	$< 3.14 \cdot 10^{-11}$	7406633701	0.01455
2022 р.	0.013073	$< 2.2 \cdot 10^{-16}$	4414890601	$2.377 \cdot 10^{-8}$
2023 р.	0.0088537	$5.561 \cdot 10^{-5}$	9012242634	0.03708

3.3 Аналіз результатів тестувань з історії України

Таблиця 3.11 – Кількість учасників тестування з історії України (за 3 роки)

2021 р.		2022 р.		2023 р.	
Кількість учасників	199373	Кількість учасників	214074	Кількість учасників	141068

Таблиця 3.12 – Процентне співвідношення чоловіків та жінок, які беруть участь у тестуванні

Стать	2021 р. (ЗНО)		2022 р. (НМТ)		2023 р. (НМТ)	
	кількість	% від загальної кількості учасників	кількість	% від загальної кількості учасників	кількість	% від загальної кількості учасників
Чоловіча	91026	45,66%	101784	47,55%	68844	48,80%
Жіноча	108347	54,34%	112290	52,45%	72224	51,20%

2021 р. (ЗНО) Історія України			
Чоловіки		Жінки	
Середнє	37,225968	Середнє	41,012986
Стандартна помилка	0,0480776	Стандартна помилка	0,0481524
Медіана	34	Медіана	37
Мода	30	Мода	32
Стандартне відхилення	14,505264	Стандартне відхилення	15,849892
Дисперсія вибірки	210,40269	Дисперсія вибірки	251,21906
Ексцес	1,6271552	Ексцес	0,6111577
Асиметричність	1,2842695	Асиметричність	1,0731163
Інтервал	94	Інтервал	94
Мінімум	0	Мінімум	0
Максимум	94	Максимум	94
Сума	3388531	Сума	4443634
Кількість	91026	Кількість	108347

Рисунок 3.29 – Показники описової статистики за 2021 р.

2022 р. (НМТ) Історія України			
Чоловіки		Жінки	
Середнє	17,248713	Середнє	18,350521
Стандартна помилка	0,0247429	Стандартна помилка	0,0248502
Медіана	16	Медіана	17
Мода	11	Мода	11
Стандартне відхилення	7,8938688	Стандартне відхилення	8,3272323
Дисперсія вибірки	62,313165	Дисперсія вибірки	69,342798
Ексцес	-0,630631	Ексцес	-0,920586
Асиметричність	0,5569518	Асиметричність	0,4113234
Інтервал	35	Інтервал	35
Мінімум	0	Мінімум	0
Максимум	35	Максимум	35
Сума	1755643	Сума	2060580
Кількість	101784	Кількість	112290

Рисунок 3.30 – Показники описової статистики за 2022 р.

2023 р. (НМТ) Історія України			
Чоловіки		Жінки	
Середнє	22,334844	Середнє	24,080486
Стандартна помилка	0,0343259	Стандартна помилка	0,0372077
Медіана	20	Медіана	21
Мода	18	Мода	18
Стандартне відхилення	9,0064715	Стандартне відхилення	9,9993999
Дисперсія вибірки	81,116529	Дисперсія вибірки	99,987997
Ексцес	1,4087104	Ексцес	0,4220343
Асиметричність	1,2552501	Асиметричність	1,0270199
Інтервал	54	Інтервал	54
Мінімум	0	Мінімум	0
Максимум	54	Максимум	54
Сума	1537620	Сума	1739189
Кількість	68844	Кількість	72224

Рисунок 3.31 – Показники описової статистики за 2023 р.

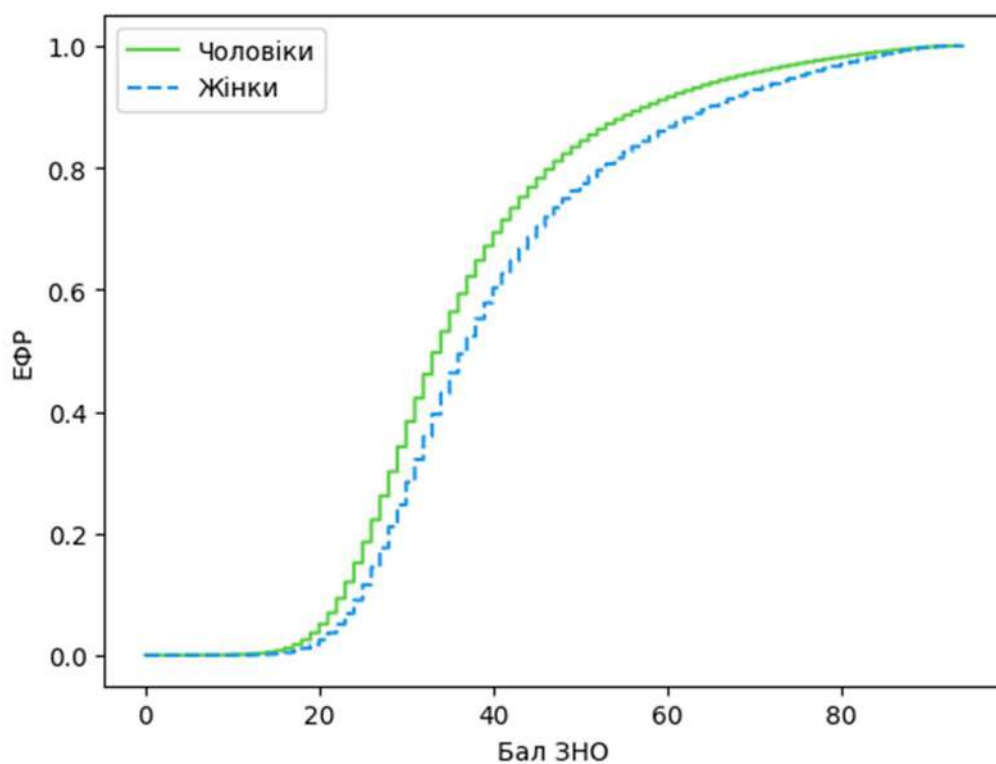


Рисунок 3.32 – Емпірична функція розподілу (Історія України 2021 р.)

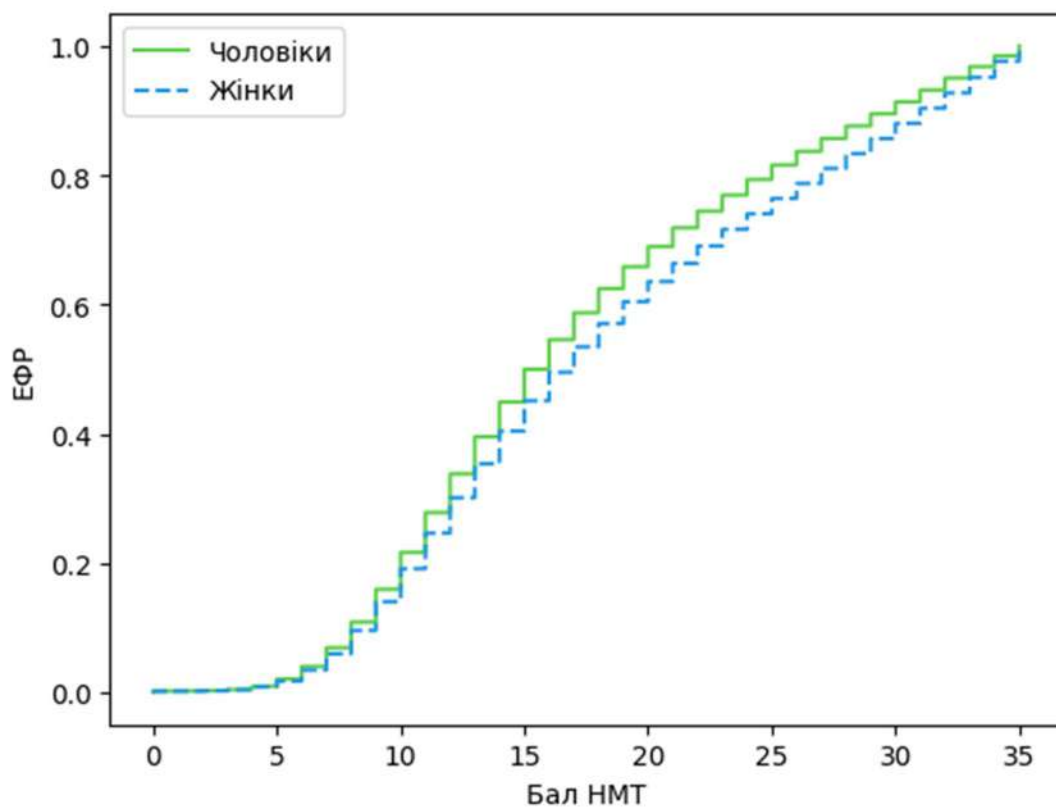


Рисунок 3.33 – Емпірична функція розподілу (Історія України 2022 р.)

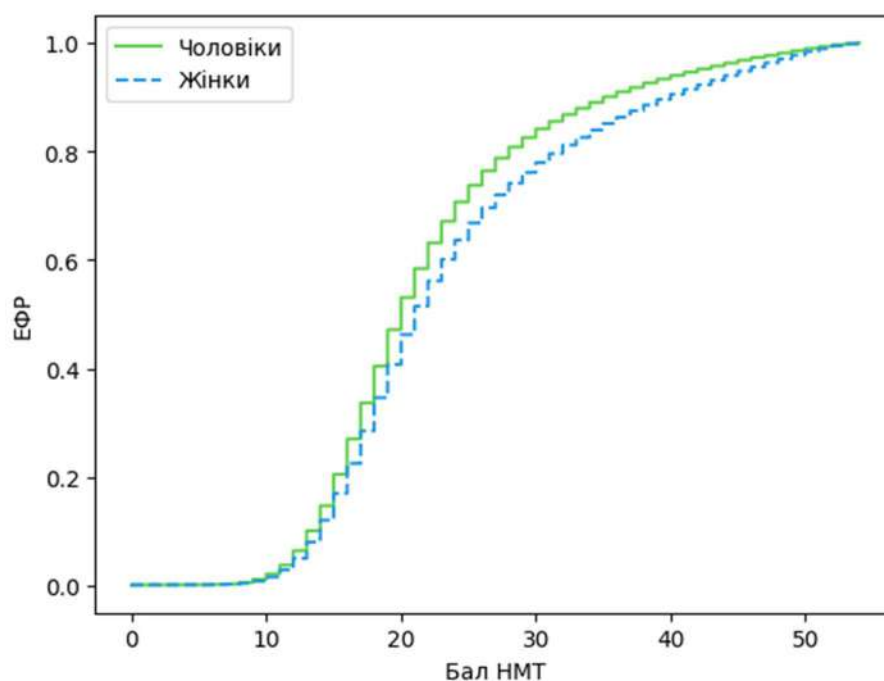


Рисунок 3.34 – Емпірична функція розподілу (Історія України 2023 р.)

Також для візуалізації даних побудуємо гістограми розподілу та boxplot діаграми:

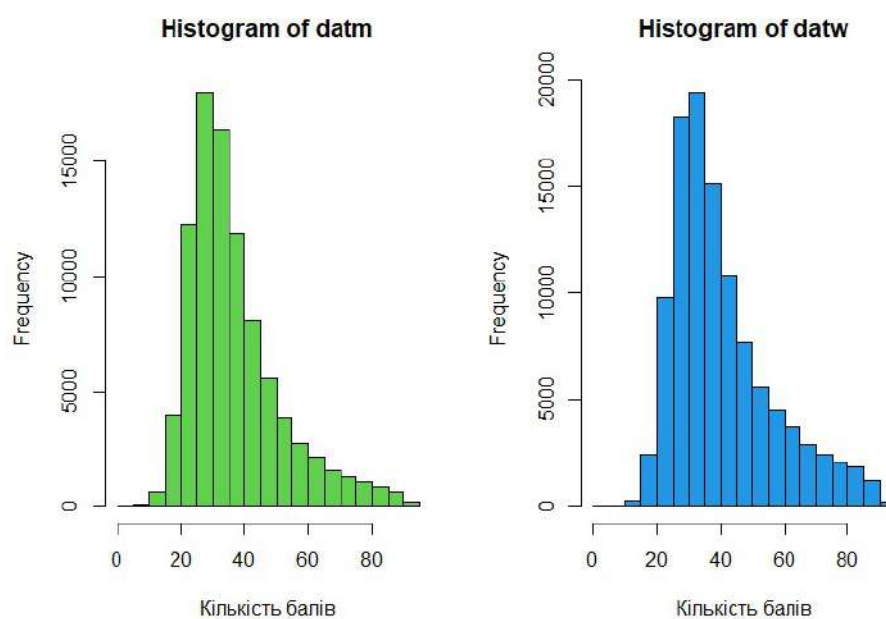


Рисунок 3.35 – Гістограма розподілу показників (Історія України 2021р.)

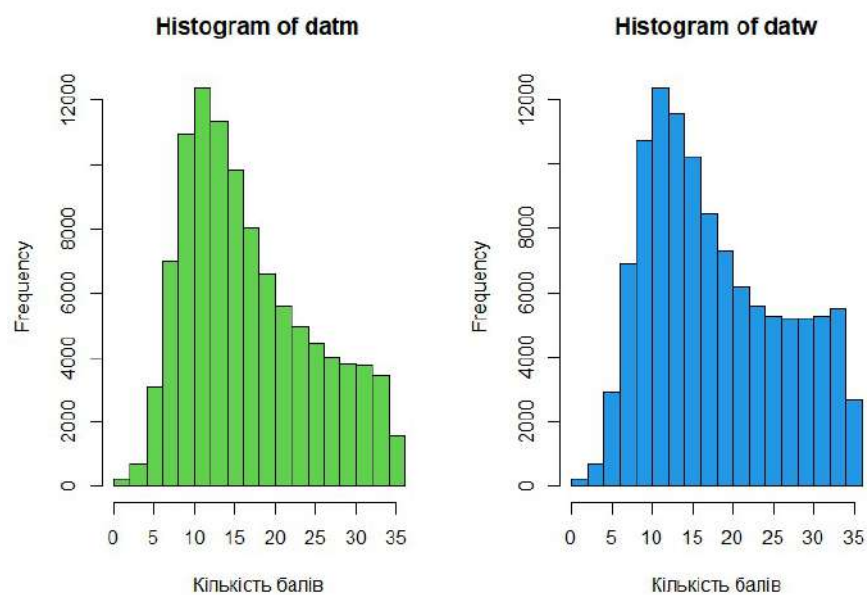


Рисунок 3.36 – Гістограма розподілу показників (Історія України 2022р.)

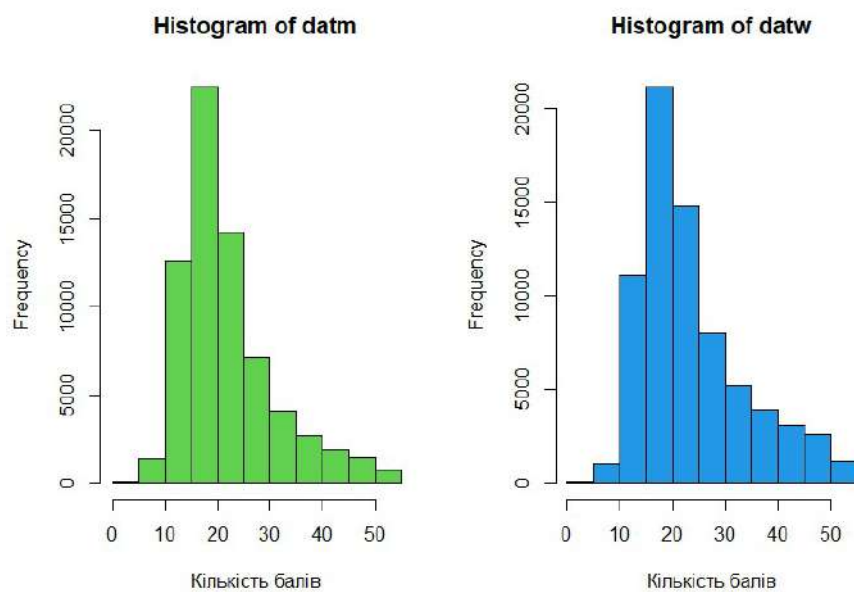


Рисунок 3.37 – Гістограма розподілу показників (Історія України 2023 р.)

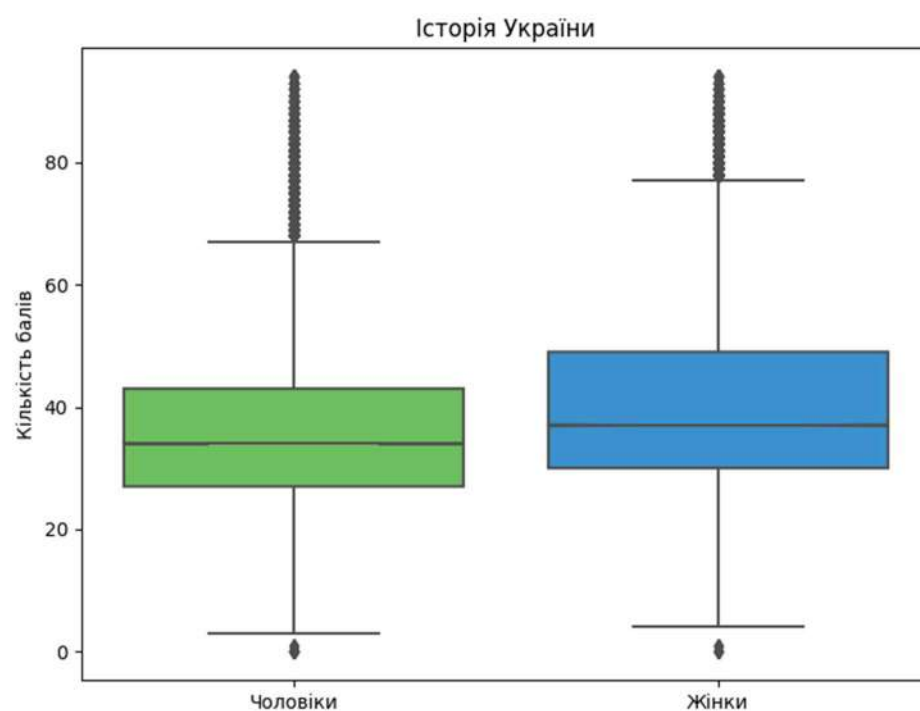


Рисунок 3.38 – Boxplot діаграми розподілу результатів за 2021 рік

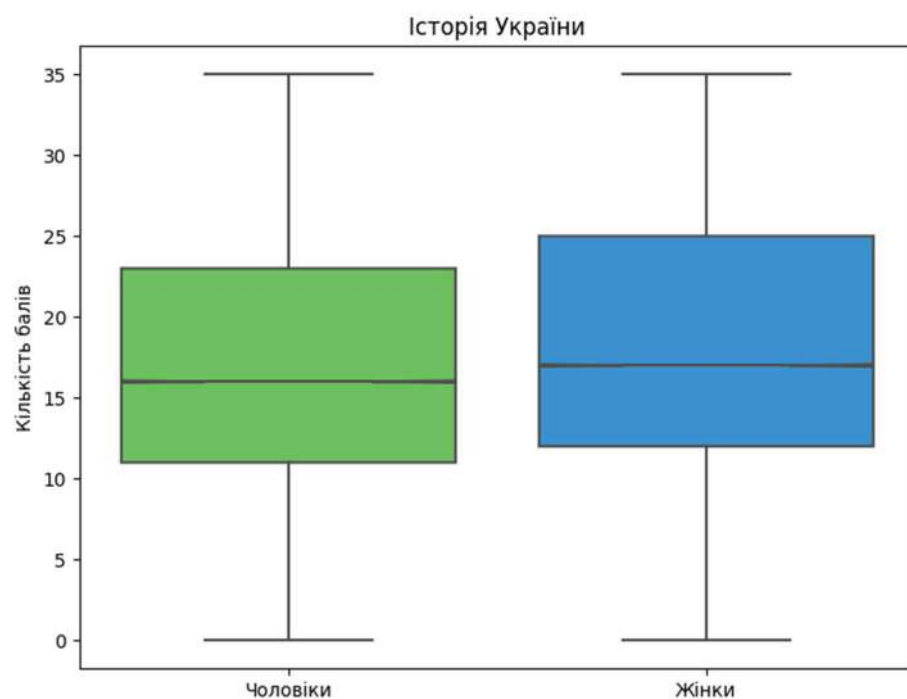


Рисунок 3.39 – Boxplot діаграми розподілу результатів за 2022 рік

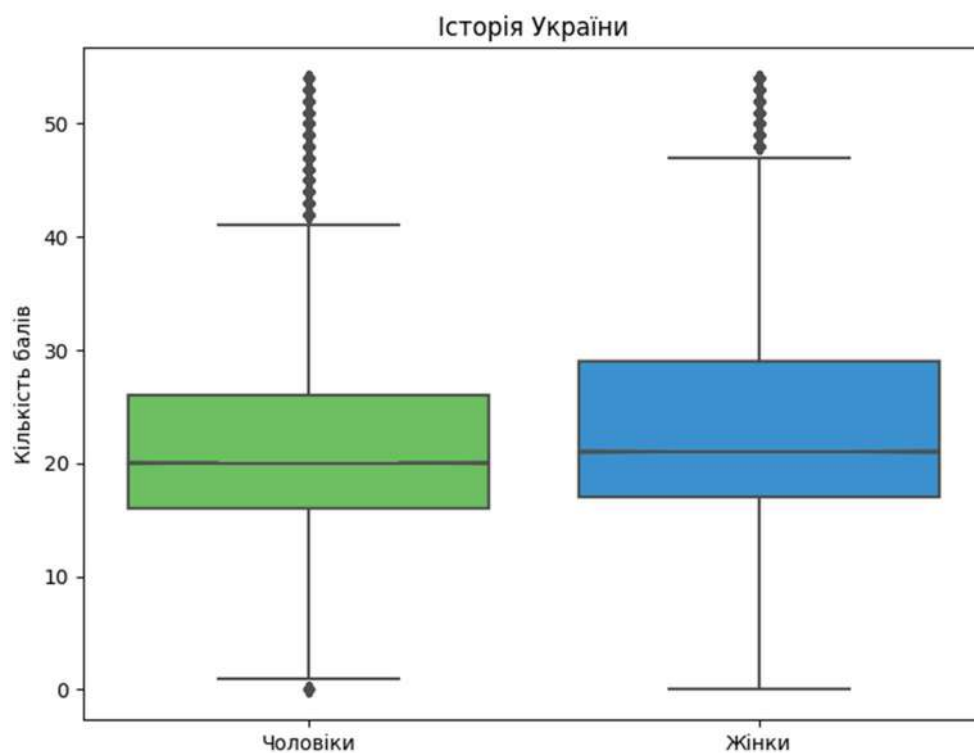


Рисунок 3.40 – Voxplot діаграми розподілу результатів за 2023 рік

Для подальшого докладного кількісного порівняння характеристик розподілів було визначено їх квантили.

Отримані результати наведено у таблиці 3.13.

Таблиця 3.13 – Квантилі розподілу даних результатів з історії України

2021 р.	10%	25%	75%	90%	Макс
Чоловіки	23	27	43	58	94
Жінки	25	30	49	65	94
2022 р.	10%	25%	75%	90%	Макс
Чоловіки	8	11	23	30	35
Жінки	9	12	25	31	35
2023 р.	10%	25%	75%	90%	Макс
Чоловіки	13	16	26	35	54
Жінки	14	17	29	40	54

Перевіримо відповідність даних нормальному розподілу за допомогою статистичних критеріїв та занесемо отримані дані до табл. 3.14.

Таблиця 3.14 – Результати перевірки нормальності

Вибірка	Критерії			
	Лілієфорса		Андерсона-Дарлінга	
	D	p-value	A	p-value
Іст.Укр, чоловіки 2021р.	0.12761	$< 2.2 \cdot 10^{-16}$	2648.5	$< 2.2 \cdot 10^{-16}$
Іст.Укр, жінки 2021р.	0.12581	$< 2.2 \cdot 10^{-16}$	3145.1	$< 2.2 \cdot 10^{-16}$
Іст.Укр, чоловіки 2022р.	0.11181	$< 2.2 \cdot 10^{-16}$	1738.4	$< 2.2 \cdot 10^{-16}$
Іст.Укр, жінки 2022р.	0.10713	$< 2.2 \cdot 10^{-16}$	1864.4	$< 2.2 \cdot 10^{-16}$
Іст.Укр, чоловіки 2023р	0.14608	$< 2.2 \cdot 10^{-16}$	2257.9	$< 2.2 \cdot 10^{-1}$
Іст.Укр, жінки 2023р.	0.14372	$< 2.2 \cdot 10^{-16}$	2286.8	$< 2.2 \cdot 10^{-16}$

Критерії Лілієфорса та Андерсона-Дарлінга показують, що розподіли результатів тестувань у чоловіків та жінок не відповідають нормальному розподілу. P-value у обох критеріях менше за 0.05, що свідчить про статистичну значущість відхилень від нормального розподілу. Тому для перевірки гіпотези про однорідність вибірок чоловіків та жінок треба використовувати непараметричні критерії, для цього застосуємо критерій Колмогорова – Смирнова та критерій Уїлкоксона.

Таблиця 3.15 – Результат перевірки однорідності показників з історії України

Вибірка	Критерії			
	Колмогорова-Смирнова		Уїлкоксона	
	D	p-value	W	p-value
2021 р.	0.10207	$< 2.2 \cdot 10^{-16}$	4197671985	$< 2.2 \cdot 10^{-16}$
2022 р.	0.055024	$< 2.2 \cdot 10^{-16}$	5299108773	$< 2.2 \cdot 10^{-16}$
2023 р.	0.070237	$< 2.2 \cdot 10^{-16}$	2242592673	$< 2.2 \cdot 10^{-16}$

3.4 Аналіз результатів тестувань з фізики

Таблиця 3.16 – Кількість учасників тестування з фізики (за 3 роки)

2021 р.		2022 р.		2023 р.	
Кількість учасників	23202	Кількість учасників	не проводилося	Кількість учасників	5254

Таблиця 3.17 – Процентне співвідношення чоловіків та жінок, які беруть участь у тестуванні

Стать	2021 р. (ЗНО)		2022 р. (НМТ)		2023 р. (НМТ)	
	кількість	% від загальної кількості учасників	кількість	% від загальної кількості учасників	кількість	% від загальної кількості учасників
Чоловіча	19809	85,38%	не проводилося		4540	86,41%
Жіноча	3393	14,62%			714	13,59%

2021 р. (ЗНО) Фізика			
Чоловіки		Жінки	
Середнє	24,250038	Середнє	24,960212
Стандартна помилка	0,0867168	Стандартна помилка	0,2254359
Медіана	21	Медіана	21
Мода	17	Мода	15
Стандартне відхилення	12,204902	Стандартне відхилення	13,131519
Дисперсія вибірки	148,95964	Дисперсія вибірки	172,4368
Ексцес	0,812657	Ексцес	-0,014299
Асиметричність	1,151298	Асиметричність	0,9056589
Інтервал	64	Інтервал	64
Мінімум	0	Мінімум	0
Максимум	64	Максимум	64
Сума	480369	Сума	84690
Кількість	19809	Кількість	3393

Рисунок 3.41 – Показники описової статистики за 2021 р.

2023 р. (НМТ) Фізика			
Чоловіки		Жінки	
Середнє	12,879515	Середнє	12,941176
Стандартна помилка	0,1073588	Стандартна помилка	0,2666819
Медіана	11	Медіана	11
Мода	7	Мода	10
Стандартне відхилення	7,2337837	Стандартне відхилення	7,1259486
Дисперсія вибірки	52,327626	Дисперсія вибірки	50,779144
Ексцес	-0,236596	Ексцес	-0,435744
Асиметричність	0,738424	Асиметричність	0,6629974
Інтервал	32	Інтервал	31
Мінімум	0	Мінімум	1
Максимум	32	Максимум	32
Сума	58473	Сума	9240
Кількість	4540	Кількість	714

Рисунок 3.42 – Показники описової статистики за 2023 р.

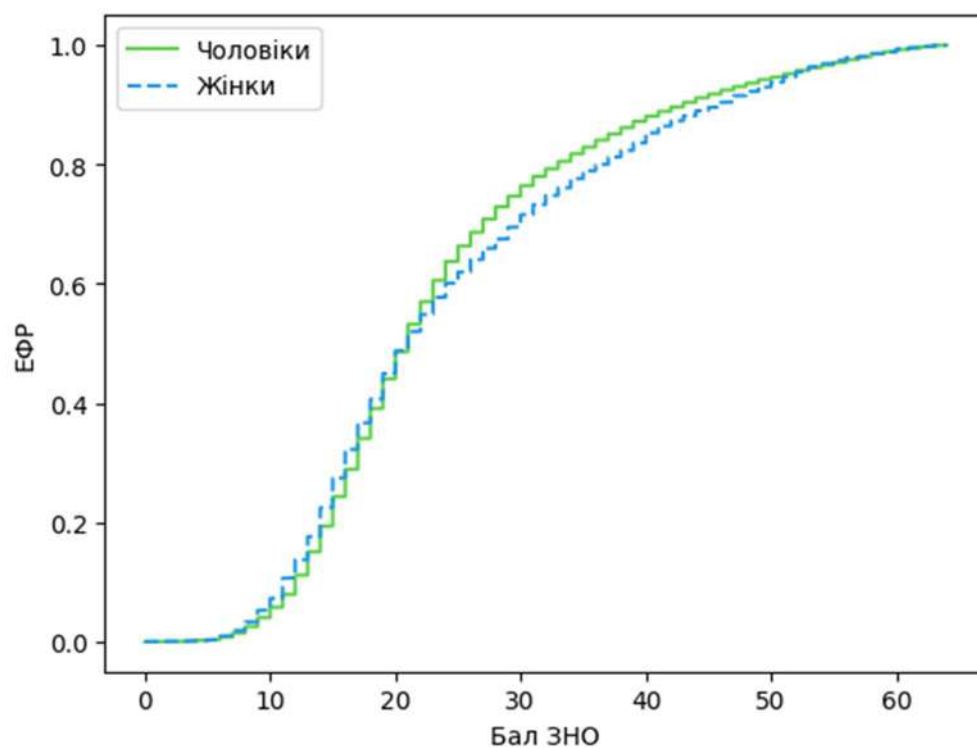


Рисунок 3.43 – Емпірична функція розподілу (Фізика 2021 р.)

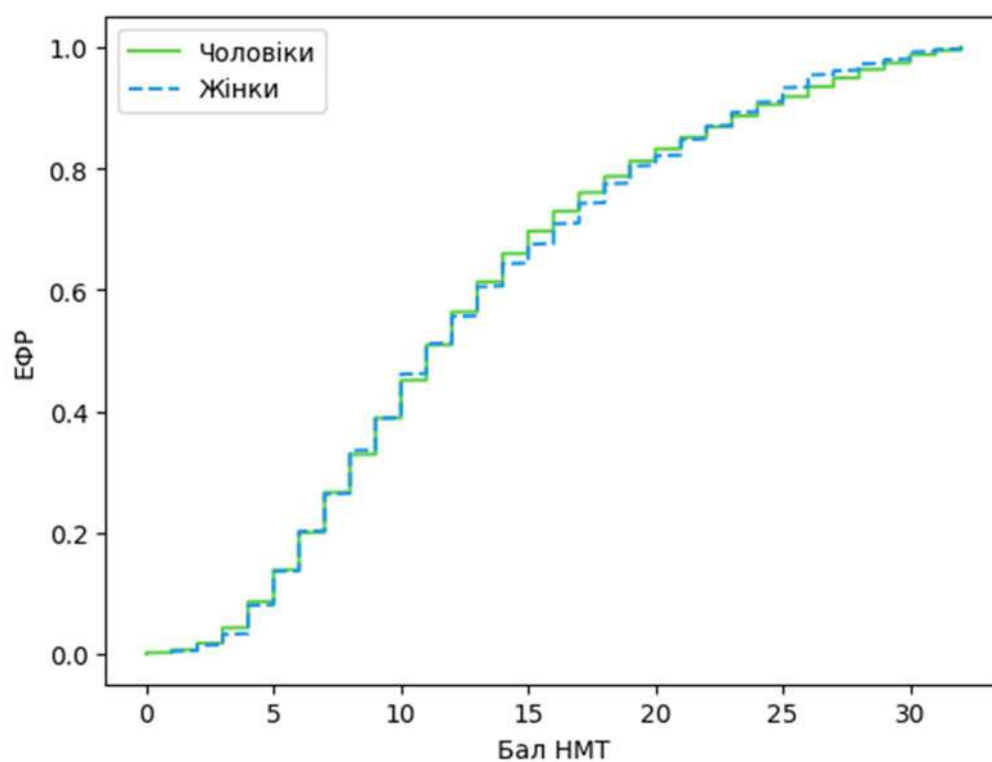


Рисунок 3.44 – Емпірична функція розподілу (Фізика 2023 р.)

Також для візуалізації даних побудуємо гістограми розподілу:

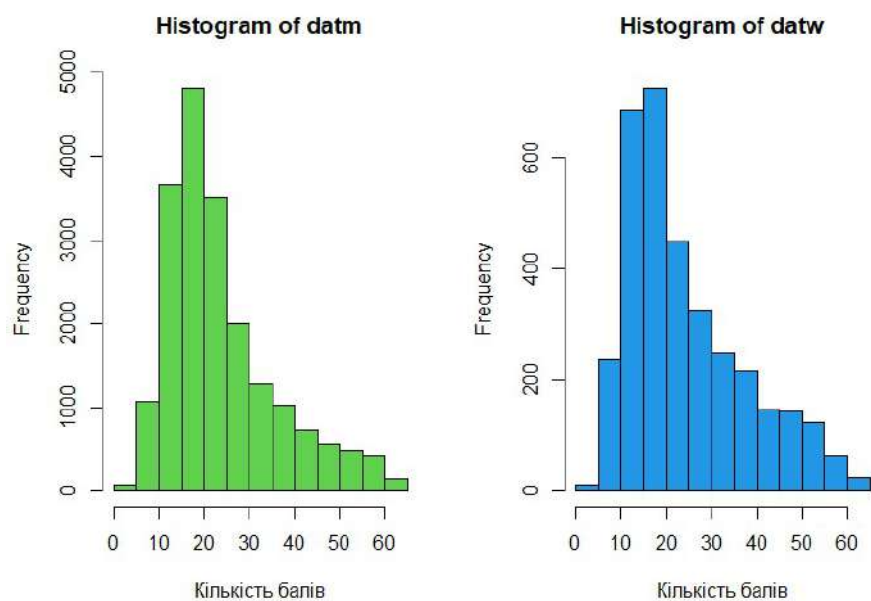


Рисунок 3.45 – Гістограма розподілу показників (Фізика 2021 р.)

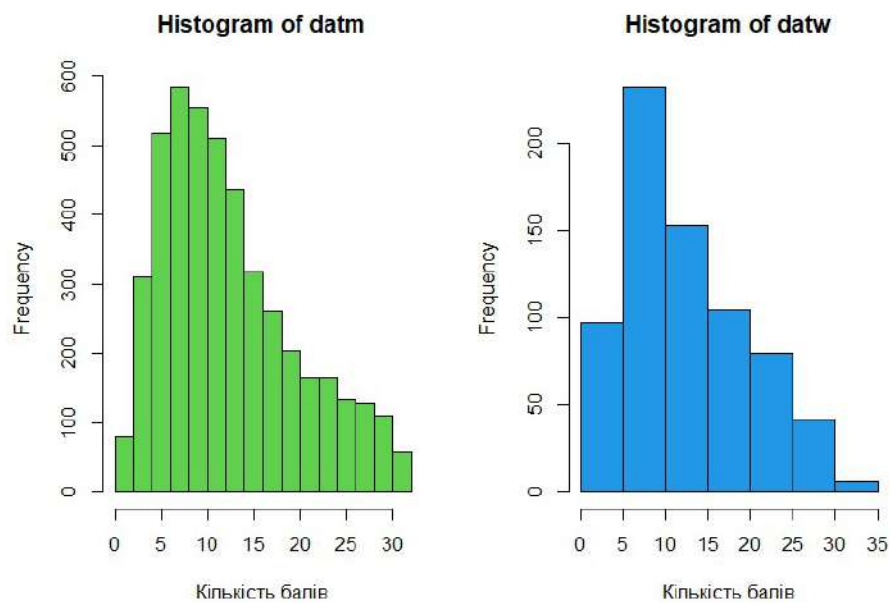


Рисунок 3.46 – Гістограма розподілу показників (Фізика 2023 р.)

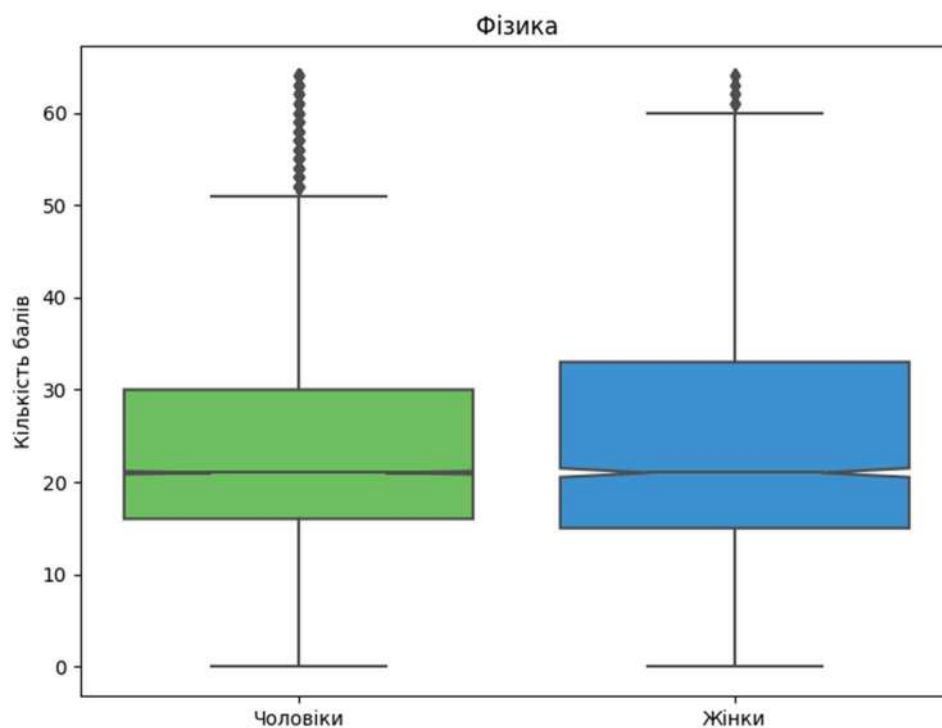


Рисунок 3.47 – Boxplot діаграма розподілу результатів (Фізика 2021 р.)

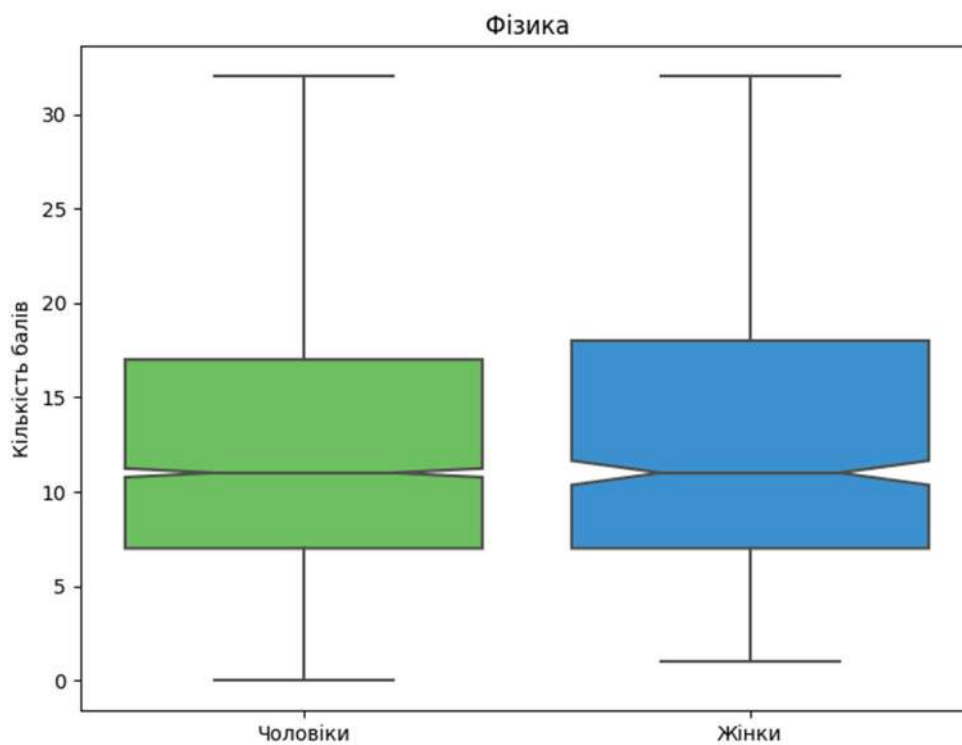


Рисунок 3.48 – Boxplot діаграма розподілу результатів (Фізика 2023 р.)

Для подальшого докладного кількісного порівняння характеристик розподілів було визначено їх квантилі.

Отримані результати наведено у таблиці 3.18.

Таблиця 3.18 – Квантилі розподілу результатів з фізики

2021 р.	10%	25%	75%	90%	Макс
Чоловіки	12	16	30	43	64
Жінки	11	15	33	46	64
2023 р.	10%	25%	75%	90%	Макс
Чоловіки	5	7	17	24	32
Жінки	5	7	18	24	32

Перевіримо відповідність даних нормальному розподілу за допомогою статистичних критеріїв та занесемо отримані дані до табл. 3.19.

Таблиця 3.19 – Результати перевірки нормальності

Вибірка	Критерії			
	Ліллєфорса		Андерсона-Дарлінга	
	D	p-value	A	p-value
Фізика, чоловіки 2021р.	0.14607	$< 2.2 \cdot 10^{-16}$	655.73	$< 2.2 \cdot 10^{-16}$
Фізика, жінки 2021р.	0.13782	$< 2.2 \cdot 10^{-16}$	97.745	$< 2.2 \cdot 10^{-16}$
Фізика, чоловіки 2023р	0.11182	$< 2.2 \cdot 10^{-16}$	84.766	$< 2.2 \cdot 10^{-16}$
Фізика, жінки 2023р.	0.12089	$< 2.2 \cdot 10^{-16}$	13.04	$< 2.2 \cdot 10^{-16}$

Критерії Ліллієфорса та Андерсона-Дарлінга показують, що розподіли результатів тестувань у чоловіків та жінок не відповідають нормальному розподілу. P-value у обох критеріях менше за 0.05, що свідчить про статистичну значущість відхилень від нормального розподілу. Тому для перевірки гіпотези про однорідність вибірок чоловіків та жінок треба використовувати непараметричні критерії.

Таблиця 3.20 – Результат перевірки однорідності показників з фізики

Вибірка	Критерії			
	Колмогорова-Смирнова		Уїлкоксона	
	D	p-value	W	p-value
2021 р.	0.054279	$7.727 \cdot 10^{-8}$	33293766	0.3862
2023 р.	0.021626	0.9351	1611101	0.797

3.5 Аналіз результатів тестувань з хімії

Таблиця 3.21 – Кількість учасників тестування з хімії (за 3 роки)

2021 р.		2022 р.		2023 р.	
Кількість учасників	9805	Кількість учасників	не проводилося	Кількість учасників	3218

Таблиця 3.22 – Процентне співвідношення чоловіків та жінок, які беруть участь у тестуванні

Стать	2021 р. (ЗНО)		2022 р. (НМТ)		2023 р. (НМТ)	
	кількість	% від загальної кількості учасників	кількість	% від загальної кількості учасників	кількість	% від загальної кількості учасників
Чоловіча	3592	36,63%	не проводилося		1085	33,72%
Жіноча	6213	63,37%			2133	66,28%

2021 р. (ЗНО) Хімія			
Чоловіки		Жінки	
Середнє	32,838808	Середнє	30,788991
Стандартна помилка	0,2802383	Стандартна помилка	0,1974721
Медіана	29	Медіана	28
Мода	19	Мода	15
Стандартне відхилення	16,795605	Стандартне відхилення	15,565258
Дисперсія вибірки	282,09236	Дисперсія вибірки	242,27726
Ексцес	-0,578962	Ексцес	-0,359848
Асиметричність	0,6174492	Асиметричність	0,6771393
Інтервал	72	Інтервал	69
Мінімум	2	Мінімум	5
Максимум	74	Максимум	74
Сума	117957	Сума	191292
Кількість	3592	Кількість	6213

Рисунок 3.49 – Показники описової статистики за 2021 р.

2023 р. (НМТ) Хімія			
Чоловіки		Жінки	
Середнє	19,184332	Середнє	17,634318
Стандартна помилка	0,3031527	Стандартна помилка	0,1955079
Медіана	17	Медіана	16
Мода	9	Мода	11
Стандартне відхилення	9,9856509	Стандартне відхилення	9,0294174
Дисперсія вибірки	99,713223	Дисперсія вибірки	81,530379
Ексцес	-1,05776	Ексцес	-0,589346
Асиметричність	0,41748	Асиметричність	0,6253793
Інтервал	40	Інтервал	40
Мінімум	0	Мінімум	0
Максимум	40	Максимум	40
Сума	20815	Сума	37614
Кількість	1085	Кількість	2133

Рисунок 3.50 – Показники описової статистики за 2023 р.

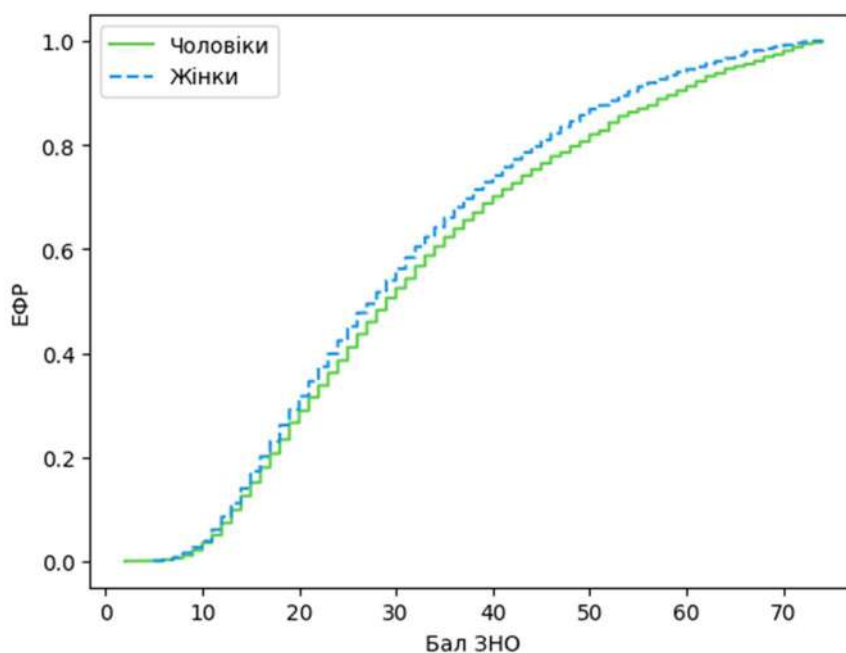


Рисунок 3.51 – Емпірична функція розподілу (Хімія 2021 р.)

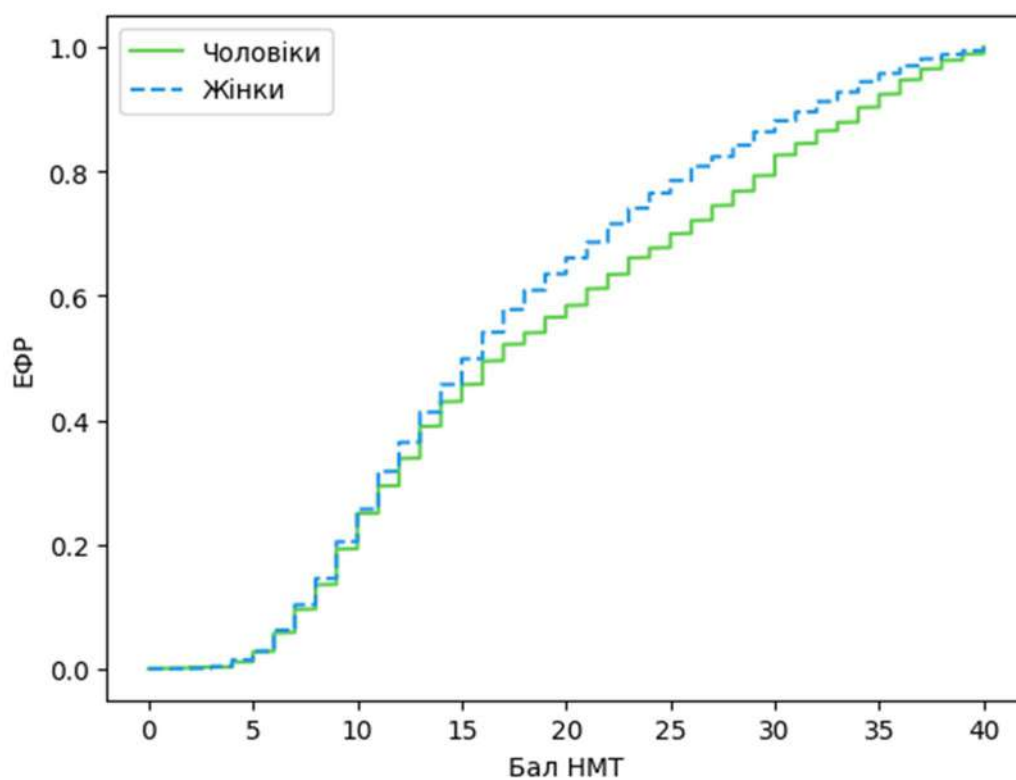


Рисунок 3.52 – Емпірична функція розподілу (Хімія 2023 р.)

Також для візуалізації даних побудуємо гістограми розподілу:

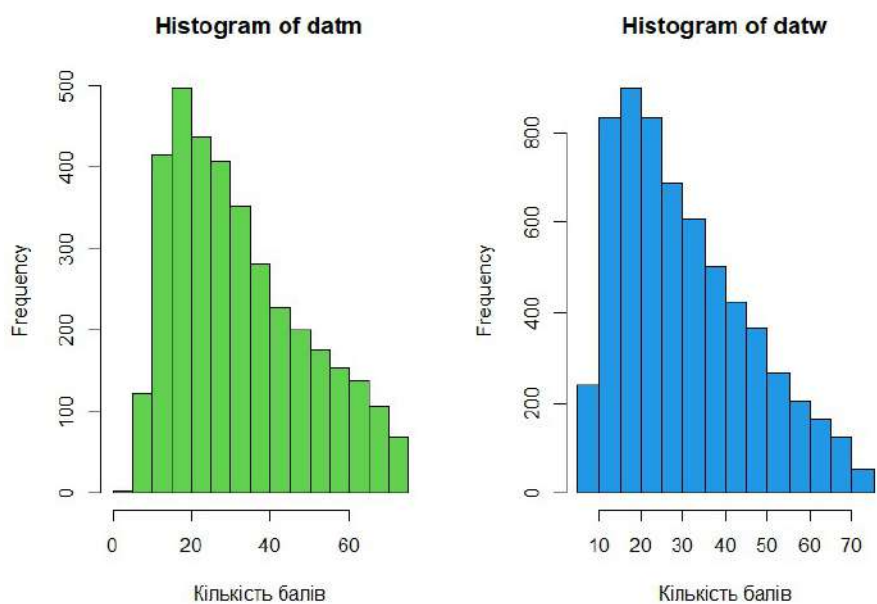


Рисунок 3.53 – Гістограма розподілу показників (Хімія 2021 р.)

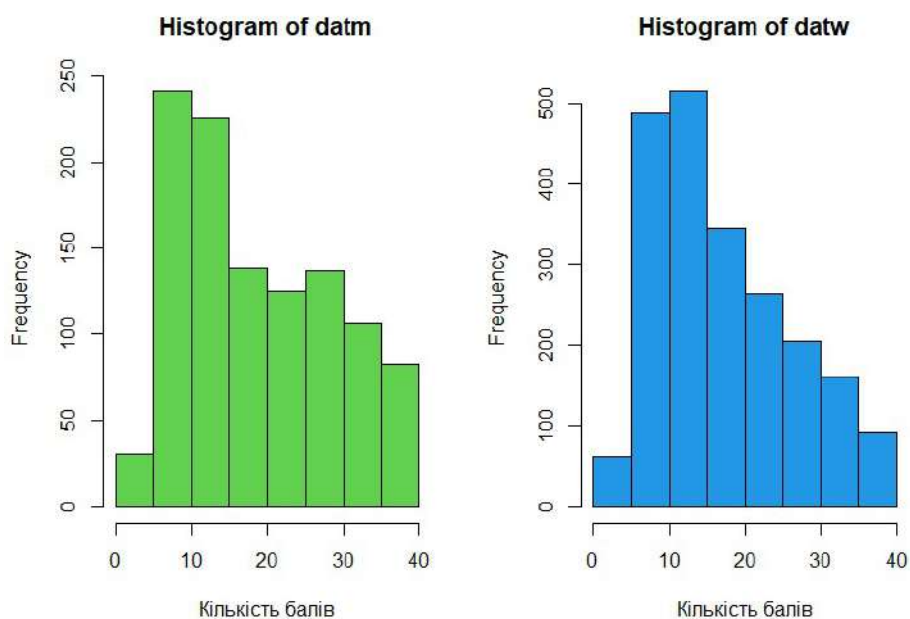


Рисунок 3.54 – Гістограма розподілу показників (Хімія 2023 р.)

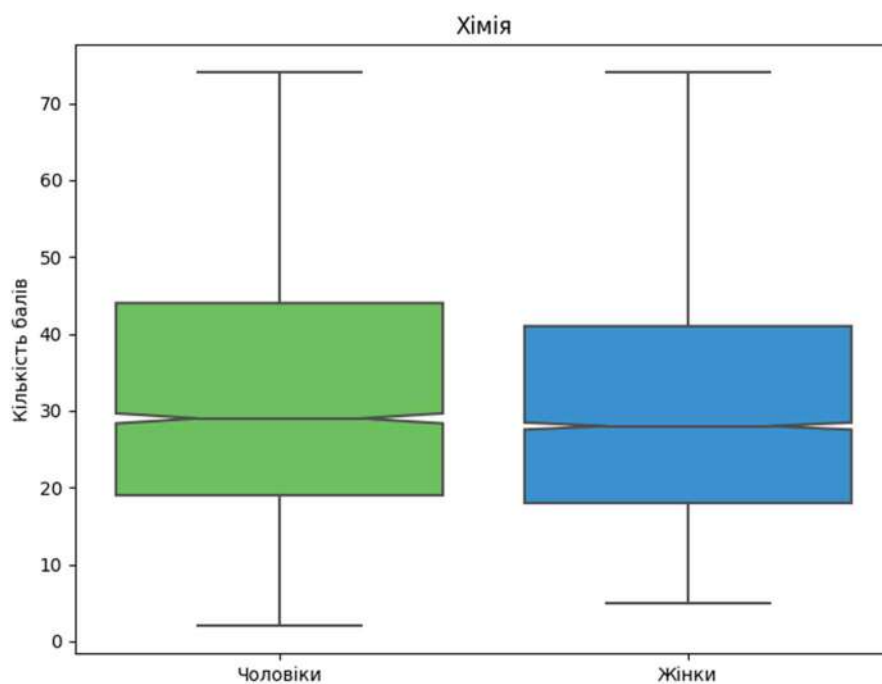


Рисунок 3.55 – Boxplot діаграма розподілу результатів чоловіків та жінок (Хімія 2021 р.)

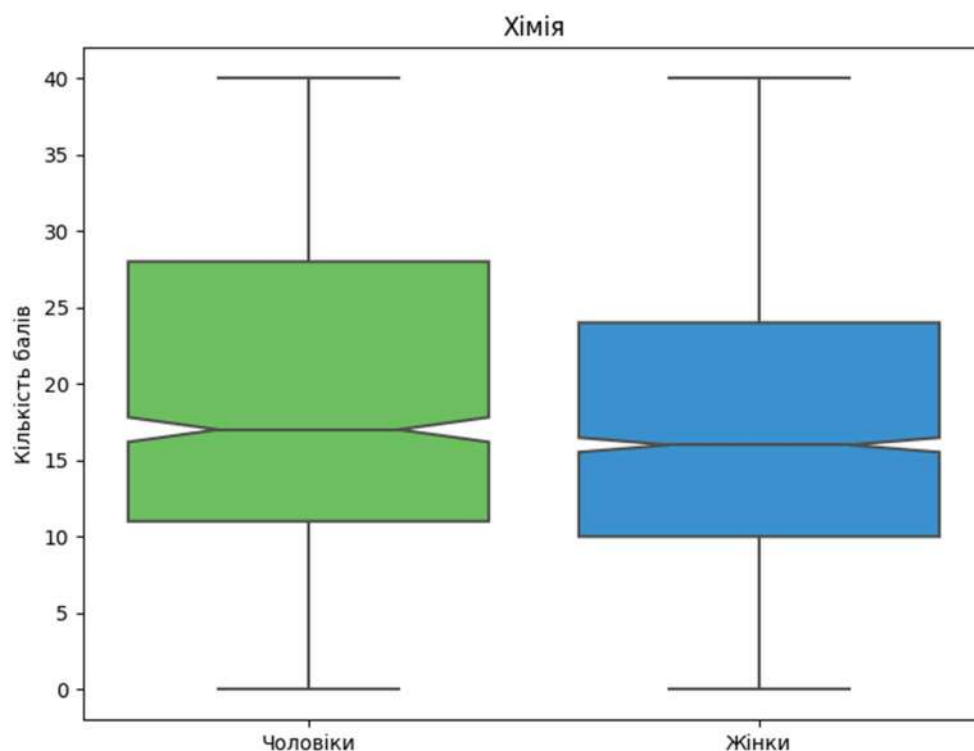


Рисунок 3.56 – Boxplot діаграма розподілу результатів чоловіків та жінок (Хімія 2023 р.)

Для подальшого докладного кількісного порівняння характеристик розподілів було визначено їх квантили.

Отримані результати наведено у таблиці 3.23.

Таблиця 3.23 – Квантили розподілу результатів з хімії

2021 р.	10%	25%	75%	90%	Макс
Чоловіки	14	19	44	59	74
Жінки	13	18	41	54	74
2023 р.	10%	25%	75%	90%	Макс
Чоловіки	8	11	28	34	40
Жінки	7	10	24	32	40

Перевіримо відповідність даних нормальному розподілу за допомогою статистичних критеріїв та занесемо отримані дані до табл. 3.24.

Таблиця 3.24 – Результати перевірки нормальності

Вибірка	Критерії			
	Ліллієфорса		Андерсона-Дарлінга	
	D	p-value	A	p-value
Хімія, чоловіки 2021р.	0.095963	$< 2.2 \cdot 10^{-16}$	62.402	$< 2.2 \cdot 10^{-16}$
Хімія, жінки 2021р.	0.097898	$< 2.2 \cdot 10^{-16}$	99.627	$< 2.2 \cdot 10^{-16}$
Хімія, чоловіки 2023р	0.12767	$< 2.2 \cdot 10^{-16}$	23.482	$< 2.2 \cdot 10^{-16}$
Хімія, жінки 2023р.	0.11344	$< 2.2 \cdot 10^{-16}$	41.366	$< 2.2 \cdot 10^{-16}$

Критерії Ліллієфорса та Андерсона-Дарлінга показують, що розподіли результатів тестувань у чоловіків та жінок не відповідають нормальному розподілу. Р-value у обох критеріях менше за 0.05, що свідчить про статистичну значущість відхилень від нормального розподілу. Тому для перевірки гіпотези про однорідність вибірок чоловіків та жінок треба використовувати непараметричні критерії, для цього застосуємо критерій Колмогорова – Смирнова та критерій Уїлкоксона.

Таблиця 3.25 – Результат перевірки однорідності показників з хімії

Вибірка	Критерії			
	Колмогорова-Смирнова		Уїлкоксона	
	D	p-value	W	p-value
2021 р.	0.050633	$1.708 \cdot 10^{-5}$	11856106	$2.384 \cdot 10^{-7}$
2023 р.	0.088153	$2.797 \cdot 10^{-5}$	1246514	0.0003321

3.6 Аналіз результатів тестувань з біології

Таблиця 3.26 – Кількість учасників тестування з біології (за 3 роки)

2021 р.		2022 р.		2023 р.	
Кількість учасників	81411	Кількість учасників	не проводилося	Кількість учасників	36009

Таблиця 3.27 – Процентне співвідношення чоловіків та жінок, які беруть участь у тестуванні

Стать	2021 р. (ЗНО)		2022 р. (НМТ)		2023 р. (НМТ)	
	кількість	% від загальної кількості учасників	кількість	% від загальної кількості учасників	кількість	% від загальної кількості учасників
Чоловіча	29410	36,13%	не проводилося		11910	33,08%
Жіноча	52001	63,87%			24099	66,92%

2021 р. (ЗНО) Біологія			
Чоловіки		Жінки	
Середнє	40,607616	Середнє	43,585796
Стандартна помилка	0,0758165	Стандартна помилка	0,0572637
Медіана	39	Медіана	41
Мода	35	Мода	36
Стандартне відхилення	13,002026	Стандартне відхилення	13,05826
Дисперсія вибірки	169,05267	Дисперсія вибірки	170,51814
Ексцес	0,4433019	Ексцес	0,0341886
Асиметричність	0,7308834	Асиметричність	0,6804247
Інтервал	82	Інтервал	82
Мінімум	0	Мінімум	0
Максимум	82	Максимум	82
Сума	1194270	Сума	2266505
Кількість	29410	Кількість	52001

Рисунок 3.57 – Показники описової статистики за 2021 р.

2023 р. (НМТ) Біологія			
Чоловіки		Жінки	
Середнє	20,66927	Середнє	22,581144
Стандартна помилка	0,0763971	Стандартна помилка	0,058134
Медіана	19	Медіана	21
Мода	15	Мода	17
Стандартне відхилення	8,3374453	Стандартне відхилення	9,0246434
Дисперсія вибірки	69,512995	Дисперсія вибірки	81,444189
Ексцес	0,4140341	Ексцес	-0,298018
Асиметричність	0,9114952	Асиметричність	0,6643053
Інтервал	46	Інтервал	46
Мінімум	0	Мінімум	0
Максимум	46	Максимум	46
Сума	246171	Сума	544183
Кількість	11910	Кількість	24099

Рисунок 3.58 – Показники описової статистики за 2023 р.

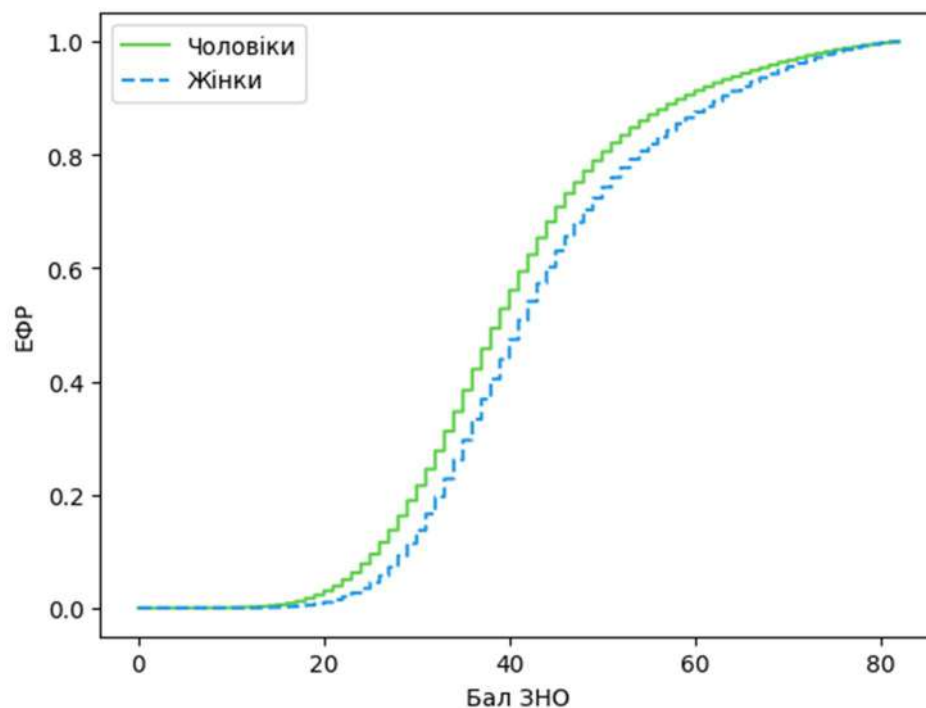


Рисунок 3.59 – Емпірична функція розподілу (Біологія 2021 р.)

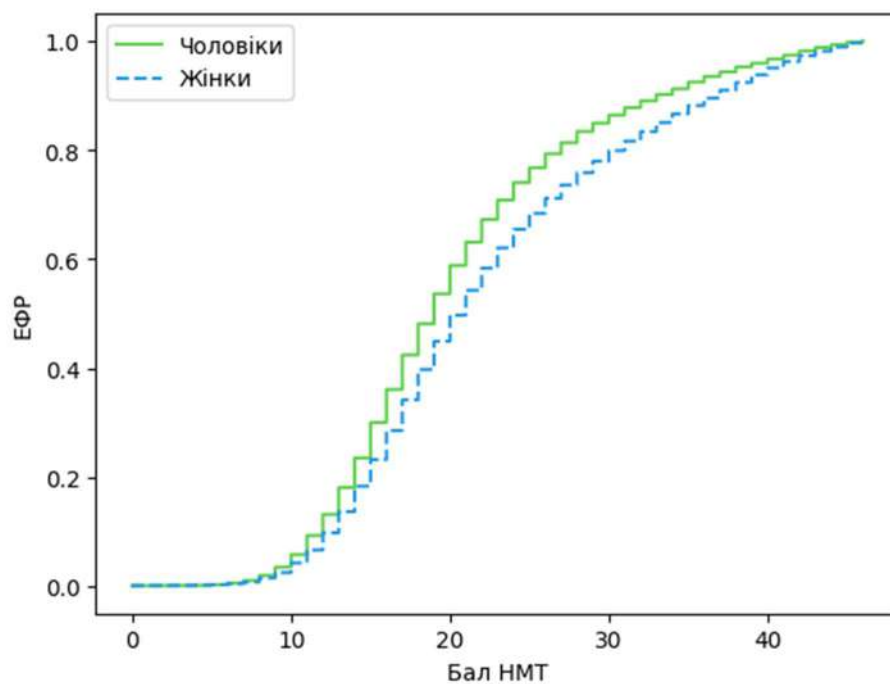


Рисунок 3.60 – Емпірична функція розподілу (Біологія 2023 р.)

Також для візуалізації даних побудуємо гістограми розподілу:

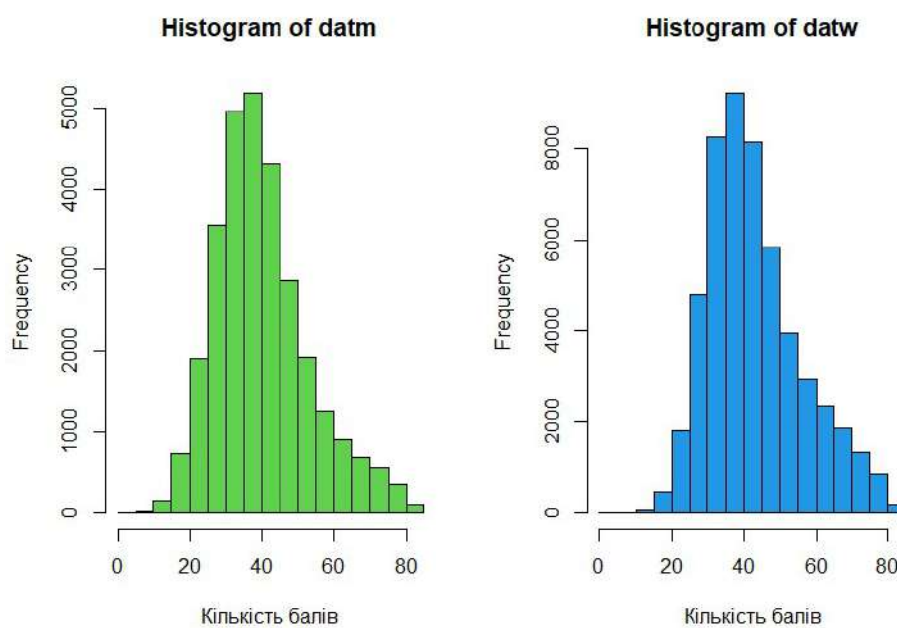


Рисунок 3.61 – Гістограма розподілу показників (Біологія 2021 р.)

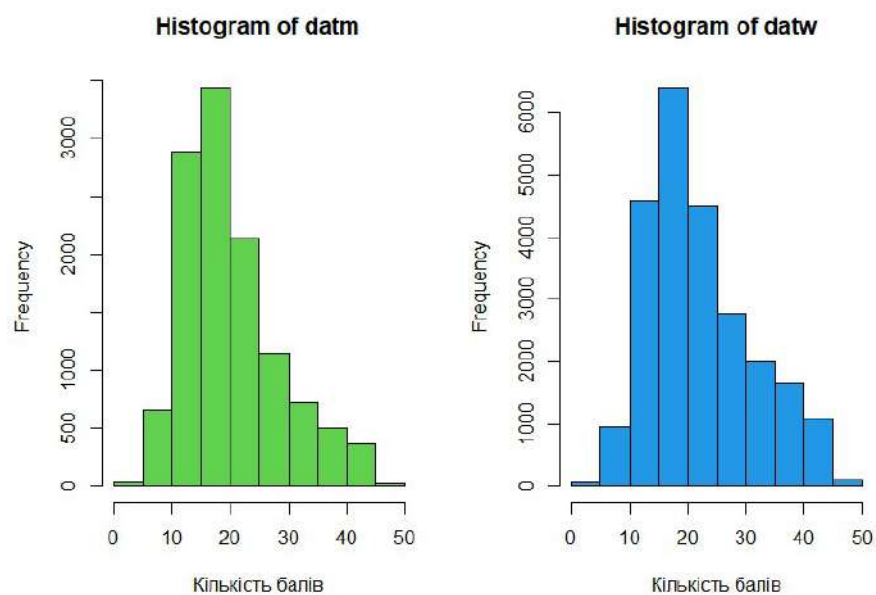


Рисунок 3.62 – Гістограма розподілу показників (Біологія 2023 р.)

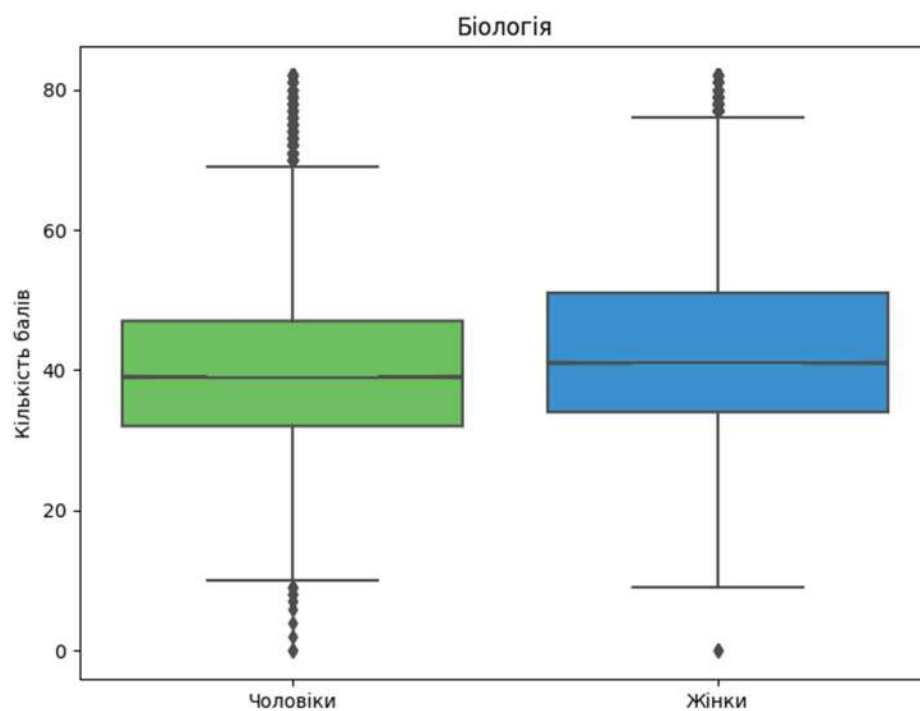


Рисунок 3.63 – Boxplot діаграма розподілу результатів чоловіків та жінок (Біологія 2021 р.)

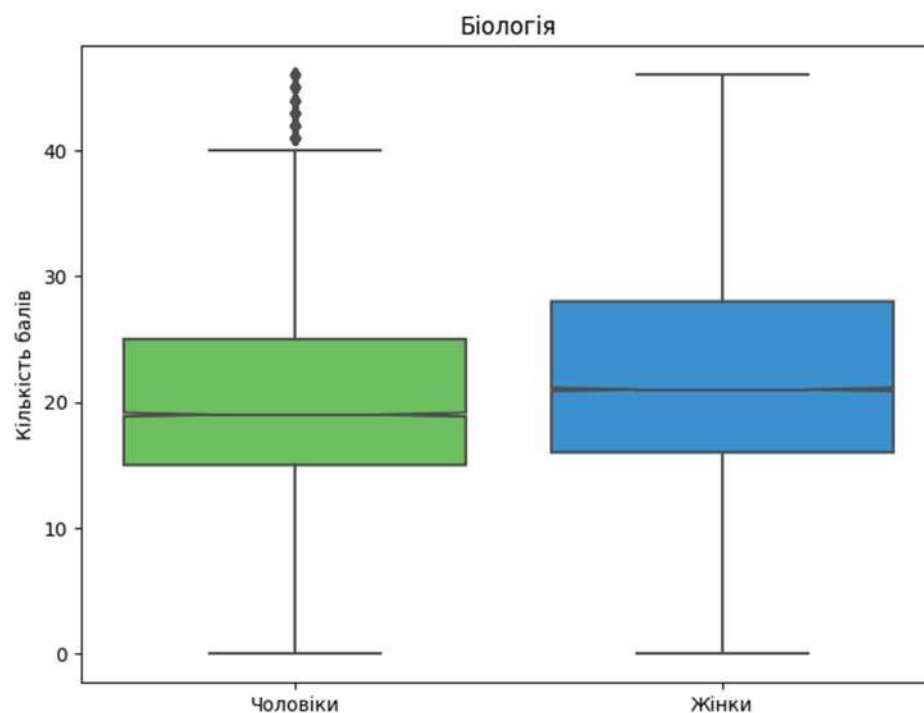


Рисунок 3.64 – Boxplot діаграма розподілу результатів чоловіків та жінок
(Біологія 2023 р.)

Для подальшого докладного кількісного порівняння характеристик розподілів було визначено їх квантили.

Отримані результати наведено у таблиці 3.28.

Таблиця 3.28 – Квантили розподілу даних результатів з біології

2021 р.	10%	25%	75%	90%	Макс
Чоловіки	26	32	47	59	82
Жінки	29	34	51	63	82
2023 р.	10%	25%	75%	90%	Макс
Чоловіки	12	15	25	33	46
Жінки	13	16	28	37	46

Перевіримо відповідність даних нормальному розподілу за допомогою більш строгих статистичних критеріїв та занесемо отримані дані до табл. 3.29.

Таблиця 3.29 – Результати перевірки нормальності

Вибірка	Критерії			
	Ліллієфорса		Андерсона-Дарлінга	
	D	p-value	A	p-value
Біологія, чоловіки 2021р.	0.081434	$< 2.2 \cdot 10^{-1}$	299.06	$< 2.2 \cdot 10^{-16}$
Біологія, жінки 2021р.	0.090245	$< 2.2 \cdot 10^{-16}$	640.43	$< 2.2 \cdot 10^{-16}$
Біологія, чоловіки 2023р	0.1199	$< 2.2 \cdot 10^{-16}$	244.64	$< 2.2 \cdot 10^{-16}$
Біологія, жінки 2023р.	0.11205	$< 2.2 \cdot 10^{-16}$	417.48	$< 2.2 \cdot 10^{-16}$

Критерії Ліллієфорса та Андерсона-Дарлінга показують, що розподіли результатів тестувань у чоловіків та жінок не відповідають нормальному розподілу. Р-value у обох критеріях менше за 0.05, що свідчить про статистичну значущість відхилень від нормального розподілу. Тому для перевірки гіпотези про однорідність вибірок чоловіків та жінок треба використовувати непараметричні критерії, для цього застосуємо критерій Колмогорова – Смирнова та критерій Уїлкоксона.

Таблиця 3.30 – Результат перевірки однорідності показників з біології

Вибірка	Критерії			
	Колмогорова-Смирнова		Уїлкоксона	
	D	p-value	W	p-value
2021 р.	0.089512	$< 2.2 \cdot 10^{-16}$	662896263	$< 2.2 \cdot 10^{-16}$
2023 р.	0.091001	$< 2.2 \cdot 10^{-16}$	125643131	$< 2.2 \cdot 10^{-16}$

3.7 Аналіз результатів тестувань з географії

Таблиця 3.31 – Кількість учасників тестування з географії (за 3 роки)

2021 р.		2022 р.		2023 р.	
Кількість учасників	113116	Кількість учасників	не проводилося	Кількість учасників	не проводилося

Таблиця 3.32 – Процентне співвідношення чоловіків та жінок, які беруть участь у тестуванні

Стать	2021 р. (ЗНО)		2022 р. (НМТ)		2023 р. (НМТ)	
	кількість	% від загальної кількості учасників	кількість	% від загальної кількості учасників	кількість	% від загальної кількості учасників
Чоловіча	58548	51,76%	не проводилося		не проводилося	
Жіноча	54568	48,24%				

2021 р. (ЗНО) Географія			
Чоловіки		Жінки	
Середнє	44,145317	Середнє	45,335783
Стандартна помилка	0,0614946	Стандартна помилка	0,063587
Медіана	43	Медіана	43
Мода	41	Мода	40
Стандартне відхилення	14,879671	Стандартне відхилення	14,853799
Дисперсія вибірки	221,40461	Дисперсія вибірки	220,63534
Ексцес	-0,324731	Ексцес	-0,374476
Асиметричність	0,3978363	Асиметричність	0,4584847
Інтервал	89	Інтервал	89
Мінімум	0	Мінімум	0
Максимум	89	Максимум	89
Сума	2584620	Сума	2473883
Кількість	58548	Кількість	54568

Рисунок 3.65 – Показники описової статистики за 2021 р.

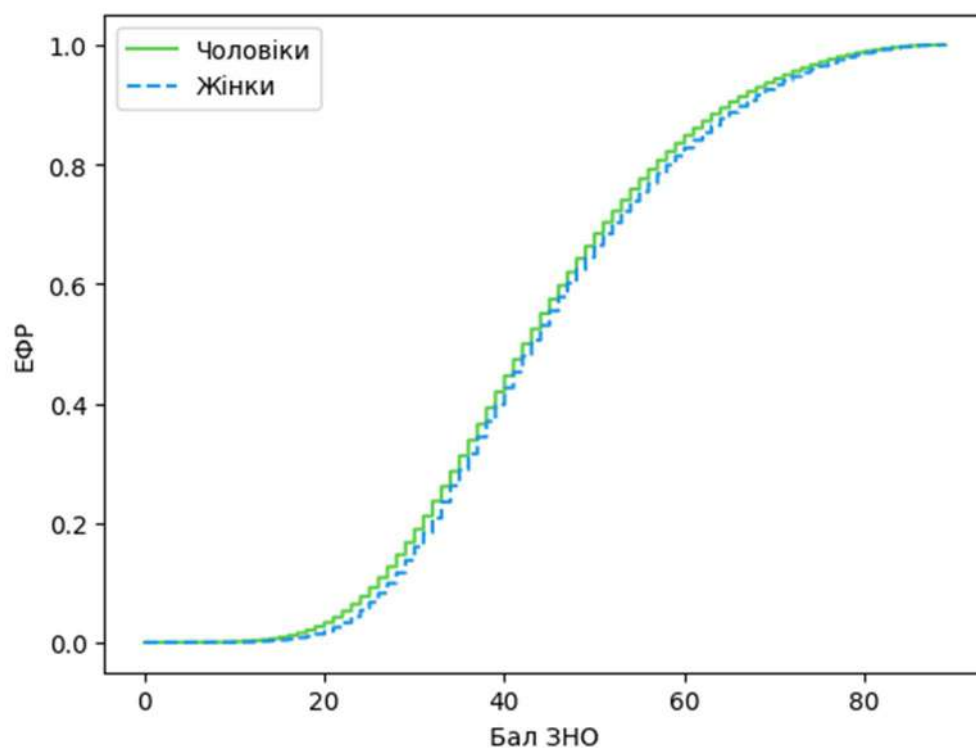


Рисунок 3.66 – Емпірична функція розподілу (Географія 2021 р.)

Також для візуалізації даних побудуємо гістограми розподілу:

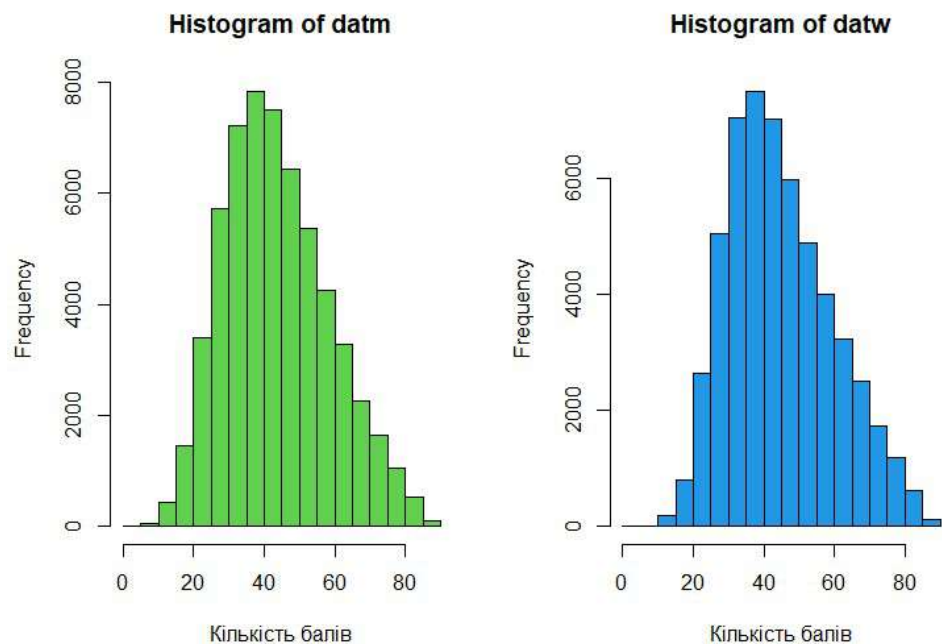


Рисунок 3.67 – Гістограма розподілу показників (Географія 2021 р.)

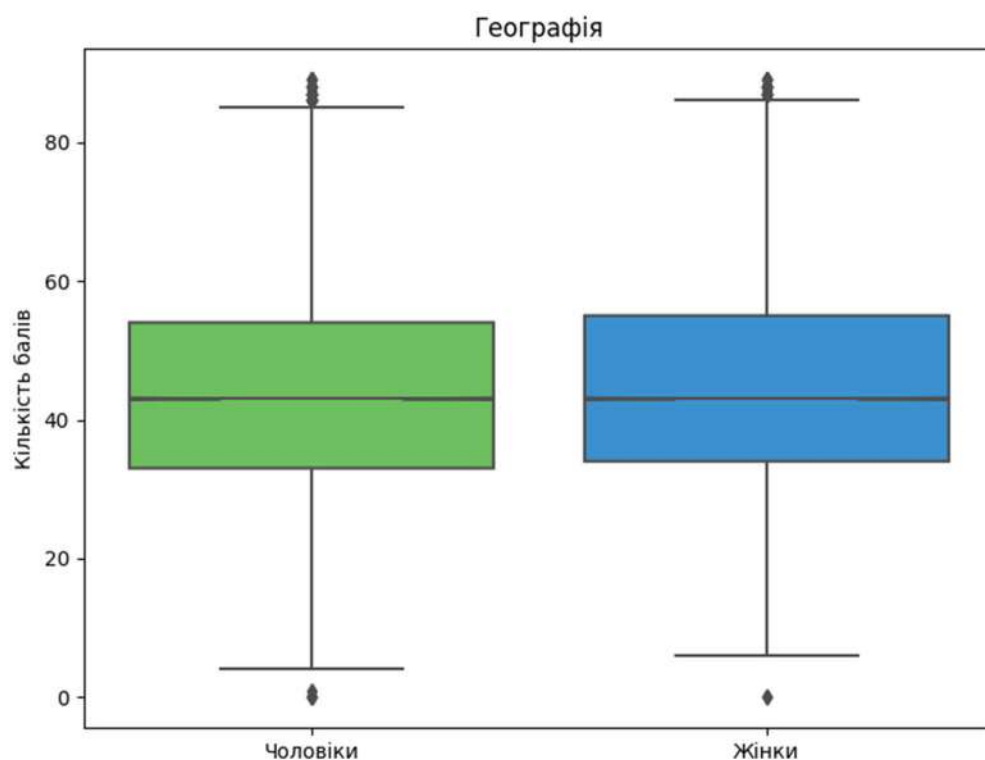


Рисунок 3.68 – Boxplot діаграма розподілу результатів чоловіків та жінок (Географія 2021 р.)

Для подальшого докладного кількісного порівняння характеристик розподілів визначимо їх квантилі.

Отримані результати наведено у таблиці 3.33.

Таблиця 3.33 – Квантилі розподілу результатів з географії

2021 р.	10%	25%	75%	90%	Макс
Чоловіки	26	33	54	65	89
Жінки	28	34	55	67	89

Перевіримо відповідність даних нормальному розподілу за допомогою статистичних критеріїв та занесемо отримані дані до табл. 3.34.

Таблиця 3.34 – Результати перевірки нормальності

Вибірка	Критерії			
	Лілієфорса		Андерсона-Дарлінга	
	D	p-value	A	p-value
Географія, чоловіки 2021р.	0.057232	$< 2.2 \cdot 10^{-16}$	258.22	$< 2.2 \cdot 10^{-16}$
Географія, жінки 2021р.	0.068176	$< 2.2 \cdot 10^{-16}$	369.71	$< 2.2 \cdot 10^{-16}$

Критерії Лілієфорса та Андерсона-Дарлінга показують, що розподіли результатів тестувань у чоловіків та жінок не відповідають нормальному розподілу. P-value у обох критеріях менше за 0.05, що свідчить про статистичну значущість відхилень від нормального розподілу. Тому для перевірки гіпотези про однорідність вибірок чоловіків та жінок треба використовувати непараметричні критерії, для цього застосуємо критерій Колмогорова – Смирнова та критерій Уїлкоксона.

Таблиця 3.35 – Результат перевірки однорідності показників з географії

Вибірка	Критерії			
	Колмогорова-Смирнова		Уїлкоксона	
	D	p-value	W	p-value
2021 р.	0.030197	$< 2.2 \cdot 10^{-16}$	$1.532 \cdot 10^9$	$< 2.2 \cdot 10^{-16}$

3.8 Аналіз результатів тестувань з англійської мови

Таблиця 3.36 – Кількість учасників тестування з англійської мови (за 3 роки)

2021 р.		2022 р.		2023 р.	
Кількість учасників	127714	Кількість учасників	не проводилося	Кількість учасників	80247

Таблиця 3.37 – Процентне співвідношення чоловіків та жінок, які беруть участь у тестуванні

Стать	2021 р. (ЗНО)		2022 р. (НМТ)		2023 р. (НМТ)	
	кількість	% від загальної кількості учасників	кількість	% від загальної кількості учасників	кількість	% від загальної кількості учасників
Чоловіча	62836	49,20%	не проводилося		41093	51,21%
Жіноча	64878	50,80%			39154	48,79%

2021 р. (ЗНО) Англійська мова			
Чоловіки		Жінки	
Середнє	38,426555	Середнє	40,882518
Стандартна помилка	0,0659166	Стандартна помилка	0,0663305
Медіана	38	Медіана	41
Мода	18	Мода	59
Стандартне відхилення	16,523397	Стандартне відхилення	16,895137
Дисперсія вибірки	273,02265	Дисперсія вибірки	285,44567
Ексцес	-1,166375	Ексцес	-1,184108
Асиметричність	0,1090841	Асиметричність	-0,008608
Інтервал	72	Інтервал	72
Мінімум	0	Мінімум	0
Максимум	72	Максимум	72
Сума	2414571	Сума	2652376
Кількість	62836	Кількість	64878

Рисунок 3.69 – Показники описової статистики за 2021 р.

2023 р. (НМТ) Англійська мова			
Чоловіки		Жінки	
Середнє	18,120483	Середнє	19,422358
Стандартна помилка	0,0374035	Стандартна помилка	0,0380154
Медіана	18	Медіана	20
Мода	12	Мода	28
Стандартне відхилення	7,5822087	Стандартне відхилення	7,5222492
Дисперсія вибірки	57,489888	Дисперсія вибірки	56,584233
Екссес	-1,124797	Екссес	-1,072695
Асиметричність	-0,032847	Асиметричність	-0,214404
Інтервал	32	Інтервал	32
Мінімум	0	Мінімум	0
Максимум	32	Максимум	32
Сума	744625	Сума	760463
Кількість	41093	Кількість	39154

Рисунок 3.70 – Показники описової статистики за 2023 р.

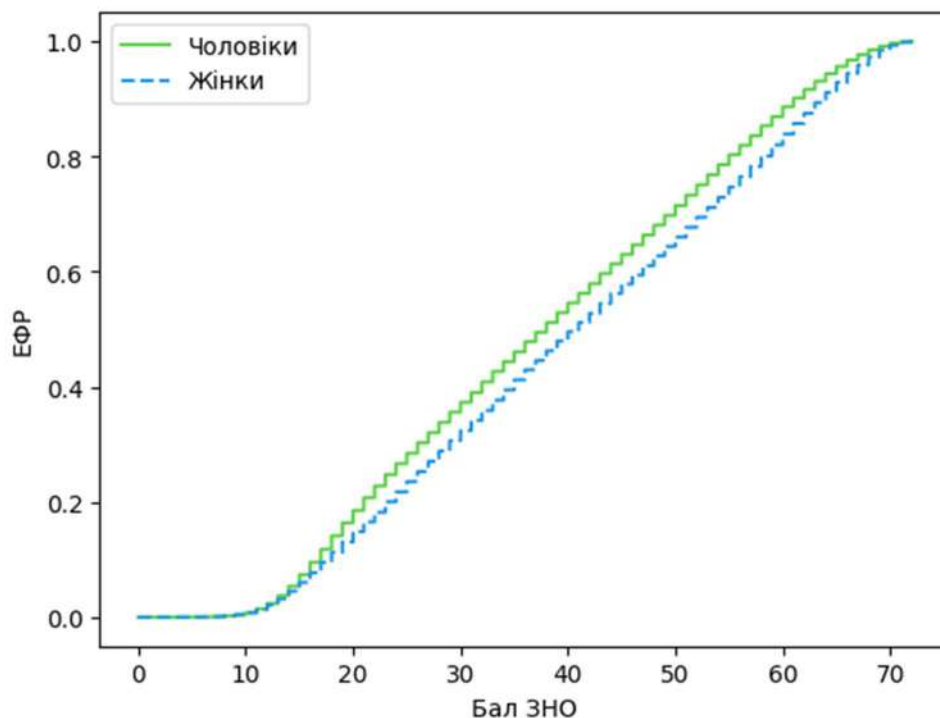


Рисунок 3.71 – Емпірична функція розподілу (Англійська мова 2021р.)

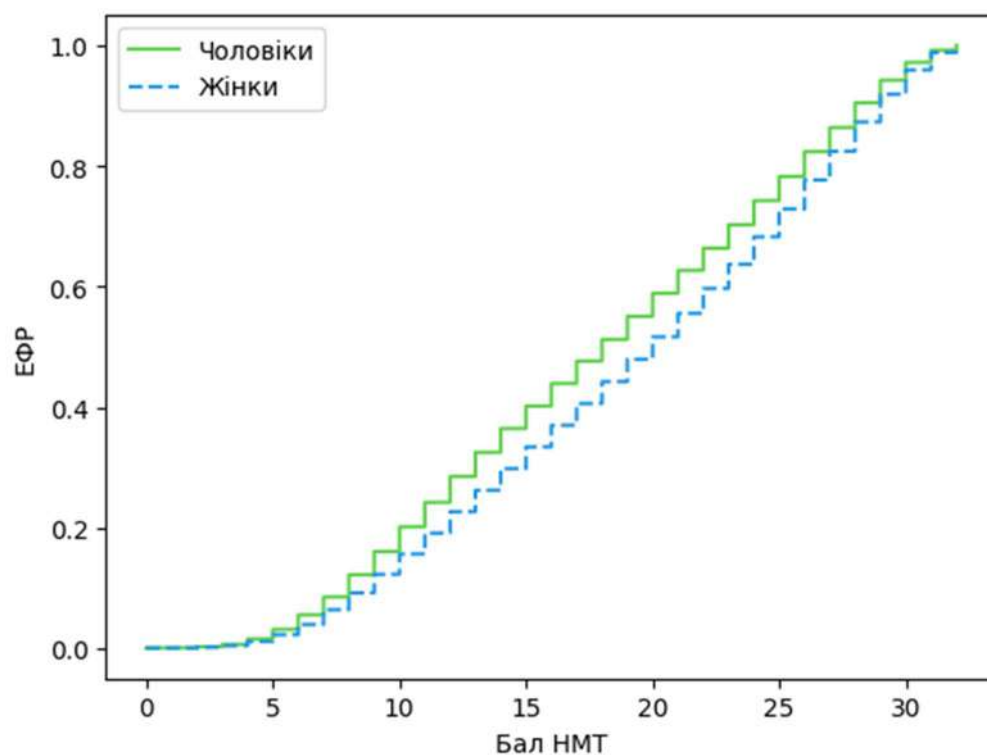


Рисунок 3.72 – Емпірична функція розподілу (Англійська мова 2023 р.)

Також для візуалізації даних побудуємо гістограми розподілу:

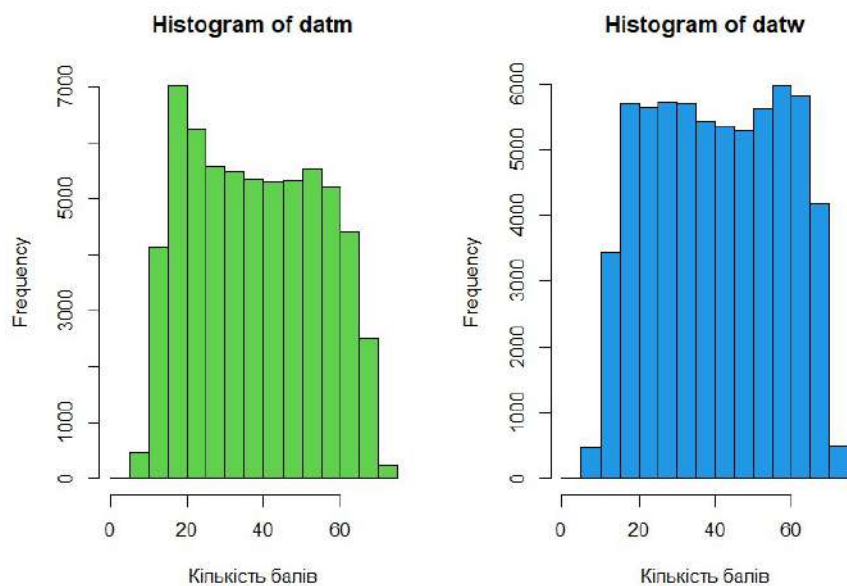


Рисунок 3.73 – Гістограма розподілу показників (Англійська мова 2021 р.)

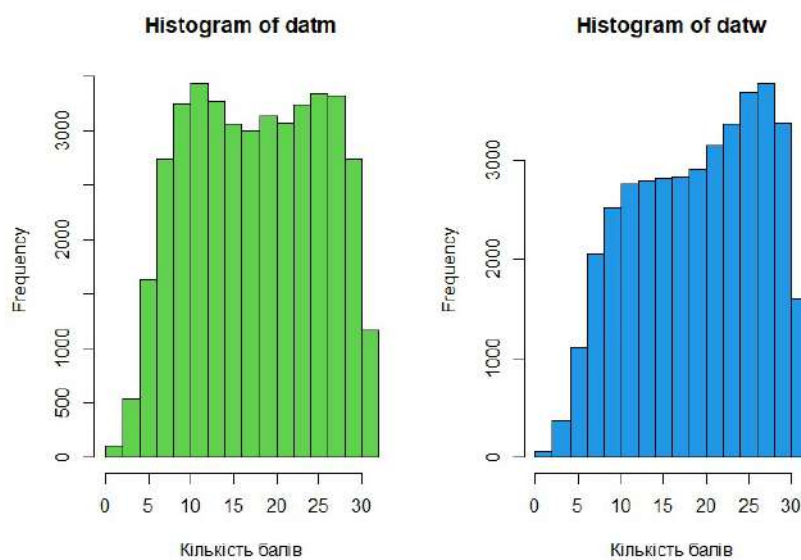


Рисунок 3.74 – Гістограма розподілу показників (Англійська мова 2023 р.)

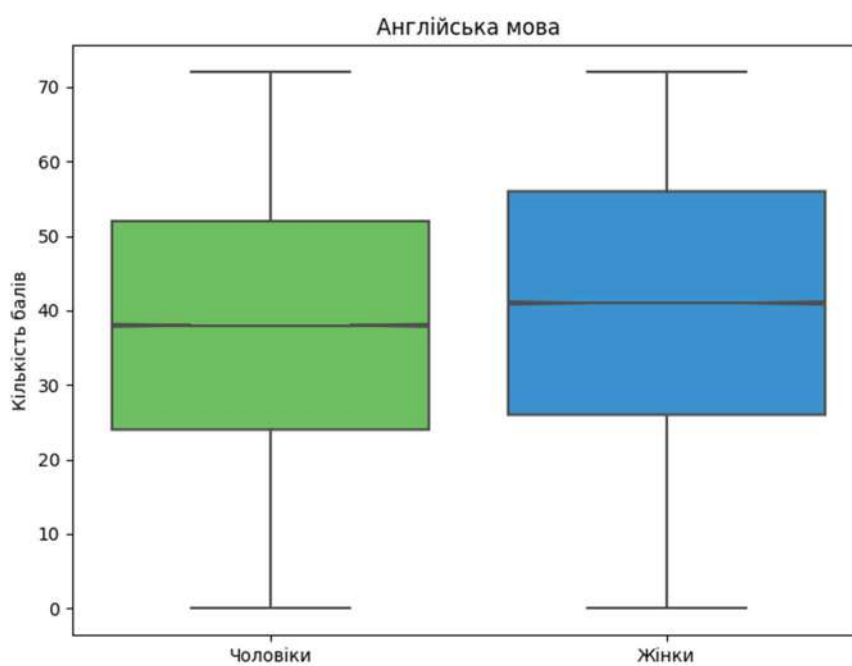


Рисунок 3.75 – Boxplot діаграма розподілу результатів чоловіків та жінок (Англійська мова 2021 р.)

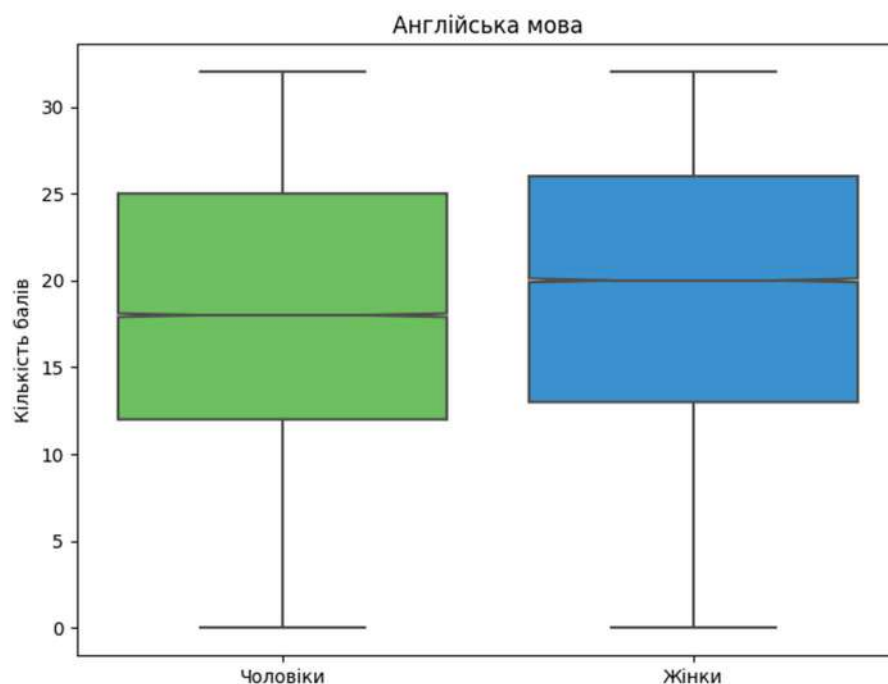


Рисунок 3.76 – Boxplot діаграма розподілу результатів чоловіків та жінок (Англійська мова 2023 р.)

Для подальшого докладного кількісного порівняння характеристик розподілів було визначено їх квантилі.

Отримані результати наведено у таблиці 3.38.

Таблиця 3.38 – Квантилі розподілу даних результатів з англійської мови

2021 р.	10%	25%	75%	90%	Макс
Чоловіки	17	24	52	61	72
Жінки	18	26	56	64	72
2023 р.	10%	25%	75%	90%	Макс
Чоловіки	8	12	25	28	32
Жінки	9	13	26	29	32

Перевіримо відповідність даних нормальному розподілу за допомогою статистичних критеріїв та занесемо отримані дані до табл. 3.39.

Таблиця 3.39 – Результати перевірки нормальності

Вибірка	Критерії			
	Ліллієфорса		Андерсона-Дарлінга	
	D	p-value	A	p-value
Англ.мова, чоловіки 2021р.	0.076669	$< 2.2 \cdot 10^{-16}$	725.69	$< 2.2 \cdot 10^{-1}$
Англ.мова, жінки 2021р.	0.070049	$< 2.2 \cdot 10^{-16}$	717.4	$< 2.2 \cdot 10^{-16}$
Англ.мова, чоловіки 2023р	0.078698	$< 2.2 \cdot 10^{-16}$	423.3	$< 2.2 \cdot 10^{-16}$
Англ.мова, жінки 2023р.	0.091667	$< 2.2 \cdot 10^{-16}$	456.15	$< 2.2 \cdot 10^{-16}$

Критерії Ліллієфорса та Андерсона-Дарлінга показують, що розподіли результатів тестувань у чоловіків та жінок не відповідають нормальному розподілу. P-value у обох критеріях менше за 0.05, що свідчить про статистичну значущість відхилень від нормального розподілу. Тому для перевірки гіпотези про однорідність вибірок чоловіків та жінок треба використовувати непараметричні критерії, для цього застосуємо критерій Колмогорова – Смирнова та критерій Уїлкоксона.

Таблиця 3.40 – Результат перевірки однорідності показників з англійської мови

Вибірка	Критерії			
	Колмогорова-Смирнова		Уїлкоксона	
	D	p-value	W	p-value
2021 р.	0.056653	$< 2.2 \cdot 10^{-16}$	1866408422	$< 2.2 \cdot 10^{-16}$
2023 р.	0.072248	$< 2.2 \cdot 10^{-16}$	724901913	$< 2.2 \cdot 10^{-16}$

3.9 Аналіз результатів тестувань з французької мови

Таблиця 3.41 – Кількість учасників тестування з французької мови (за 3 роки)

2021 р.		2022 р.		2023 р.	
Кількість учасників	355	Кількість учасників	не проводилося	Кількість учасників	273

Таблиця 3.42 – Процентне співвідношення чоловіків та жінок, які беруть участь у тестуванні

Стать	2021 р. (ЗНО)		2022 р. (НМТ)		2023 р. (НМТ)	
	кількість	% від загальної кількості учасників	кількість	% від загальної кількості учасників	кількість	% від загальної кількості учасників
Чоловіча	99	27,89%	не проводилося		86	31,50%
Жіноча	256	72,11%			187	68,50%

2021 р. (ЗНО) Французька мова			
Чоловіки		Жінки	
Середнє	42,848485	Середнє	43,941406
Стандартна помилка	1,65776	Стандартна помилка	0,9894211
Медіана	43	Медіана	47
Мода	62	Мода	58
Стандартне відхилення	16,494503	Стандартне відхилення	15,830737
Дисперсія вибірки	272,06865	Дисперсія вибірки	250,61224
Ексцес	-1,03322	Ексцес	-0,990064
Асиметричність	-0,287376	Асиметричність	-0,404141
Інтервал	58	Інтервал	58
Мінімум	11	Мінімум	11
Максимум	69	Максимум	69
Сума	4242	Сума	11249
Кількість	99	Кількість	256

Рисунок 3.77 – Показники описової статистики за 2021 р.

2023 р. (НМТ) Французька мова			
Чоловіки		Жінки	
Середнє	15,732558	Середнє	17
Стандартна помилка	0,7790739	Стандартна помилка	0,5399885
Медіана	14	Медіана	16
Мода	11	Мода	12
Стандартне відхилення	7,2248337	Стандартне відхилення	7,384232
Дисперсія вибірки	52,198222	Дисперсія вибірки	54,526882
Ексцес	-0,781211	Ексцес	-1,116001
Асиметричність	0,5621984	Асиметричність	0,1418817
Інтервал	26	Інтервал	28
Мінімум	6	Мінімум	3
Максимум	32	Максимум	31
Сума	1353	Сума	3179
Кількість	86	Кількість	187

Рисунок 3.78 – Показники описової статистики за 2023 р.

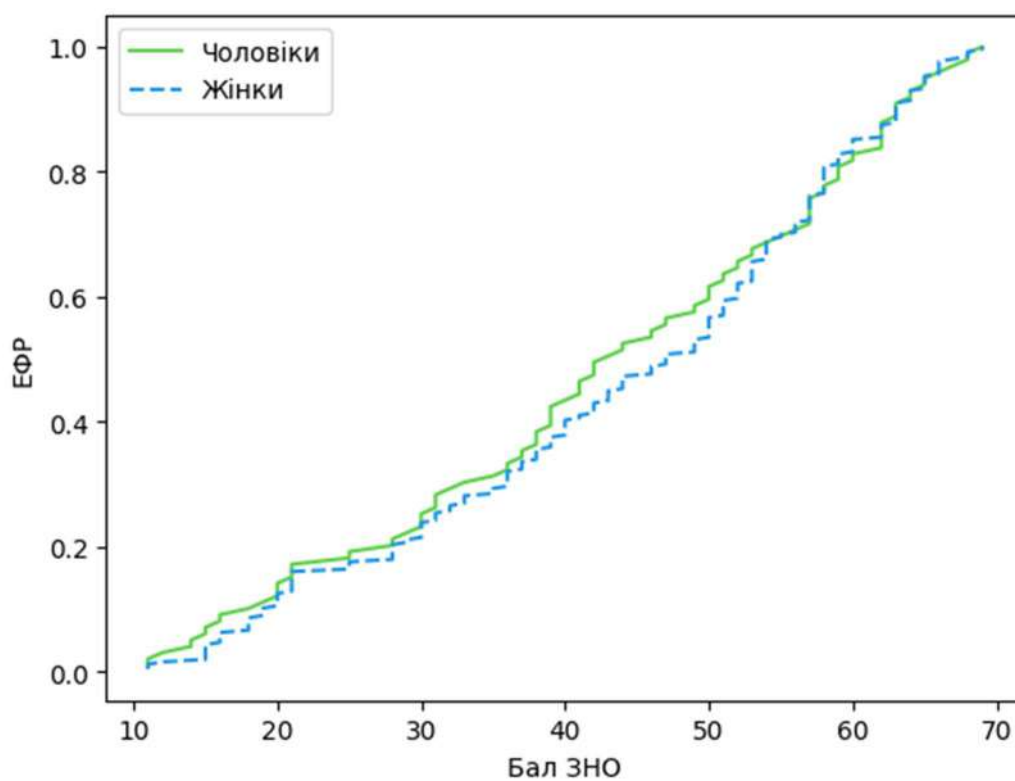


Рисунок 3.79 – Емпірична функція розподілу (Французька мова 2021 р.)

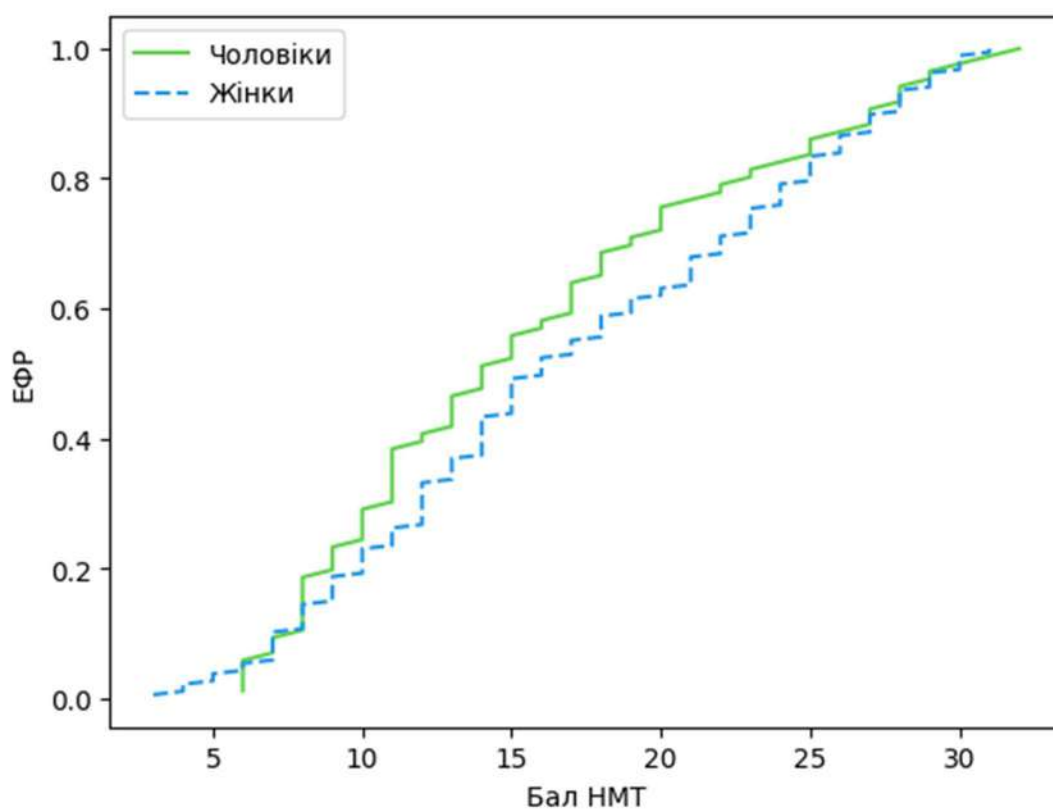


Рисунок 3.80 – Емпірична функція розподілу (Французька мова 2023 р.)

Також для візуалізації даних побудуємо гістограми розподілу:

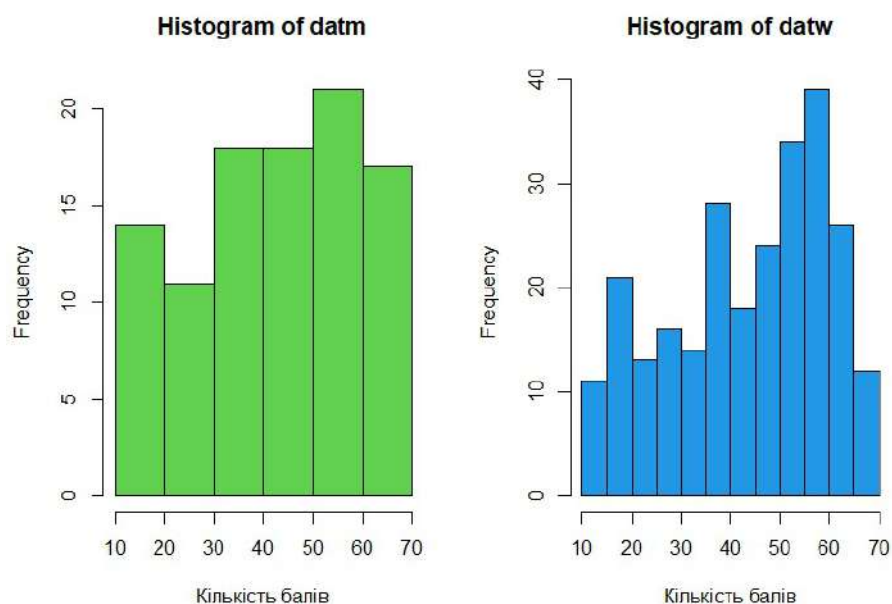


Рисунок 3.81 – Гістограма розподілу показників (Французька мова 2021 р.)

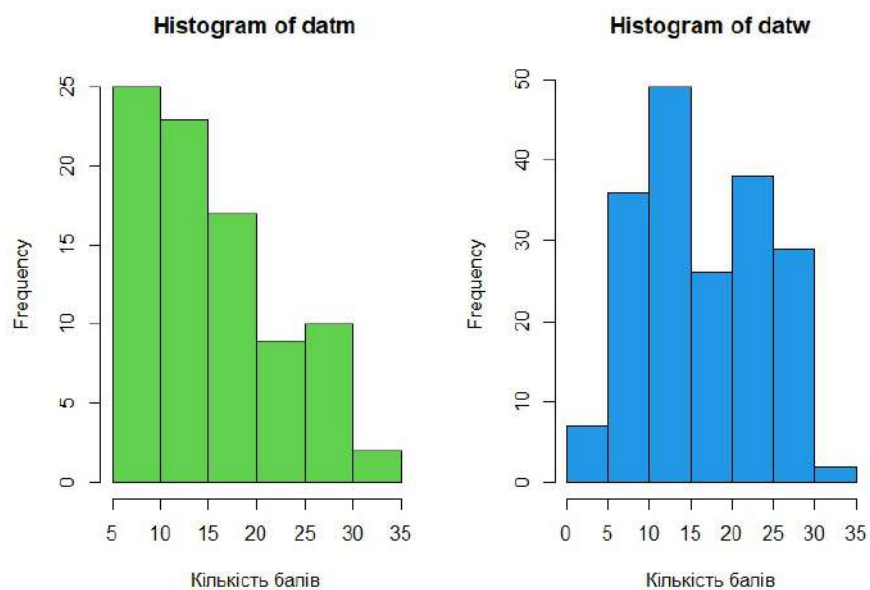


Рисунок 3.82 – Гістограма розподілу показників (Французька мова 2023 р.)

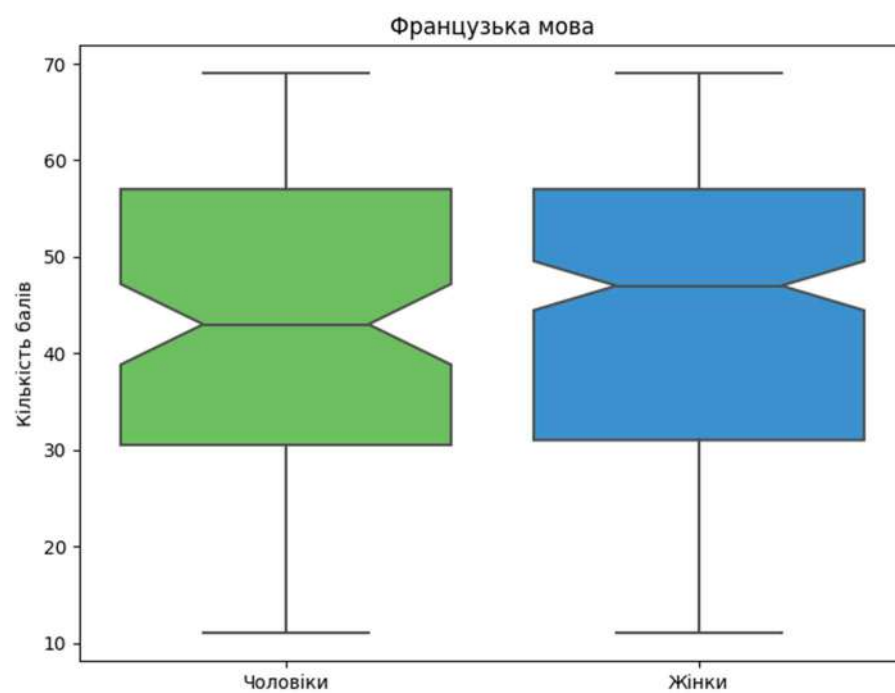


Рисунок 3.83 – Boxplot діаграма розподілу результатів чоловіків та жінок (Французька мова 2021р.)

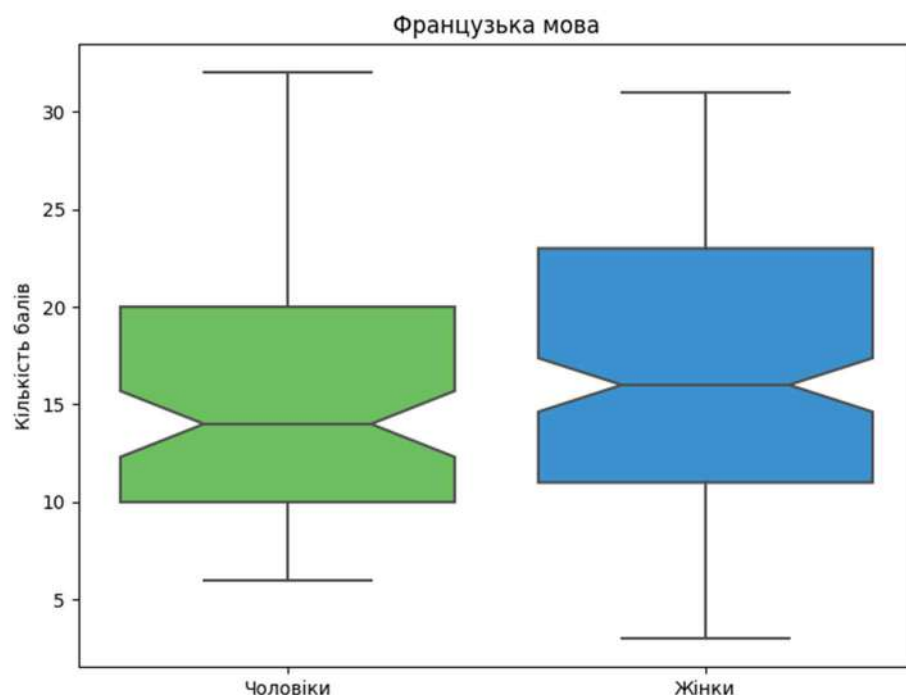


Рисунок 3.84 – Boxplot діаграма розподілу результатів чоловіків та жінок (Французька мова 2023р.)

Для подальшого докладного кількісного порівняння характеристик розподілів було визначено їх квантілі.

Отримані результати наведено у таблиці 3.43.

Таблиця 3.43 – Квантілі розподілу результатів з французької мови

2021 р.	10%	25%	75%	90%	Макс
Чоловіки	18.8	30.5	57	63	69
Жінки	19.5	31	57	63	69
2023 р.	10%	25%	75%	90%	Макс
Чоловіки	8	10	20	27	32
Жінки	7.6	11	23	27.4	31

Перевіримо відповідність даних нормальному розподілу за допомогою статистичних критеріїв та занесемо отримані дані до табл. 3.44.

Таблиця 3.44 – Результати перевірки нормальності

Вибірка	Критерії			
	Ліллієфорса		Андерсона-Дарлінга	
	D	p-value	A	p-value
Фр.мова, чоловіки 2021р.	0.09747	0.02145	1.2593	0.00268
Фр.мова, жінки 2021р.	0.11778	$3.544 \cdot 10^{-9}$	4.4066	$5.776 \cdot 10^{-11}$
Фр.мова, чоловіки 2023р	0.1275	0.001475	1.7791	0.0001371
Фр.мова, жінки 2023р.	0.098724	0.0001356	2.2695	$< 8.99 \cdot 10^{-6}$

Критерії Ліллієфорса та Андерсона-Дарлінга показують, що розподіли результатів тестувань у чоловіків та жінок не відповідають нормальному розподілу. P-value у обох критеріях менше за 0.05, що свідчить про статистичну значущість відхилень від нормального розподілу. Тому для перевірки гіпотези про однорідність вибірок чоловіків та жінок треба використовувати непараметричні критерії, для цього застосуємо критерій Колмогорова – Смирнова та критерій Уїлкоксона.

Таблиця 3.45 – Результат перевірки однорідності показників з французької мови

Вибірка	Критерії			
	Колмогорова-Смирнова		Уїлкоксона	
	D	p-value	W	p-value
2021 р.	0.065262	0.9214	12215	0.5984
2023 р.	0.1248	0.318	7187	0.1586

3.10 Аналіз результатів тестувань з німецької мови

Таблиця 3.46 – Кількість учасників тестування з німецької мови (за 3 роки)

2021 р.		2022 р.		2023 р.	
Кількість учасників	1825	Кількість учасників	не проводилося	Кількість учасників	1913

Таблиця 3.47 – Процентне співвідношення чоловіків та жінок, які беруть участь у тестуванні

Стать	2021 р. (ЗНО)		2022 р. (НМТ)		2023 р. (НМТ)	
	кількість	% від загальної кількості учасників	кількість	% від загальної кількості учасників	кількість	% від загальної кількості учасників
Чоловіча	627	34,36%	не проводилося		794	41,51%
Жіноча	1198	65,64%			1119	58,49%

2021 р. (ЗНО) Німецька мова			
Чоловіки		Жінки	
Середнє	39,746411	Середнє	41,980801
Стандартна помилка	0,6859681	Стандартна помилка	0,4891936
Медіана	41	Медіана	43
Мода	60	Мода	60
Стандартне відхилення	17,176619	Стандартне відхилення	16,932037
Дисперсія вибірки	295,03623	Дисперсія вибірки	286,69387
Ексцес	-1,220697	Ексцес	-1,196659
Асиметричність	-0,114305	Асиметричність	-0,189468
Інтервал	71	Інтервал	63
Мінімум	0	Мінімум	8
Максимум	71	Максимум	71
Сума	24921	Сума	50293
Кількість	627	Кількість	1198

Рисунок 3.85 – Показники описової статистики за 2021 р.

2023 р. (НМТ) Німецька мова			
Чоловіки		Жінки	
Середнє	17,649874	Середнє	18,925827
Стандартна помилка	0,2574613	Стандартна помилка	0,2229171
Медіана	17	Медіана	19
Мода	15	Мода	24
Стандартне відхилення	7,2547471	Стандартне відхилення	7,4569032
Дисперсія вибірки	52,631356	Дисперсія вибірки	55,605406
Ексцес	-0,946419	Ексцес	-0,947517
Асиметричність	0,0017659	Асиметричність	-0,222385
Інтервал	32	Інтервал	32
Мінімум	0	Мінімум	0
Максимум	32	Максимум	32
Сума	14014	Сума	21178
Кількість	794	Кількість	1119

Рисунок 3.86 – Показники описової статистики за 2023 р.

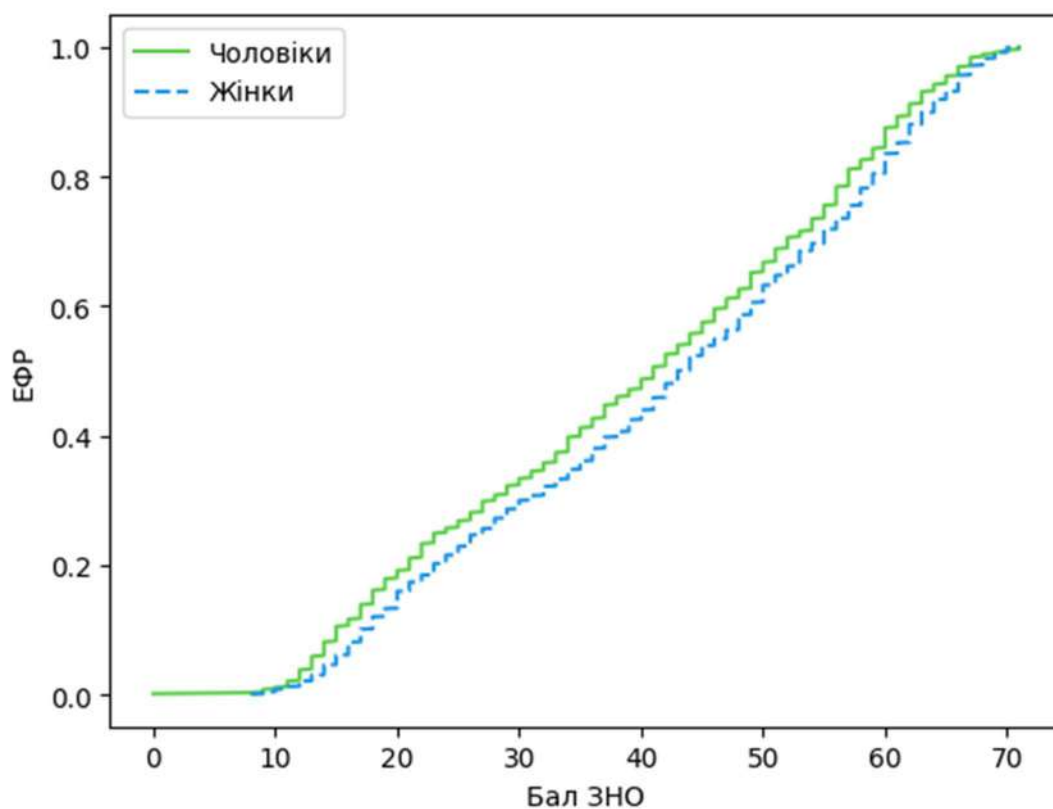


Рисунок 3.87 – Емпірична функція розподілу (Німецька мова 2021 р.)

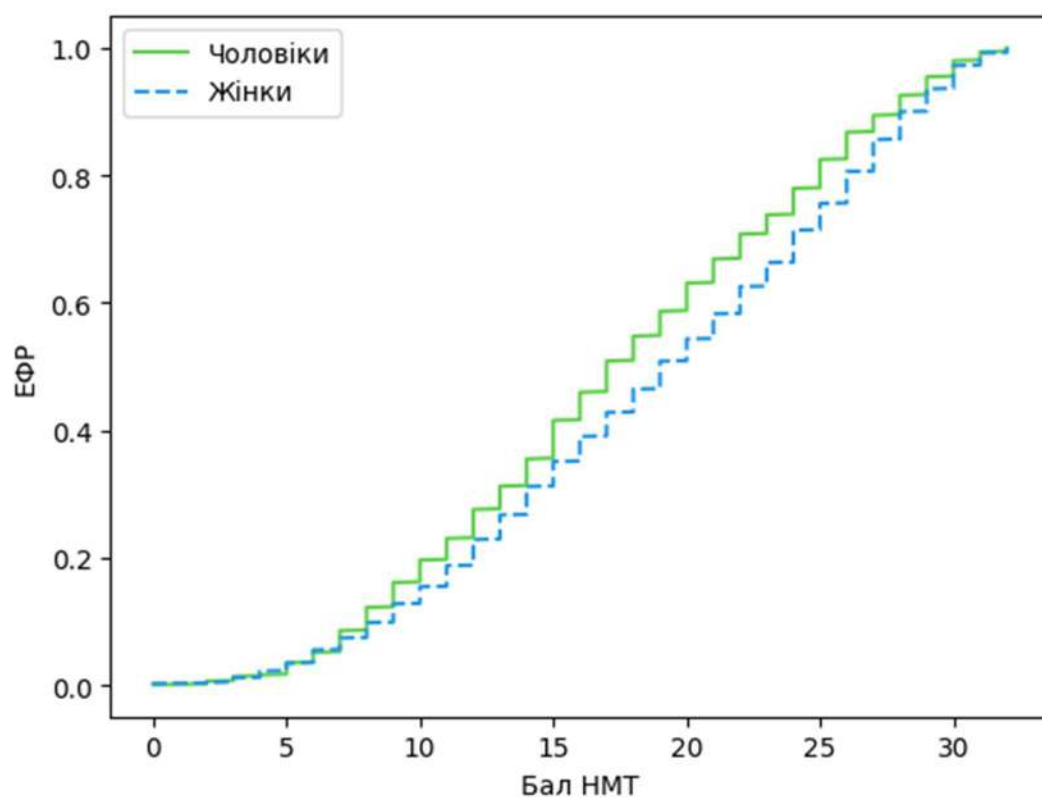


Рисунок 3.88 – Емпірична функція розподілу (Німецька мова 2023 р.)

Також для візуалізації даних побудуємо гістограми розподілу:

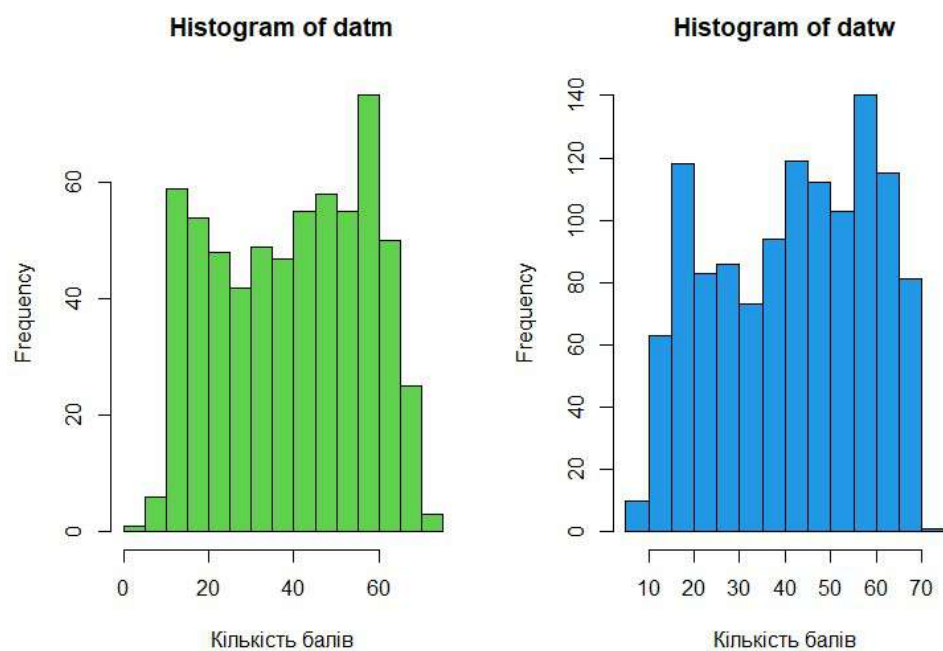


Рисунок 3.89 – Гістограма розподілу показників (Німецька мова 2021 р.)

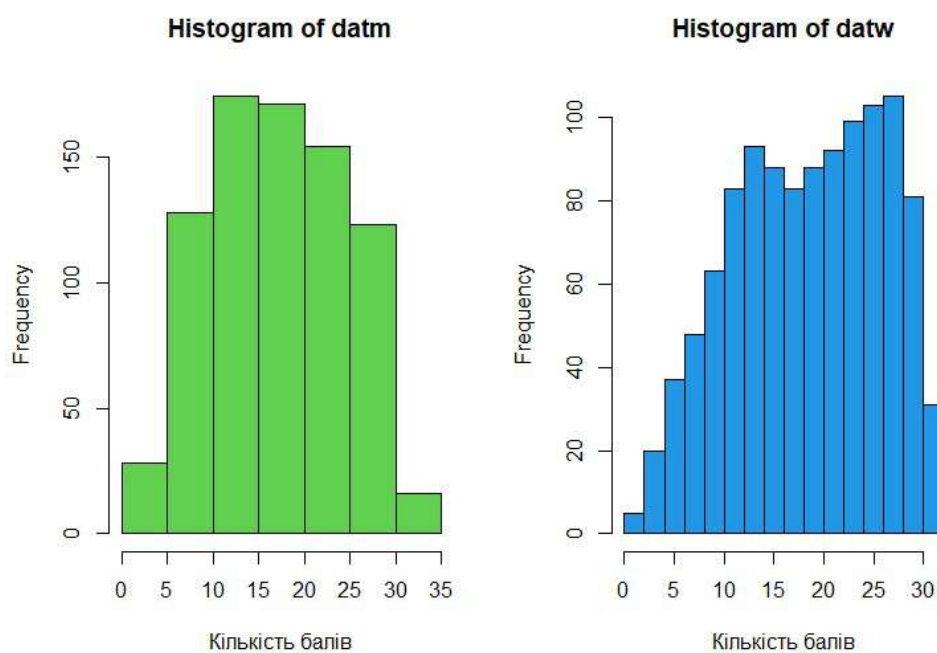


Рисунок 3.90 – Гістограма розподілу показників (Німецька мова 2023 р.)

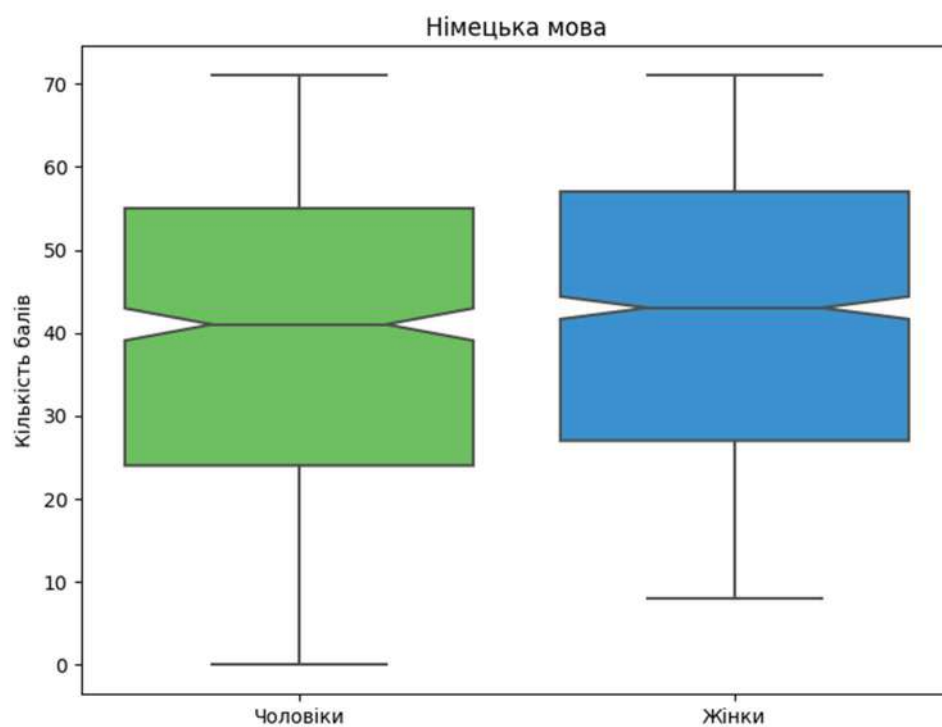


Рисунок 3.91 – Boxplot діаграма розподілу результатів чоловіків та жінок (Німецька мова 2021 р.)

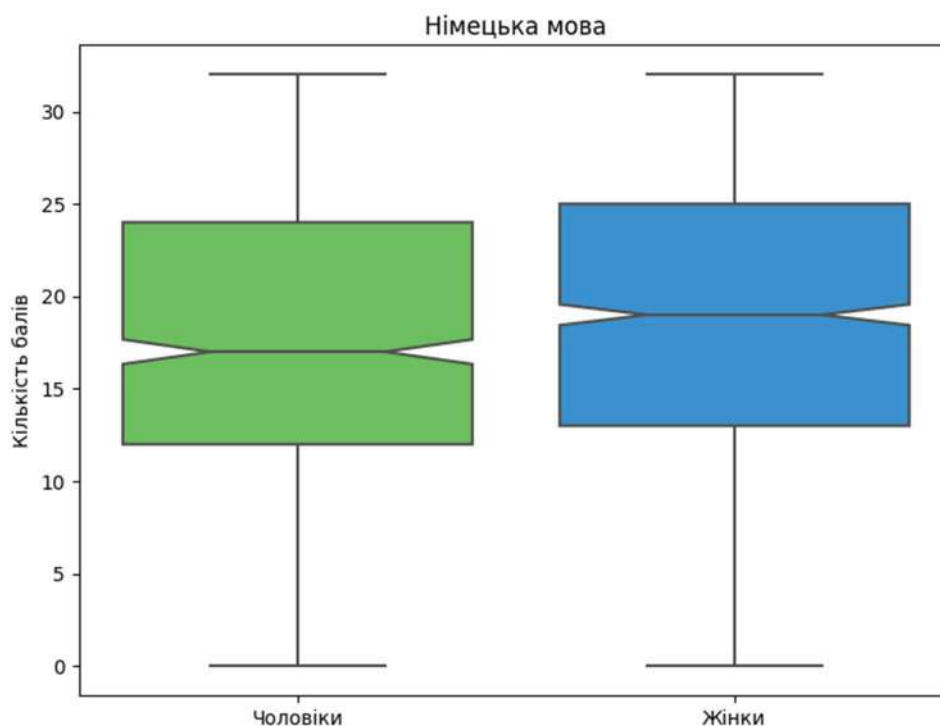


Рисунок 3.92 – Boxplot діаграма розподілу результатів чоловіків та жінок (Німецька мова 2023 р.)

Для подальшого докладного кількісного порівняння характеристик розподілів було визначено їх квантілі.

Отримані результати наведено у таблиці 3.48.

Таблиця 3.48 – Квантілі розподілу даних результатів з німецької мови

2021 р.	10%	25%	75%	90%	Макс
Чоловіки	15	24	55	62	71
Жінки	17	27	57	64	71
2023 р.	10%	25%	75%	90%	Макс
Чоловіки	8	12	24	28	32
Жінки	9	13	25	28.2	32

Перевіримо відповідність даних нормальному розподілу за допомогою статистичних критеріїв та занесемо отримані дані до табл. 3.48.

Таблиця 3.48 – Результати перевірки нормальності

Вибірка	Критерії			
	Лілієфорса		Андерсона-Дарлінга	
	D	p-value	A	p-value
Нім.мова, чоловіки 2021р.	0.084012	$2.333 \cdot 10^{-11}$	8.946	$< 2.2 \cdot 10^{-16}$
Нім.мова, жінки 2021р.	0.082031	$< 2.2 \cdot 10^{-16}$	17.363	$< 2.2 \cdot 10^{-16}$
Нім.мова, чоловіки 2023р	0.071261	$3.035 \cdot 10^{-10}$	4.8735	$4.565 \cdot 10^{-12}$
Нім.мова, жінки 2023р.	0.088803	$< 2.2 \cdot 10^{-16}$	9.9786	$< 2.2 \cdot 10^{-16}$

Критерії Лілієфорса та Андерсона-Дарлінга показують, що розподіли результатів тестувань у чоловіків та жінок не відповідають нормальному розподілу. P-value у обох критеріях менше за 0.05, що свідчить про статистичну значущість відхилень від нормального розподілу. Тому для перевірки гіпотези про однорідність вибірок чоловіків та жінок треба використовувати непараметричні критерії, для цього застосуємо критерій Колмогорова – Смирнова та критерій Уїлкоксона.

Таблиця 3.49 – Результат перевірки однорідності показників з німецької мови

Вибірка	Критерії			
	Колмогорова-Смирнова		Уїлкоксона	
	D	p-value	W	p-value
2021 р.	0.056377	0.1461	346899	0.007308
2023 р.	0.08764	0.001594	398764	0.0001317

3.11 Аналіз результатів тестувань НМТ з іспанської мови

Таблиця 3.50 – Кількість учасників тестування з іспанської мови (за 3 роки)

2021 р.		2022 р.		2023 р.	
Кількість учасників	150	Кількість учасників	не проводилося	Кількість учасників	146

Таблиця 3.51 – Процентне співвідношення чоловіків та жінок, які беруть участь у тестуванні

Стать	2021 р. (ЗНО)		2022 р. (НМТ)		2023 р. (НМТ)	
	кількість	% від загальної кількості учасників	кількість	% від загальної кількості учасників	кількість	% від загальної кількості учасників
Чоловіча	39	26%	не проводилося		53	36,30 %
Жіноча	111	74%			93	63,70%

2021 р. (ЗНО) Іспанська мова			
Чоловіки		Жінки	
Середнє	44,333333	Середнє	44,162162
Стандартна помилка	2,9214487	Стандартна помилка	1,9603164
Медіана	48	Медіана	51
Мода	56	Мода	65
Стандартне відхилення	18,244442	Стандартне відхилення	20,653215
Дисперсія вибірки	332,85965	Дисперсія вибірки	426,55528
Ексцес	-0,936522	Ексцес	-1,439084
Асиметричність	-0,373516	Асиметричність	-0,323329
Інтервал	62	Інтервал	62
Мінімум	9	Мінімум	10
Максимум	71	Максимум	72
Сума	1729	Сума	4902
Кількість	39	Кількість	111

Рисунок 3.93 – Показники описової статистики за 2021 р.

2023 р. (НМТ) Іспанська мова			
Чоловіки		Жінки	
Середнє	16,358491	Середнє	18,935484
Стандартна помилка	1,133695	Стандартна помилка	0,8455818
Медіана	16	Медіана	20
Мода	9	Мода	27
Стандартне відхилення	8,2534243	Стандартне відхилення	8,1544952
Дисперсія вибірки	68,119013	Дисперсія вибірки	66,495792
Екссес	-1,35727	Екссес	-1,128383
Асиметричність	0,06331	Асиметричність	-0,239769
Інтервал	26	Інтервал	28
Мінімум	4	Мінімум	4
Максимум	30	Максимум	32
Сума	867	Сума	1761
Кількість	53	Кількість	93

Рисунок 3.94 – Показники описової статистики за 2023 р.

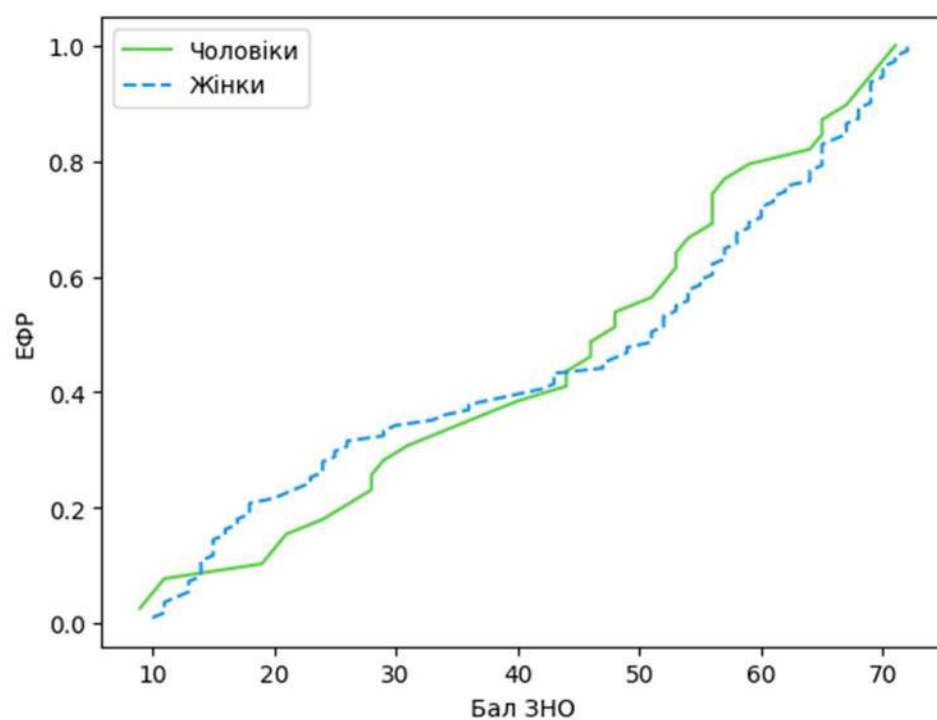


Рисунок 3.95 – Емпірична функція розподілу (Іспанська мова 2021 р.)

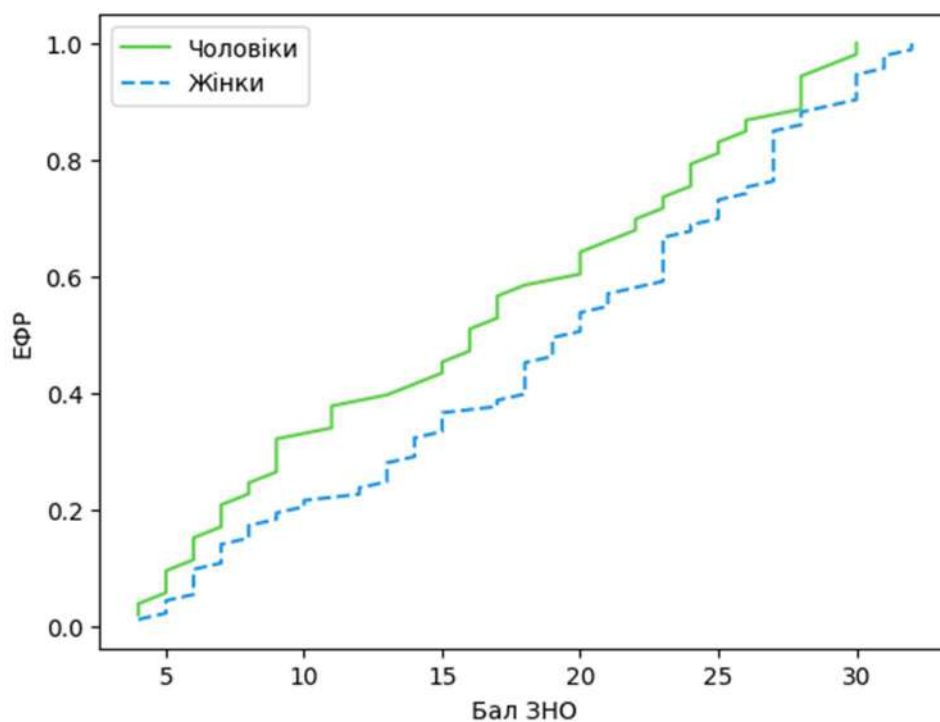


Рисунок 3.96 – Емпірична функція розподілу (Іспанська мова 2023 р.)

Також для візуалізації даних побудуємо гістограми розподілу:

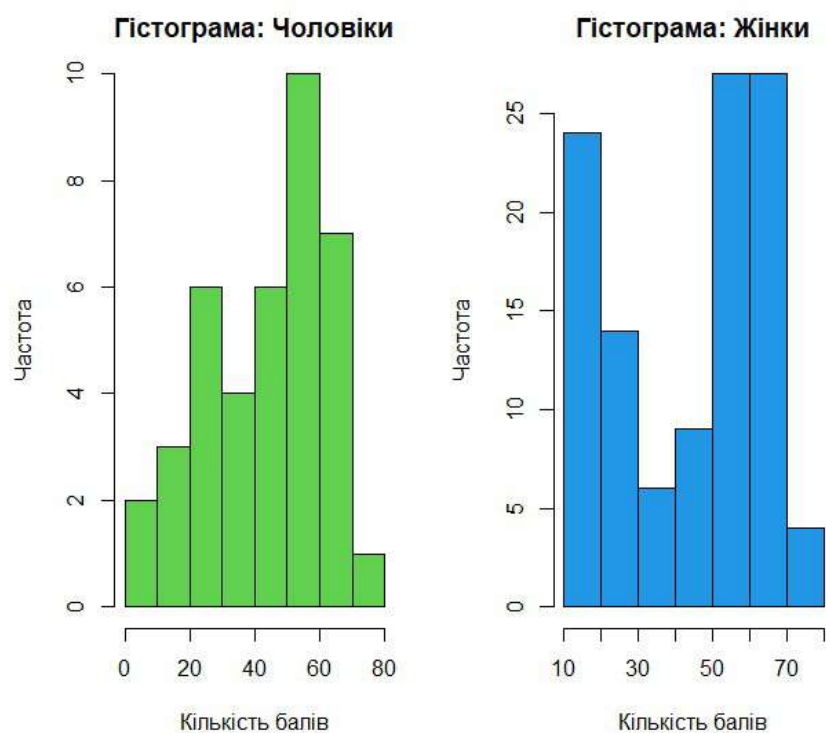


Рисунок 3.97 – Гістограма розподілу показників (Іспанська мова 2021 р.)

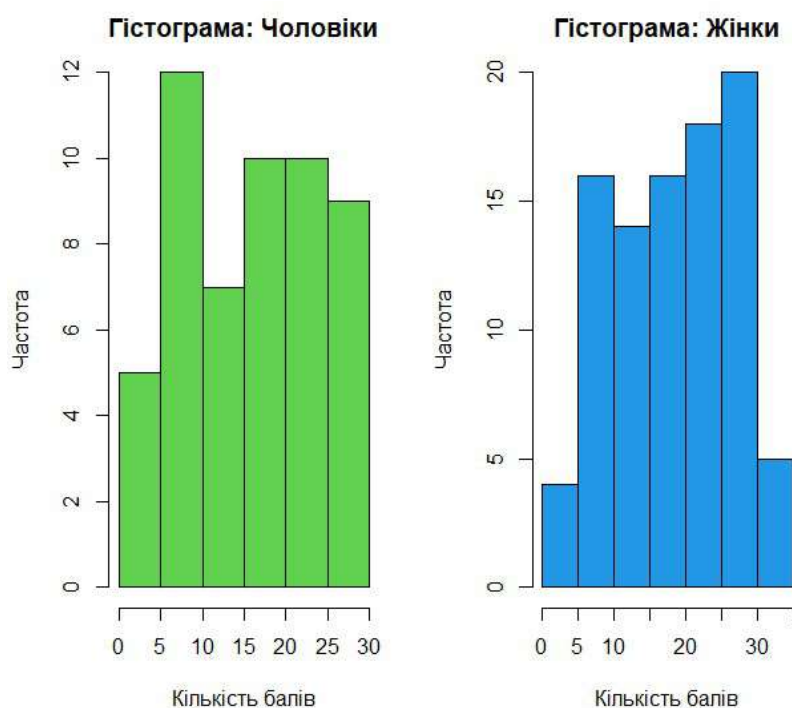


Рисунок 3.98 – Гістограма розподілу показників (Іспанська мова 2023 р.)

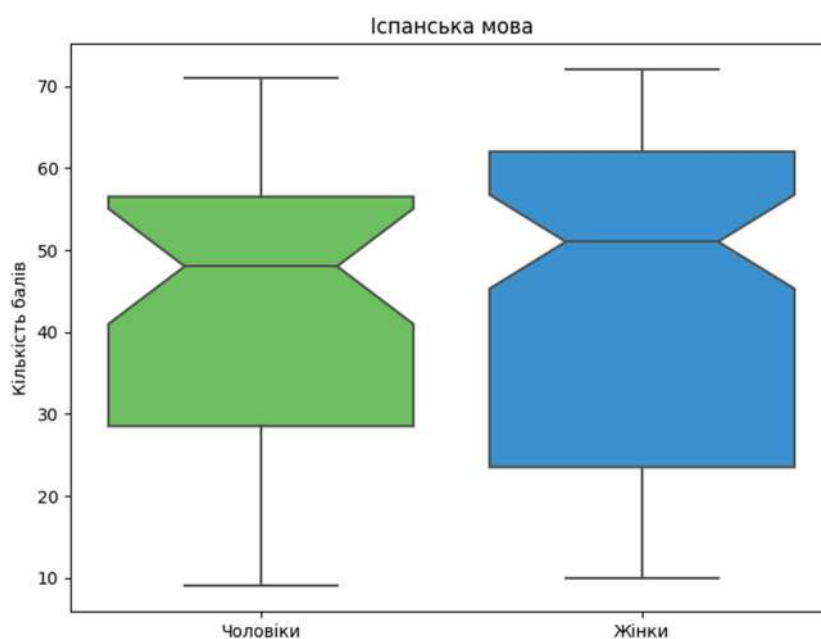


Рисунок 3.99 – Вохplot діаграма розподілу результатів чоловіків та жінок (Іспанська мова 2021 р.)

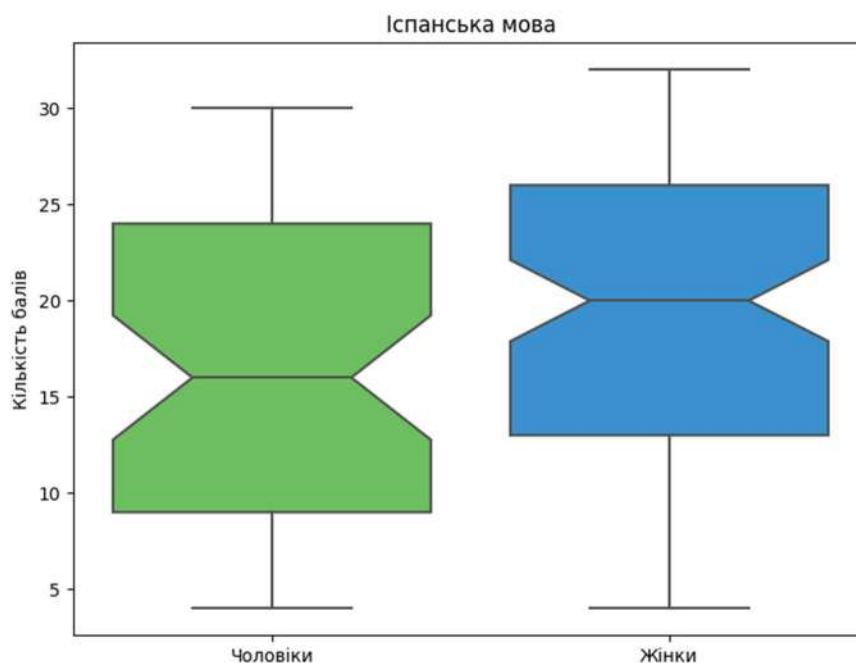


Рисунок 3.100 – Boxplot діаграма розподілу результатів чоловіків та жінок (Іспанська мова 2023 р.)

Для подальшого докладного кількісного порівняння характеристик розподілів було визначено їх квантілі.

Отримані результати наведено у таблиці 3.52.

Таблиця 3.52 – Квантілі розподілу даних результатів з іспанської мови

2021 р.	10%	25%	75%	90%	Макс
Чоловіки	19.8	28.5	56.5	67.2	71
Жінки	14	23.5	62	69	72
2023 р.	10%	25%	75%	90%	Макс
Чоловіки	6	9	24	28	30
Жінки	7	13	26	29.8	32

Перевіримо відповідність даних нормальному розподілу за допомогою статистичних критеріїв та занесемо отримані дані до табл. 3.4

Таблиця 3.53 – Результати перевірки нормальності

Вибірка	Критерії			
	Ліллієфорса		Андерсона-Дарлінга	
	D	p-value	A	p-value
Ісп.мова, чоловіки 2021р.	0.1081	0.2993	0.61931	0.09949
Ісп.мова, жінки 2021р.	0.15223	$1.188 \cdot 10^{-6}$	4.3074	$9.12 \cdot 10^{-11}$
Ісп.мова, чоловіки 2023р	0.13444	0.01798	1.0736	0.007407
Ісп.мова, жінки 2023р.	0.11027	0.007178	1.4043	0.001167

Критерії Ліллієфорса та Андерсона-Дарлінга показують, що розподіли результатів тестувань у чоловіків та жінок не відповідають нормальному розподілу. Р-value у обох критеріях менше за 0.05, що свідчить про статистичну значущість відхилень від нормального розподілу. Тому для перевірки гіпотези про однорідність вибірок чоловіків та жінок треба використовувати непараметричні критерії, для цього застосуємо критерій Колмогорова – Смирнова та критерій Уїлкоксона.

Таблиця 3.54 – Результат перевірки однорідності показників з іспанської мови

Вибірка	Критерії			
	Колмогорова-Смирнова		Уїлкоксона	
	D	p-value	W	p-value
2021 р.	0.13028	0.5868	2114.5	0.832
2023 р.	0.17894	0.1475	2033	0.07916

3.12 Аналіз різниці середніх балів між чоловіками та жінками

Таблиця 3.55 – Середні бали результатів тестувань ЗНО та НМТ за 2021-2023 роки

Предмет	Рік					
	2021		2022		2023	
	чоловіки	жінки	чоловіки	жінки	чоловіки	жінки
Українська мова	49.77	62.89	17.5	20.49	21.52	25.49
Математика	19.64	19.56	14.14	13.93	11.99	11.92
Історія України	37.22	41.01	17.24	18.35	22.33	24.08
Фізика	24.25	24.96	не проводилося		12.87	12.94
Хімія	32.83	30.78	не проводилося		19.18	17.63
Біологія	43.58	40.60	не проводилося		20.66	22.58
Географія	44.14	45.33	не проводилося		не проводилося	
Англійська мова	38.42	40.88	не проводилося		18.12	19.42
Французька мова	42.84	43.94	не проводилося		15.73	17
Німецька мова	39.74	41.98	не проводилося		17.64	18.92
Іспанська мова	44.33	44.16	не проводилося		16.35	18.93

Таблиця 3.56 – Різниця середніх балів результатів тестувань ЗНО та НМТ за 2021-2023 роки

Предмет	Рік		
	2021	2022	2023
Українська мова	13.12	2.99	3.97
Математика	0.08	0.21	0.07
Історія України	3.79	1.11	1.75
Фізика	0.71	не проводилося	0.07

Продовження табл.3.56

Предмет	Рік		
	2021	2022	2023
Хімія	2.05	не проводилося	1.54
Біологія	3.25	не проводилося	1.92
Географія	1.19	не проводилося	не проводилося
Англійська мова	2.46	не проводилося	1.3
Французька мова	1.1	не проводилося	1.27
Німецька мова	2.24	не проводилося	1.28
Іспанська мова	0.17	не проводилося	2.58

ВИСНОВКИ

В дипломній роботі було розглянуто проблему гендерних відмінностей результатів навчання, зокрема у контексті результатів ЗНО та НМТ в Україні за 2021-2023 роки. Було проаналізовано результати з таких предметів як: українська мова, математика, історія України, фізика, хімія, біологія, географія, англійська мова, французька мова, німецька мова, іспанська мова.

На основі результатів ЗНО та НМТ за 2021-2023 роки найбільш великий розрив спостерігається між чоловіками та жінками з української мови на користь жінок. Результати статистичного аналізу з української мови свідчать про відхилення нульової гіпотези про відсутність гендерних відмінностей у результатах тестів. Зокрема, низькі значення p -value для обох критеріїв свідчать про значущі розходження між розподілами результатів чоловіків та жінок. Отже, можна визначити, що гендерні відмінності у результатах учасників ЗНО та НМТ є статистично значущими на протязі трьох років. Різниця середніх балів між чоловіками та жінками є практично значущою та перевищує 0,5 балів. Це може бути зумовлено різними підходами до вивчення мови чоловіків та жінок, а також можливим впливом гендерних стереотипів. Тестування з української мови є обов'язковою умовою для вступу до вищих навчальних закладів, тому можна очікувати, що при вступі на бюджет за спеціальностями з високими прохідними балами ця відмінність може бути істотною.

Результати аналізу результатів з математики також показують статистичну значущість різниці між результатами чоловіків та жінок, хоча середні бали результатів мають незначні розходження на користь чоловіків.

З історії України жінки мають вищі результати, проведений статистичний аналіз підтверджує статистичну значущість відмінностей між результатами чоловіків та жінок. Значення P -value свідчить про значущі розходження між розподілами. Різниця у середніх балах перевищує 0,5 та є практично значущою на протязі трьох років.

Результати перевірки однорідності показників у хімії, біології, географії, англійської та німецькій мовах підтверджують статистичну значущість розбіжностей між результатами чоловіків та жінок. У хімії виявлено, що чоловіки мають вищі середні бали в обох роках спостереження. Стосовно біології, у 2021 році чоловіки показали кращі результати, з розривом у 3,25 бали, але в 2023 році жінки перевершили чоловіків на 1,92 бали. У географії жінки також показали результати на 1,19 бали вищі за чоловіків. В англійській та німецьких мовах жінки також мають кращі результати.

З французької та іспанської мови значення p -value більше 0,05 та на підставі цих результатів немає достатніх доказів для відхилення нульової гіпотези про однорідність. З французької мови, у 2021 році жінки випереджають чоловіків на 1,1 бали, в 2023 році цей показник збільшився до 1,27, тому можна казати про практичну значущість відмінностей. В іспанській мові жінки показали більш високі результати, різниця середніх балів у 2021 році склала 0,17, а в 2023 році цей показник збільшився до 2,58.

Загальна тенденція вказує на те, що гендерні відмінності в результатах ЗНО та НМТ з деяких предметів є стійкими та зберігаються протягом розглянутого періоду за 2021-2023 роки. Це може викликати серйозні питання щодо рівності у навчанні та вимагає системного підходу для розуміння та вирішення цих проблем.

ПЕРЕЛІКИ ПОСИЛАНЬ

1. Основне про національний мультипредметний тест [Електронний ресурс]. – Режим доступу: <https://testportal.gov.ua/osnovne-pro-nmt/>.
2. UN. 17 Goals to Transform Our World [Електронний ресурс]. – Режим доступу: <https://un.org/sustainabledevelopment/>.
3. UNESCO. Gender equality and education [Електронний ресурс]. – Режим доступу: <https://unesco.org/en/gender-equality/education>.
4. School enrollment, primary (% net) [Електронний ресурс]. – Режим доступу: <https://data.worldbank.org/indicator/SE.PRM.NENR.MA>.
5. Countries by mean years of schooling [Електронний ресурс]. – Режим доступу: <http://data.uis.unesco.org/index.aspx?queryid=3803>.
6. Sex Differences in Education: an overview [Електронний ресурс]. – Режим доступу: <https://www.tandfonline.com/doi/abs/10.1080/0305006870230102>.
7. OECD Employment Outlook 2012 [Електронний ресурс]. – Режим доступу: https://www.oecd-ilibrary.org/employment/oecd-employment-outlook-2012/summary/estonian_empl_outlook-2012-sum-et.
8. Are boys discriminated in Swedish high schools? [Електронний ресурс]. – Режим доступу: <https://www.sciencedirect.com/science/article/abs/pii/S0272775711000227>.
9. Classroom Interactions: Gender of Teacher, Gender of Student, and Classroom Subject [Електронний ресурс]. – Режим доступу: <https://link.springer.com/article/10.1023/A:1014892408105>.
10. TIMSS-2019 International Results in Mathematics and Science [Електронний ресурс]. – Режим доступу: <https://timss2019.org/reports/>.
11. PIRLS [Електронний ресурс]. – Режим доступу: <https://www.iea.nl/studies/iea/pirls>.

12. PISA [Электронный ресурс]. – Режим доступа: <https://www.oecd.org/pisa/>.
13. The Gender-Equality Paradox in Science, Technology, Engineering, and Mathematics Education [Электронный ресурс]. – Режим доступа: <https://journals.sagepub.com/doi/10.1177/0956797617741719>.
14. Why are boys doing badly at school? [Электронный ресурс]. – Режим доступа: <https://www.economist.com/the-economist-explains/2022/11/23/why-are-boys-doing-badly-at-school>.

ДОДАТОК А.

Код програми

Ukr_2021.ipynb

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kstest, anderson, ks_2samp, wilcoxon
file_path = '/content/drive/MyDrive/Ukr_2021.csv'
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=130580)
datm = np.array(dat.iloc[0:91668, 0])
datw = np.array(dat.iloc[0:130580, 1])
plt.figure(figsize=(8, 6))
sns.boxplot(x=["Чоловіки"] * len(datm) + ["Жінки"] * len(datw),
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
plt.ylabel("Кількість балів")
plt.title("Українська мова")
plt.show()
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(datm, bins=8, color="#61d04f", alpha=0.7)
plt.title("Гістограма: Чоловіки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.subplot(1, 2, 2)
plt.hist(datw, bins=8, color="#2297e6", alpha=0.7)
plt.title("Гістограма: Жінки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.tight_layout()
plt.show()
erfm = np.linspace(1/len(datm), 1, len(datm))
erfw = np.linspace(1/len(datw), 1, len(datw))
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--")
plt.xlabel("Бал ЗНО")
plt.ylabel("ЕФР")
plt.legend()
plt.show()
print("Квантилі для чоловіків:", np.percentile(datm, [10, 25, 75, 90, 100]))
print("Квантилі для жінок:", np.percentile(datw, [10, 25, 75, 90, 100]))

```

```

result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)
print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)
p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Ліллієфорса для чоловіків:", p_value_lilliefors_m)
print("Тест Ліллієфорса для жінок:", p_value_lilliefors_w)
print("KS-тест:", ks_2samp(datm, datw))
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона:", wilcoxon(datm[:common_length], datw[:common_length],
alternative='two-sided'))

```

Ukr_2022.ipynb

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kstest, anderson, ks_2samp, wilcoxon
file_path = '/content/drive/MyDrive/Ukr_NMT.csv'
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=112290)
datm = np.array(dat.iloc[0:101784, 0])
datw = np.array(dat.iloc[0:112290, 1])
plt.figure(figsize=(8, 6))
sns.boxplot(x=["Чоловіки"] * len(datm) + ["Жінки"] * len(datw),
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
plt.ylabel("Кількість балів")
plt.title("Українська мова")
plt.show()
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(datm, bins=8, color="#61d04f", alpha=0.7)
plt.title("Гістограма: Чоловіки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")

```

```

plt.subplot(1, 2, 2)
plt.hist(datw, bins=8, color="#2297e6", alpha=0.7)
plt.title("Гістограма: Жінки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.tight_layout()
plt.show()
erfm = np.linspace(1/len(datm), 1, len(datm))
erfw = np.linspace(1/len(datw), 1, len(datw))
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--")
plt.xlabel("Бал НМТ")
plt.ylabel("ЕФР")
plt.legend()
plt.show()
print("Квантилі для чоловіків:", np.percentile(datm, [10, 25, 75, 90, 100]))
print("Квантилі для жінок:", np.percentile(datw, [10, 25, 75, 90, 100]))
result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)
print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)
p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Ліллієфорса для чоловіків:", p_value_lilliefors_m)
print("Тест Ліллієфорса для жінок:", p_value_lilliefors_w)
print("KS-тест:", ks_2samp(datm, datw))
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона:", wilcoxon(datm[:common_length], datw[:common_length],
alternative='two-sided'))

```

Ukr_2023.ipynb

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kstest, anderson, ks_2samp, wilcoxon

```

```

file_path = '/content/drive/MyDrive/Ukr_NMT2023.csv'
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=139723)
datm = np.array(dat.iloc[0:128405, 0])
datw = np.array(dat.iloc[0:139723, 1])
plt.figure(figsize=(8, 6))
sns.boxplot(x=["Чоловіки"] * len(datm) + ["Жінки"] * len(datw),
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
plt.ylabel("Кількість балів")
plt.title("Українська мова")
plt.show()
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(datm, bins=8, color="#61d04f", alpha=0.7)
plt.title("Гістограма: Чоловіки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.subplot(1, 2, 2)
plt.hist(datw, bins=8, color="#2297e6", alpha=0.7)
plt.title("Гістограма: Жінки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.tight_layout()
plt.show()
erfm = np.linspace(1/len(datm), 1, len(datm))
erfw = np.linspace(1/len(datw), 1, len(datw))
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--")
plt.xlabel("Бал НМТ")
plt.ylabel("ЕФР")
plt.legend()
plt.show()
print("Квантилі для чоловіків:", np.percentile(datm, [10, 25, 75, 90, 100]))
print("Квантилі для жінок:", np.percentile(datw, [10, 25, 75, 90, 100]))
result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)
print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)

```

```

p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Ліллієфорса для чоловіків:", p_value_lilliefors_m)
print("Тест Ліллієфорса для жінок:", p_value_lilliefors_w)
print("KS-тест:", ks_2samp(datm, datw))
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона:", wilcoxon(datm[:common_length], datw[:common_length],
alternative='two-sided'))

```

Math_2021.ipynb

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kstest, anderson, ks_2samp, wilcoxon
file_path = '/content/drive/MyDrive/Math_2021.csv'
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=125319)
datm = np.array(dat.iloc[0:125319, 0])
datw = np.array(dat.iloc[0:118883, 1])
plt.figure(figsize=(8, 6))
sns.boxplot(x=["Чоловіки"] * len(datm) + ["Жінки"] * len(datw),
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
plt.ylabel("Кількість балів")
plt.title("Математика")
plt.show()
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(datm, bins=8, color="#61d04f", alpha=0.7)
plt.title("Гістограма: Чоловіки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.subplot(1, 2, 2)
plt.hist(datw, bins=8, color="#2297e6", alpha=0.7)
plt.title("Гістограма: Жінки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.tight_layout()
plt.show()
erfm = np.linspace(1/len(datm), 1, len(datm))
erfw = np.linspace(1/len(datw), 1, len(datw))
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--")
plt.xlabel("Бал ЗНО")

```

```

plt.ylabel("ЕФР")
plt.legend()
plt.show()
print("Квантили для чоловіків:", np.percentile(datm, [10, 25, 75, 90, 100]))
print("Квантили для жінок:", np.percentile(datw, [10, 25, 75, 90, 100]))
result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)
print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)
p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Ліллієфорса для чоловіків:", p_value_lilliefors_m)
print("Тест Ліллієфорса для жінок:", p_value_lilliefors_w)
print("KS-тест:", ks_2samp(datm, datw))
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона:", wilcoxon(datm[:common_length], datw[:common_length],
alternative='two-sided'))

```

Math_2022.ipynb

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kstest, anderson, ks_2samp, wilcoxon
file_path = '/content/drive/MyDrive/Math_NMT.csv'
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=112290)
datm = np.array(dat.iloc[0:101784, 0])
datw = np.array(dat.iloc[0:112290, 1])
plt.figure(figsize=(8, 6))
sns.boxplot(x=["Чоловіки"] * len(datm) + ["Жінки"] * len(datw),
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
plt.ylabel("Кількість балів")
plt.title("Математика")
plt.show()
plt.figure(figsize=(12, 6))

```

```

plt.subplot(1, 2, 1)
plt.hist(datm, bins=8, color="#61d04f", alpha=0.7)
plt.title("Гістограма: Чоловіки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.subplot(1, 2, 2)
plt.hist(datw, bins=8, color="#2297e6", alpha=0.7)
plt.title("Гістограма: Жінки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.tight_layout()
plt.show()
erfm = np.linspace(1/len(datm), 1, len(datm))
erfw = np.linspace(1/len(datw), 1, len(datw))
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--")
plt.xlabel("Бал НМТ")
plt.ylabel("ЕФР")
plt.legend()
plt.show()
print("Квантилі для чоловіків:", np.percentile(datm, [10, 25, 75, 90, 100]))
print("Квантилі для жінок:", np.percentile(datw, [10, 25, 75, 90, 100]))
result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)
print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)
p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Ліллієфорса для чоловіків:", p_value_lilliefors_m)
print("Тест Ліллієфорса для жінок:", p_value_lilliefors_w)
print("KS-тест:", ks_2samp(datm, datw))
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона:", wilcoxon(datm[:common_length], datw[:common_length],
alternative='two-sided'))

```

Math_2023.ipynb

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kstest, anderson, ks_2samp, wilcoxon
file_path = '/content/drive/MyDrive/Math23.csv'
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=112290)
datm = np.array(dat.iloc[0:128405, 0])
datw = np.array(dat.iloc[0:139723, 1])
plt.figure(figsize=(8, 6))
sns.boxplot(x=["Чоловіки"] * len(datm) + ["Жінки"] * len(datw),
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
plt.ylabel("Кількість балів")
plt.title("Математика")
plt.show()
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(datm, bins=8, color="#61d04f", alpha=0.7)
plt.title("Гістограма: Чоловіки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.subplot(1, 2, 2)
plt.hist(datw, bins=8, color="#2297e6", alpha=0.7)
plt.title("Гістограма: Жінки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.tight_layout()
plt.show()
erfm = np.linspace(1/len(datm), 1, len(datm))
erfw = np.linspace(1/len(datw), 1, len(datw))
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--")
plt.xlabel("Бал НМТ")
plt.ylabel("ЕФР")
plt.legend()
plt.show()
print("Квантилі для чоловіків:", np.percentile(datm, [10, 25, 75, 90, 100]))
print("Квантилі для жінок:", np.percentile(datw, [10, 25, 75, 90, 100]))
result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
```

```

print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)
print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)
p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Ліллієфорса для чоловіків:", p_value_lilliefors_m)
print("Тест Ліллієфорса для жінок:", p_value_lilliefors_w)
print("KS-тест:", ks_2samp(datm, datw))
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона:", wilcoxon(datm[:common_length], datw[:common_length],
alternative='two-sided'))

```

HistUkr_2021.ipynb

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kstest, anderson, ks_2samp, wilcoxon
file_path = '/content/drive/MyDrive/HistUkr_2021.csv'
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=108347)
datm = np.array(dat.iloc[0:91026, 0])
datw = np.array(dat.iloc[0:108347, 1])
plt.figure(figsize=(8, 6))
sns.boxplot(x=["Чоловіки"] * len(datm) + ["Жінки"] * len(datw),
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
plt.ylabel("Кількість балів")
plt.title("Історія України")
plt.show()
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(datm, bins=8, color="#61d04f", alpha=0.7)
plt.title("Гістограма: Чоловіки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.subplot(1, 2, 2)
plt.hist(datw, bins=8, color="#2297e6", alpha=0.7)
plt.title("Гістограма: Жінки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")

```

```

plt.tight_layout()
plt.show()
erfm = np.linspace(1/len(datm), 1, len(datm))
erfw = np.linspace(1/len(datw), 1, len(datw))
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--")
plt.xlabel("Бал ЗНО")
plt.ylabel("ЕФР")
plt.legend()
plt.show()
print("Квантилі для чоловіків:", np.percentile(datm, [10, 25, 75, 90, 100]))
print("Квантилі для жінок:", np.percentile(datw, [10, 25, 75, 90, 100]))
result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)
print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)
p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Ліллієфорса для чоловіків:", p_value_lilliefors_m)
print("Тест Ліллієфорса для жінок:", p_value_lilliefors_w)
print("KS-тест:", ks_2samp(datm, datw))
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона:", wilcoxon(datm[:common_length], datw[:common_length],
alternative='two-sided'))

```

HistUkr_2022.ipynb

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kstest, anderson, ks_2samp, wilcoxon
file_path = '/content/drive/MyDrive/HistUkr_NMT.csv'
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=112290)
datm = np.array(dat.iloc[0:101784, 0])
datw = np.array(dat.iloc[0:112290, 1])

```

```

plt.figure(figsize=(8, 6))
sns.boxplot(x=["Чоловіки"] * len(datm) + ["Жінки"] * len(datw),
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
plt.ylabel("Кількість балів")
plt.title("Історія України")
plt.show()
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(datm, bins=8, color="#61d04f", alpha=0.7)
plt.title("Гістограма: Чоловіки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.subplot(1, 2, 2)
plt.hist(datw, bins=8, color="#2297e6", alpha=0.7)
plt.title("Гістограма: Жінки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.tight_layout()
plt.show()
erfm = np.linspace(1/len(datm), 1, len(datm))
erfw = np.linspace(1/len(datw), 1, len(datw))
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--")
plt.xlabel("Бал НМТ")
plt.ylabel("ЕФР")
plt.legend()
plt.show()
print("Квантилі для чоловіків:", np.percentile(datm, [10, 25, 75, 90, 100]))
print("Квантилі для жінок:", np.percentile(datw, [10, 25, 75, 90, 100]))
result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)
print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)
p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Ліллієфорса для чоловіків", p_value_lilliefors_m)
print("Тест Ліллієфорса для жінок", p_value_lilliefors_w)

```

```
print("KS-тест:", ks_2samp(datm, datw))
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона:", wilcoxon(datm[:common_length], datw[:common_length],
alternative='two-sided'))
```

HistUkr_2023.ipynb

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kstest, anderson, ks_2samp, wilcoxon
file_path = '/content/drive/MyDrive/HistUkr23.csv'
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=72225)
datm = np.array(dat.iloc[0:68844, 0])
datw = np.array(dat.iloc[0:72225, 1])
plt.figure(figsize=(8, 6))
sns.boxplot(x=["Чоловіки"] * len(datm) + ["Жінки"] * len(datw),
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
plt.ylabel("Кількість балів")
plt.title("Історія України")
plt.show()
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(datm, bins=8, color="#61d04f", alpha=0.7)
plt.title("Гістограма: Чоловіки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.subplot(1, 2, 2)
plt.hist(datw, bins=8, color="#2297e6", alpha=0.7)
plt.title("Гістограма: Жінки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.tight_layout()
plt.show()
erfm = np.linspace(1/len(datm), 1, len(datm))
erfw = np.linspace(1/len(datw), 1, len(datw))
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--")
plt.xlabel("Бал НМТ")
plt.ylabel("ЕФР")
plt.legend()
plt.show()
print("Квантили для чоловіків:", np.percentile(datm, [10, 25, 75, 90, 100]))
print("Квантили для жінок:", np.percentile(datw, [10, 25, 75, 90, 100]))
```

```

result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)
print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)
p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Ліллієфорса для чоловіків:", p_value_lilliefors_m)
print("Тест Ліллієфорса для жінок:", p_value_lilliefors_w)
print("KS-тест:", ks_2samp(datm, datw))
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона:", wilcoxon(datm[:common_length], datw[:common_length],
alternative='two-sided'))

```

Phys_2021.ipynb

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kstest, anderson, ks_2samp, wilcoxon
file_path = '/content/drive/MyDrive/Phys_2021.csv'
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=19810)
datm = np.array(dat.iloc[0:19810, 0])
datw = np.array(dat.iloc[0:3393, 1])
plt.figure(figsize=(8, 6))
sns.boxplot(x=["Чоловіки"] * len(datm) + ["Жінки"] * len(datw),
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
plt.ylabel("Кількість балів")
plt.title("Фізика")
plt.show()
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(datm, bins=8, color="#61d04f", alpha=0.7)
plt.title("Гістограма: Чоловіки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")

```

```

plt.subplot(1, 2, 2)
plt.hist(datw, bins=8, color="#2297e6", alpha=0.7)
plt.title("Гістограма: Жінки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.tight_layout()
plt.show()
erfm = np.linspace(1/len(datm), 1, len(datm))
erfw = np.linspace(1/len(datw), 1, len(datw))
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--")
plt.xlabel("Бал ЗНО")
plt.ylabel("ЕФР")
plt.legend()
plt.show()
print("Квантилі для чоловіків:", np.percentile(datm, [10, 25, 75, 90, 100]))
print("Квантилі для жінок:", np.percentile(datw, [10, 25, 75, 90, 100]))
result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)
print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)
p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Ліллієфорса для чоловіків:", p_value_lilliefors_m)
print("Тест Ліллієфорса для жінок:", p_value_lilliefors_w)
print("KS-тест:", ks_2samp(datm, datw))
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона:", wilcoxon(datm[:common_length], datw[:common_length],
alternative='two-sided'))

```

Phys_2023.ipynb

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kstest, anderson, ks_2samp, wilcoxon

```

```

file_path = '/content/drive/MyDrive/Phys23.csv'
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=4540)
datm = np.array(dat.iloc[0:4540, 0])
datw = np.array(dat.iloc[0:714, 1])
plt.figure(figsize=(8, 6))
sns.boxplot(x=["Чоловіки"] * len(datm) + ["Жінки"] * len(datw),
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
plt.ylabel("Кількість балів")
plt.title("Фізика")
plt.show()
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(datm, bins=8, color="#61d04f", alpha=0.7)
plt.title("Гістограма: Чоловіки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.subplot(1, 2, 2)
plt.hist(datw, bins=8, color="#2297e6", alpha=0.7)
plt.title("Гістограма: Жінки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.tight_layout()
plt.show()
erfm = np.linspace(1/len(datm), 1, len(datm))
erfw = np.linspace(1/len(datw), 1, len(datw))
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--")
plt.xlabel("Бал НМТ")
plt.ylabel("ЕФР")
plt.legend()
plt.show()
print("Квантилі для чоловіків:", np.percentile(datm, [10, 25, 75, 90, 100]))
print("Квантилі для жінок:", np.percentile(datw, [10, 25, 75, 90, 100]))
result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)
print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)

```

```

p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Ліллієфорса для чоловіків:", p_value_lilliefors_m)
print("Тест Ліллієфорса для жінок:", p_value_lilliefors_w)
print("KS-тест:", ks_2samp(datm, datw))
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона:", wilcoxon(datm[:common_length], datw[:common_length],
alternative='two-sided'))

```

Chemistry_2021.ipynb

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kstest, anderson, ks_2samp, wilcoxon
file_path = '/content/drive/MyDrive/Chemistry_2021.csv'
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=6213)
datm = np.array(dat.iloc[0:3592, 0])
datw = np.array(dat.iloc[0:6213, 1])
plt.figure(figsize=(8, 6))
sns.boxplot(x=["Чоловіки"] * len(datm) + ["Жінки"] * len(datw),
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
plt.ylabel("Кількість балів")
plt.title("Хімія")
plt.show()
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(datm, bins=8, color="#61d04f", alpha=0.7)
plt.title("Гістограма: Чоловіки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.subplot(1, 2, 2)
plt.hist(datw, bins=8, color="#2297e6", alpha=0.7)
plt.title("Гістограма: Жінки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.tight_layout()
plt.show()
erfm = np.linspace(1/len(datm), 1, len(datm))
erfw = np.linspace(1/len(datw), 1, len(datw))
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--")
plt.xlabel("Бал ЗНО")
plt.ylabel("ЕФР")

```

```

plt.legend()
plt.show()
print("Квантили для чоловіків:", np.percentile(datm, [10, 25, 75, 90, 100]))
print("Квантили для жінок:", np.percentile(datw, [10, 25, 75, 90, 100]))
result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)
print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)
p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Ліллієфорса для чоловіків:", p_value_lilliefors_m)
print("Тест Ліллієфорса для жінок:", p_value_lilliefors_w)
print("KS-тест:", ks_2samp(datm, datw))
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона:", wilcoxon(datm[:common_length], datw[:common_length],
alternative='two-sided'))

```

Chemistry_2023.ipynb

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kstest, anderson, ks_2samp, wilcoxon
file_path = '/content/drive/MyDrive/Chem23.csv'
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=2133)
datm = np.array(dat.iloc[0:1085, 0])
datw = np.array(dat.iloc[0:2133, 1])
plt.figure(figsize=(8, 6))
sns.boxplot(x=["Чоловіки"] * len(datm) + ["Жінки"] * len(datw),
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
plt.ylabel("Кількість балів")
plt.title("Хімія")
plt.show()
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(datm, bins=8, color="#61d04f", alpha=0.7)

```

```

plt.title("Гістограма: Чоловіки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.subplot(1, 2, 2)
plt.hist(datw, bins=8, color="#2297e6", alpha=0.7)
plt.title("Гістограма: Жінки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.tight_layout()
plt.show()
erfm = np.linspace(1/len(datm), 1, len(datm))
erfw = np.linspace(1/len(datw), 1, len(datw))
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--")
plt.xlabel("Бал НМТ")
plt.ylabel("ЕФР")
plt.legend()
plt.show()
print("Квантилі для чоловіків:", np.percentile(datm, [10, 25, 75, 90, 100]))
print("Квантилі для жінок:", np.percentile(datw, [10, 25, 75, 90, 100]))
result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)
print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)
p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Ліллієфорса для чоловіків:", p_value_lilliefors_m)
print("Тест Ліллієфорса для жінок:", p_value_lilliefors_w)
print("KS-тест:", ks_2samp(datm, datw))
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона:", wilcoxon(datm[:common_length], datw[:common_length],
alternative='two-sided'))

```

Biology_2021.ipynb

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kstest, anderson, ks_2samp, wilcoxon
file_path = '/content/drive/MyDrive/Biology_2021.csv'
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=52001)
datm = np.array(dat.iloc[0:29410, 0])
datw = np.array(dat.iloc[0:52001, 1])
plt.figure(figsize=(8, 6))
sns.boxplot(x=["Чоловіки"] * len(datm) + ["Жінки"] * len(datw),
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
plt.ylabel("Кількість балів")
plt.title("Біологія")
plt.show()
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(datm, bins=8, color="#61d04f", alpha=0.7)
plt.title("Гістограма: Чоловіки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.subplot(1, 2, 2)
plt.hist(datw, bins=8, color="#2297e6", alpha=0.7)
plt.title("Гістограма: Жінки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.tight_layout()
plt.show()
erfm = np.linspace(1/len(datm), 1, len(datm))
erfw = np.linspace(1/len(datw), 1, len(datw))
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--")
plt.xlabel("Бал ЗНО")
plt.ylabel("ЕФР")
plt.legend()
plt.show()
print("Квантилі для чоловіків:", np.percentile(datm, [10, 25, 75, 90, 100]))
print("Квантилі для жінок:", np.percentile(datw, [10, 25, 75, 90, 100]))
result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
```

```

print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)
print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)
p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Ліллієфорса для чоловіків:", p_value_lilliefors_m)
print("Тест Ліллієфорса для жінок:", p_value_lilliefors_w)
print("KS-тест:", ks_2samp(datm, datw))
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона:", wilcoxon(datm[:common_length], datw[:common_length],
alternative='two-sided'))

```

Biology_2023.ipynb

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kstest, anderson, ks_2samp, wilcoxon
file_path = '/content/drive/MyDrive/Biology23.csv'
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=24099)
datm = np.array(dat.iloc[0:11910, 0])
datw = np.array(dat.iloc[0:24099, 1])
plt.figure(figsize=(8, 6))
sns.boxplot(x=["Чоловіки"] * len(datm) + ["Жінки"] * len(datw),
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
plt.ylabel("Кількість балів")
plt.title("Біологія")
plt.show()
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(datm, bins=8, color="#61d04f", alpha=0.7)
plt.title("Гістограма: Чоловіки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.subplot(1, 2, 2)
plt.hist(datw, bins=8, color="#2297e6", alpha=0.7)
plt.title("Гістограма: Жінки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")

```

```

plt.tight_layout()
plt.show()
erfm = np.linspace(1/len(datm), 1, len(datm))
erfw = np.linspace(1/len(datw), 1, len(datw))
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--")
plt.xlabel("Бал НМТ")
plt.ylabel("ЕФР")
plt.legend()
plt.show()
print("Квантилі для чоловіків:", np.percentile(datm, [10, 25, 75, 90, 100]))
print("Квантилі для жінок:", np.percentile(datw, [10, 25, 75, 90, 100]))
result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)
print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)
p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Лілієфорса для чоловіків:", p_value_lilliefors_m)
print("Тест Лілієфорса для жінок:", p_value_lilliefors_w)
print("KS-тест:", ks_2samp(datm, datw))
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона:", wilcoxon(datm[:common_length], datw[:common_length],
alternative='two-sided'))

```

Geography_2021.ipynb

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kstest, anderson, ks_2samp, wilcoxon
file_path = '/content/drive/MyDrive/Geography_2021.csv'
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=58548)
datm = np.array(dat.iloc[0:58548, 0])
datw = np.array(dat.iloc[0:54568, 1])

```

```

plt.figure(figsize=(8, 6))
sns.boxplot(x=["Чоловіки"] * len(datm) + ["Жінки"] * len(datw),
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
plt.ylabel("Кількість балів")
plt.title("Географія")
plt.show()
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(datm, bins=8, color="#61d04f", alpha=0.7)
plt.title("Гістограма: Чоловіки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.subplot(1, 2, 2)
plt.hist(datw, bins=8, color="#2297e6", alpha=0.7)
plt.title("Гістограма: Жінки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.tight_layout()
plt.show()
erfm = np.linspace(1/len(datm), 1, len(datm))
erfw = np.linspace(1/len(datw), 1, len(datw))
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--")
plt.xlabel("Бал ЗНО")
plt.ylabel("ЕФР")
plt.legend()
plt.show()
print("Квантилі для чоловіків:", np.percentile(datm, [10, 25, 75, 90, 100]))
print("Квантилі для жінок:", np.percentile(datw, [10, 25, 75, 90, 100]))
result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)
print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)
p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Ліллієфорса для чоловіків", p_value_lilliefors_m)
print("Тест Ліллієфорса для жінок", p_value_lilliefors_w)

```

```
print("KS-тест:", ks_2samp(datm, datw))
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона:", wilcoxon(datm[:common_length], datw[:common_length],
alternative='two-sided'))
```

English_2021.ipynb

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kstest, anderson, ks_2samp, wilcoxon
file_path = '/content/drive/MyDrive/English_2021.csv'
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=64878)
datm = np.array(dat.iloc[0:62836, 0])
datw = np.array(dat.iloc[0:64878, 1])
plt.figure(figsize=(8, 6))
sns.boxplot(x=["Чоловіки"] * len(datm) + ["Жінки"] * len(datw),
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
plt.ylabel("Кількість балів")
plt.title("Англійська мова")
plt.show()
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(datm, bins=8, color="#61d04f", alpha=0.7)
plt.title("Гістограма: Чоловіки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.subplot(1, 2, 2)
plt.hist(datw, bins=8, color="#2297e6", alpha=0.7)
plt.title("Гістограма: Жінки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.tight_layout()
plt.show()
erfm = np.linspace(1/len(datm), 1, len(datm))
erfw = np.linspace(1/len(datw), 1, len(datw))
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--")
plt.xlabel("Бал ЗНО")
plt.ylabel("ЕФР")
plt.legend()
plt.show()
print("Квантилі для чоловіків:", np.percentile(datm, [10, 25, 75, 90, 100]))
print("Квантилі для жінок:", np.percentile(datw, [10, 25, 75, 90, 100]))
```

```

result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)
print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)
p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Ліллієфорса для чоловіків:", p_value_lilliefors_m)
print("Тест Ліллієфорса для жінок:", p_value_lilliefors_w)
print("KS-тест:", ks_2samp(datm, datw))
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона:", wilcoxon(datm[:common_length], datw[:common_length],
alternative='two-sided'))

```

English_2023.ipynb

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kstest, anderson, ks_2samp, wilcoxon
file_path = '/content/drive/MyDrive/English23.csv'
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=41093)
datm = np.array(dat.iloc[0:41093, 0])
datw = np.array(dat.iloc[0:39154, 1])
plt.figure(figsize=(8, 6))
sns.boxplot(x=["Чоловіки"] * len(datm) + ["Жінки"] * len(datw),
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
plt.ylabel("Кількість балів")
plt.title("Англійська мова")
plt.show()
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(datm, bins=8, color="#61d04f", alpha=0.7)
plt.title("Гістограма: Чоловіки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")

```

```

plt.subplot(1, 2, 2)
plt.hist(datw, bins=8, color="#2297e6", alpha=0.7)
plt.title("Гістограма: Жінки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.tight_layout()
plt.show()
erfm = np.linspace(1/len(datm), 1, len(datm))
erfw = np.linspace(1/len(datw), 1, len(datw))
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--")
plt.xlabel("Бал НМТ")
plt.ylabel("ЕФР")
plt.legend()
plt.show()
print("Квантилі для чоловіків:", np.percentile(datm, [10, 25, 75, 90, 100]))
print("Квантилі для жінок:", np.percentile(datw, [10, 25, 75, 90, 100]))
result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)
print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)
p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Ліллієфорса для чоловіків:", p_value_lilliefors_m)
print("Тест Ліллієфорса для жінок:", p_value_lilliefors_w)
print("KS-тест:", ks_2samp(datm, datw))
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона:", wilcoxon(datm[:common_length], datw[:common_length],
alternative='two-sided'))

```

French_2021.ipynb

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

```

```

from scipy.stats import kstest, anderson, ks_2samp, wilcoxon
file_path = '/content/drive/MyDrive/French_2021.csv'
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=256)
datm = np.array(dat.iloc[0:99, 0])
datw = np.array(dat.iloc[0:256, 1])
plt.figure(figsize=(8, 6))
sns.boxplot(x=["Чоловіки" * len(datm) + ["Жінки" * len(datw)],
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
plt.ylabel("Кількість балів")
plt.title("Французька мова")
plt.show()
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(datm, bins=8, color="#61d04f", alpha=0.7)
plt.title("Гістограма: Чоловіки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.subplot(1, 2, 2)
plt.hist(datw, bins=8, color="#2297e6", alpha=0.7)
plt.title("Гістограма: Жінки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.tight_layout()
plt.show()
erfm = np.linspace(1/len(datm), 1, len(datm))
erfw = np.linspace(1/len(datw), 1, len(datw))
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--")
plt.xlabel("Бал ЗНО")
plt.ylabel("ЕФР")
plt.legend()
plt.show()
print("Квантилі для чоловіків:", np.percentile(datm, [10, 25, 75, 90, 100]))
print("Квантилі для жінок:", np.percentile(datw, [10, 25, 75, 90, 100]))
result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)

```

```

print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)
p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Ліллієфорса для чоловіків:", p_value_lilliefors_m)
print("Тест Ліллієфорса для жінок:", p_value_lilliefors_w)
print("KS-тест:", ks_2samp(datm, datw))
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона:", wilcoxon(datm[:common_length], datw[:common_length],
alternative='two-sided'))

```

French_2023.ipynb

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kstest, anderson, ks_2samp, wilcoxon
file_path = '/content/drive/MyDrive/French23.csv'
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=187)
datm = np.array(dat.iloc[0:86, 0])
datw = np.array(dat.iloc[0:187, 1])
plt.figure(figsize=(8, 6))
sns.boxplot(x=["Чоловіки"] * len(datm) + ["Жінки"] * len(datw),
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
plt.ylabel("Кількість балів")
plt.title("Французька мова")
plt.show()
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(datm, bins=8, color="#61d04f", alpha=0.7)
plt.title("Гістограма: Чоловіки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.subplot(1, 2, 2)
plt.hist(datw, bins=8, color="#2297e6", alpha=0.7)
plt.title("Гістограма: Жінки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.tight_layout()
plt.show()
erfm = np.linspace(1/len(datm), 1, len(datm))
erfw = np.linspace(1/len(datw), 1, len(datw))
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--")

```

```

plt.xlabel("Бал НМТ")
plt.ylabel("ЕФР")
plt.legend()
plt.show()
print("Квантилі для чоловіків:", np.percentile(datm, [10, 25, 75, 90, 100]))
print("Квантилі для жінок:", np.percentile(datw, [10, 25, 75, 90, 100]))
result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)
print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)
p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Ліллієфорса для чоловіків:", p_value_lilliefors_m)
print("Тест Ліллієфорса для жінок:", p_value_lilliefors_w)
print("KS-тест:", ks_2samp(datm, datw))
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона:", wilcoxon(datm[:common_length], datw[:common_length],
alternative='two-sided'))

```

Deutsch_2021.ipynb

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kstest, anderson, ks_2samp, wilcoxon
file_path = '/content/drive/MyDrive/Deutsch_2021.csv'
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=1198)
datm = np.array(dat.iloc[1:627, 0])
datw = np.array(dat.iloc[1:1198, 1])
plt.figure(figsize=(8, 6))
sns.boxplot(x=["Чоловіки"] * len(datm) + ["Жінки"] * len(datw),
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
plt.ylabel("Кількість балів")
plt.title("Німецька мова")
plt.show()

```

```

plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(datm, bins=8, color="#61d04f", alpha=0.7)
plt.title("Гістограма: Чоловіки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.subplot(1, 2, 2)
plt.hist(datw, bins=8, color="#2297e6", alpha=0.7)
plt.title("Гістограма: Жінки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.tight_layout()
plt.show()
erfm = np.linspace(1/len(datm), 1, len(datm))
erfw = np.linspace(1/len(datw), 1, len(datw))
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--") #
Встановлюємо пунктирний стиль для лінії жінок
plt.xlabel("Бал ЗНО")
plt.ylabel("ЕФР")
plt.legend()
plt.show()
print("Квантили для чоловіків:", np.percentile(datm, [10, 25, 75, 90, 100]))
print("Квантили для жінок:", np.percentile(datw, [10, 25, 75, 90, 100]))
result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)
print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)
p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Ліллієфорса для чоловіків:", p_value_lilliefors_m)
print("Тест Ліллієфорса для жінок:", p_value_lilliefors_w)
print("KS-тест:", ks_2samp(datm, datw))
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона:", wilcoxon(datm[:common_length], datw[:common_length],
alternative='two-sided'))

```

Deutsch_2023.ipynb

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kstest, anderson, ks_2samp, wilcoxon
file_path = '/content/drive/MyDrive/Deutsch23.csv'
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=1119)
datm = np.array(dat.iloc[0:794, 0])
datw = np.array(dat.iloc[0:1119, 1])
plt.figure(figsize=(8, 6))
sns.boxplot(x=["Чоловіки"] * len(datm) + ["Жінки"] * len(datw),
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
plt.ylabel("Кількість балів")
plt.title("Німецька мова")
plt.show()
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(datm, bins=8, color="#61d04f", alpha=0.7)
plt.title("Гістограма: Чоловіки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.subplot(1, 2, 2)
plt.hist(datw, bins=8, color="#2297e6", alpha=0.7)
plt.title("Гістограма: Жінки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.tight_layout()
plt.show()
erfm = np.linspace(1/len(datm), 1, len(datm))
erfw = np.linspace(1/len(datw), 1, len(datw))
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--") #
Встановлюємо пунктирний стиль для лінії жінок
plt.xlabel("Бал НМТ")
plt.ylabel("ЕФР")
plt.legend()
plt.show()
print("Квантилі для чоловіків:", np.percentile(datm, [10, 25, 75, 90, 100]))
print("Квантилі для жінок:", np.percentile(datw, [10, 25, 75, 90, 100]))
result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
```

```

print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)
print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)
p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Ліллієфорса для чоловіків:", p_value_lilliefors_m)
print("Тест Ліллієфорса для жінок:", p_value_lilliefors_w)
print("KS-тест:", ks_2samp(datm, datw))
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона:", wilcoxon(datm[:common_length], datw[:common_length],
alternative='two-sided'))

```

Esp_2021.ipynb

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kstest, anderson, ks_2samp, wilcoxon
file_path = '/content/drive/MyDrive/Esp_2021.csv'
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=111)
datm = np.array(dat.iloc[0:39, 0])
datw = np.array(dat.iloc[0:111, 1])
plt.figure(figsize=(8, 6))
sns.boxplot(x=["Чоловіки"] * len(datm) + ["Жінки"] * len(datw),
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
plt.ylabel("Кількість балів")
plt.title("Іспанська мова")
plt.show()
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(datm, bins=8, color="#61d04f", alpha=0.7)
plt.title("Гістограма: Чоловіки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.subplot(1, 2, 2)
plt.hist(datw, bins=8, color="#2297e6", alpha=0.7)
plt.title("Гістограма: Жінки")

```

```

plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.tight_layout()
plt.show()
erfm = np.linspace(1/len(datm), 1, len(datm))
erfw = np.linspace(1/len(datw), 1, len(datw))
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--")
plt.xlabel("Бал ЗНО")
plt.ylabel("ЕФР")
plt.legend()
plt.show()
print("Квантилі для чоловіків:", np.percentile(datm, [10, 25, 75, 90, 100]))
print("Квантилі для жінок:", np.percentile(datw, [10, 25, 75, 90, 100]))
result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)
print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)
p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Ліллієфорса для чоловіків:", p_value_lilliefors_m)
print("Тест Ліллієфорса для жінок:", p_value_lilliefors_w)
print("KS-тест:", ks_2samp(datm, datw))
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона.", wilcoxon(datm[:common_length], datw[:common_length],
alternative='two-sided'))

```

Esp_2023.ipynb

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import kstest, anderson, ks_2samp, wilcoxon
file_path = '/content/drive/MyDrive/Esp23.csv'
dat = pd.read_csv(file_path, sep=';', decimal=',', header=0, nrows=93)

```

```

datm = np.array(dat.iloc[0:53, 0])
datw = np.array(dat.iloc[0:93, 1])
plt.figure(figsize=(8, 6))
sns.boxplot(x=["Чоловіки"] * len(datm) + ["Жінки"] * len(datw),
y=np.concatenate([datm, datw]), notch=True, palette=["#61d04f", "#2297e6"])
plt.ylabel("Кількість балів")
plt.title("Іспанська мова")
plt.show()
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.hist(datm, bins=8, color="#61d04f", alpha=0.7)
plt.title("Гістограма: Чоловіки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.subplot(1, 2, 2)
plt.hist(datw, bins=8, color="#2297e6", alpha=0.7)
plt.title("Гістограма: Жінки")
plt.xlabel("Кількість балів")
plt.ylabel("Частота")
plt.tight_layout()
plt.show()
erfm = np.linspace(1/len(datm), 1, len(datm))
erfw = np.linspace(1/len(datw), 1, len(datw))
plt.plot(np.sort(datm), erfm, label="Чоловіки", color="#61d04f")
plt.plot(np.sort(datw), erfw, label="Жінки", color="#2297e6", linestyle="--") #
Встановлюємо пунктирний стиль для лінії жінок
plt.xlabel("Бал ЗНО")
plt.ylabel("ЕФР")
plt.legend()
plt.show()
print("Квантилі для чоловіків:", np.percentile(datm, [10, 25, 75, 90, 100]))
print("Квантилі для жінок:", np.percentile(datw, [10, 25, 75, 90, 100]))
result_anderson_m = anderson(datm, dist='norm')
result_anderson_w = anderson(datw, dist='norm')
print("Тест Андерсона-Дарлінга для чоловіків", result_anderson_m.statistic)
print("Critical values for Anderson- Darling test (men):",
result_anderson_m.critical_values)
print("Significance level for Anderson- Darling test (men):",
result_anderson_m.significance_level)
print("Тест Андерсона-Дарлінга для жінок:", result_anderson_w.statistic)
print("Critical values for Anderson- Darling test (women):",
result_anderson_w.critical_values)
print("Significance level for Anderson- Darling test (women):",
result_anderson_w.significance_level)

```

```
p_value_lilliefors_m = kstest(datm, 'norm').pvalue
p_value_lilliefors_w = kstest(datw, 'norm').pvalue
print("Тест Ліллієфорса для чоловіків:", p_value_lilliefors_m)
print("Тест Ліллієфорса для жінок:", p_value_lilliefors_w)
print("KS-тест:", ks_2samp(datm, datw))
common_length = min(len(datm), len(datw))
print("Тест Уїлкоксона:", wilcoxon(datm[:common_length], datw[:common_length],
alternative='two-sided'))
```