

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА

Факультет комп'ютерних наук та кібернетики
Кафедра дослідження операцій

Кваліфікаційна робота
на здобуття ступеня магістра

за спеціальністю 113 Прикладна математика
на тему:

Аналіз даних типу залишкового часу життя

Виконала студентка 2-го курсу
Козак Дарина Дмитрівна

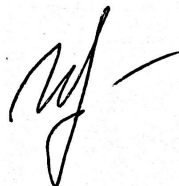


Науковий керівник:
професор, доктор фізико-математичних наук
Мацак Іван Каленикович



Робота заслухана на засіданні кафедри дослідження операцій та рекомендована до захисту в ЕК, протокол № 8 від "03" травня 2023 р.

Завідувач кафедри ДО



проф. Іксанов О.М.

Київ-2023

Зміст

Вступ	3
Проблема існування межі людського життя	6
Теоретичні основи дослідження тривалості життя	8
Бази даних довгожителів	10
Візуалізація даних	15
Проблеми моделювання розподілу	20
Алгоритм оптимізації	22
Розподіл Вейбула	25
Розподіл Гомперца	28
Дослідження розподілу	30
Висновки	37
Література	38
Додатки	40

Вступ

Загальний контекст досліджень тривалості життя в біології відносно широкий і комплексний. Старіння та віково-залежні захворювання є однією з основних проблем сучасного світу, і тому біологи та медики вкладають значні зусилля в дослідження цих процесів.

Однією з головних задач у дослідженнях тривалості життя є пошук факторів, які впливають на тривалість життя організму. Для цього дослідники аналізують взаємозв'язок між різними факторами, такими як генетика, середовищні та поведінкові фактори. Крім того, статистичний аналіз даних є невід'ємною частиною цього дослідження.

Дослідження розподілу залишкового часу життя довгожителів може бути корисним для багатьох областей, таких як медицина, страхування, пенсійне планування та соціальна політика.

У медицині дослідження розподілу залишкового часу життя довгожителів може допомогти встановити фактори, що впливають на довготривалість життя та здоров'я старшого покоління, та розробити стратегії щодо профілактики та лікування захворювань, що зменшують якість життя у цієї групи людей.

У страхуванні та пенсійному плануванні дослідження розподілу залишкового часу життя може допомогти зрозуміти ризики та збитки, пов'язані з виплатами пенсій та страхових виплат довгожителю, і дозволити страховим компаніям та пенсійним фондам краще оцінювати ризики та розробляти більш ефективні та стійкі

фінансові стратегії.

У соціальній політиці дослідження розподілу залишкового часу життя може допомогти розуміти потреби та вимоги старшого покоління, допомогти у розробці та впровадженні програм та послуг для старших людей, а також забезпечити стабільність та рівномірність розподілу соціальних ресурсів для підтримки якісного життя довгожителів.

У цій роботі ми намагатимемось розглянути старіння та тривалість життя як кількісну проблему, що може бути вирішена за допомогою статистичного аналізу даних.

Зокрема, ми зосередимось на дослідженні залишкової тривалості життя, після досягнення деякого досить високого (критичного) віку. Згідно з дослідженнями, тривалість життя після досягнення 80-90 років може бути істотно відрізнитися від загальної тривалості життя. Наприклад, дослідження показують, що при досягненні 90 років люди мають вищу ймовірність досягнути 100-річного віку, ніж ті, хто молодший за 90 років. Це може пояснюватися тим, що ті, хто досягнув цього віку, мають міцніший стан організму та меншу кількість генетичних дефектів, що можуть впливати на тривалість життя.

Тривалість життя після досягнення 100 років, також вивчається у дослідженнях, і є досить важливою темою в контексті здорового старіння та довголіття. Дослідження показали, що люди, які досягають віку понад 100 років, мають відмінності в генах та біохімічних маркерах, які можуть допомогти пояснити, чому вони доживають до такого віку. Такі дослідження можуть допомогти в розробці стратегій для збереження здоров'я на пізніших стадіях життя та підвищення тривалості життя. Важливо те, що дослідження показують, що розподіл тривалості життя людей після досягнення високого віку, наприклад 80-100 років, може значно

відрізнятися від розподілу загальної тривалості життя.

Крім того, статистичні дослідження також показують, що тривалість життя після досягнення віку високого віку може бути вплинута різними факторами. Наприклад, дослідження, проведені в Італії, показали, що серед людей, які досягли віку 105 років і старших, ті, хто живуть у сільській місцевості, мають значно більшу тривалість життя, ніж ті, хто живе в містах. Також було показано, що у жінок тривалість життя після досягнення віку високого віку може бути довшою, ніж у чоловіків.

Узагалі, вивчення тривалості життя після досягнення високого віку є важливим напрямком статистичних досліджень, оскільки це може допомогти встановити фактори, які впливають на тривалість життя, і допомогти людям підтримувати своє здоров'я та жити довше і здоровіше. Окрім цього, дана тема не є вичерпно дослідженою. Наприклад, спірним питанням залишається існування та величина максимального віку людини, хоча багато досліджень показали, що середня тривалість життя зростає та люди все частіше досягають високого віку.

Крім того, вчені не можуть точно визначити причини тривалого життя та чому деякі люди живуть до 100 років та більше. Деякі дослідження показали, що довголітність може бути спричинена генетичними факторами, а інші дослідження підтверджують, що на тривалість життя впливають здоровий спосіб життя, харчування та рівень фізичної активності.

Проблема існування межі людського життя

Проблема існування межі людського життя є важливою та спірною темою сьогодення. Дослідження, проведені в різних країнах, показують, що середня тривалість життя людей значно збільшилась в порівнянні з минулим століттям. Однак, чи існує межа, за якою людина не може жити довше, залишається питанням.

Деякі наукові дослідження схиляються до того, що існує можливість нескінченного збільшення максимальної тривалості життя людини. Однак, інші стверджують, що існує межа максимальної тривалості життя людини, яка залежить від генетичних факторів та здоров'я [1].

Дослідження в галузі математики дозволяють розглядати проблему максимальної тривалості людського життя з нової перспективи. Згідно з математичними моделями, які використовуються для аналізу процесу старіння та механізмів регуляції тривалості життя, більшість науковців схиляються до того, що максимальна межа тривалості життя людини існує.

Одним з ключових факторів, який визначає межу тривалості життя, є процес старіння. За даними досліджень, старіння є наслідком накопичення пошкоджень у клітинах та організмах в цілому. Ці пошкодження викликають порушення в роботі біологічних систем та зменшення регенеративних здібностей організму. Цей процес обмежує максимальну тривалість життя, якої може досягти людина.

Дослідження також вказують на те, що генетичні фактори можуть мати велике значення у визначенні тривалості життя [3].

Відомо, що деякі гени можуть впливати на ступінь вразливості організму до пошкоджень, що в свою чергу може відобразитися на тривалості життя.

За даними наукових досліджень, максимальна межа тривалості життя може становити близько 120 років. Це пов'язано з фізіологічними обмеженнями, зокрема з обмеженням регенерації тканин та клітин, а також з погіршенням роботи органів та систем в організмі [4].

Проте, оскільки чіткої межі максимальної тривалості життя людини не існує, ми будемо використовувати для моделювання розподіли, що можуть приймати необмежені значення, для яких, проте, імовірність потрапляння у кожен інтервал значень зменшується із зростанням номера інтервалу, аналогічно з тим як для моделювання росту людини використовується нормальний розподіл.

Теоретичні основи дослідження тривалості життя

Залежність тривалості життя людини від віку є фундаментальною у вивченні процесу старіння людського організму. Ця теорія базується на концепції залишкової тривалості життя, яка описує розподіл тривалості життя певної популяції людей, що досягли віку x , тобто скільки часу вони можуть очікувати жити після досягнення віку x [5].

Однією з основних математичних моделей, яка використовується для опису залишкової тривалості життя, є модель Гомперца-Мейкхама. У цій моделі передбачається, що ризик смерті збільшується з віком експоненційно. Формула цієї моделі має вигляд:

$$\lambda(x) = \lambda_0 e^{bx}$$

де $\lambda(x)$ - інтенсивність смертності, λ_0 - інтенсивність смертності відповідно до початкового віку, b - параметр, який характеризує швидкість зростання інтенсивності смертності з віком.

Іншою моделлю, яка використовується для аналізу тривалості життя, є модель Кендалла, яка базується на концепції кумулятивного ризику смерті. У цій моделі передбачається, що інтенсивність смертності є сталим параметром, а ризик смерті збільшується зі збільшенням віку. Формула для цієї моделі має вигляд:

$$S(x) = S_0 e^{-\int_0^x \mu(t) dt}$$

Для моделі Кендалла інтенсивність смертності $\mu(x)$ є сталим параметром, а ризик смерті збільшується зі збільшенням віку.

Ризик смерті можна виразити через функцію виживання $S(x)$, яка відображає ймовірність того, що людина виживе до віку x .

Щоб визначити параметри моделі Кендалла, можна використовувати метод максимальної правдоподібності, який полягає в тому, щоб знайти такі значення параметрів, при яких спостережувані дані мають найбільшу ймовірність.

Зокрема, для аналізу даних про тривалість життя людей зазвичай використовують табличний метод, що полягає у побудові таблиць виживання та розподілів смертності для різних груп віку. На основі цих таблиць можна оцінити параметри моделей та порівняти рівень смертності для різних груп населення [2].

Бази даних довгожителів

В дослідженні залишкового часу життя людей після досягнення високого віку використовуються різні бази даних, що дозволяють збирати інформацію про довгожителів і аналізувати її для дослідження динаміки старіння та факторів, що впливають на тривалість життя. Однак, в цьому питанні все ще присутня певна проблема даних, оскільки баз даних довгожителів не так багато, і вони не є дуже об'ємними.

Однією з найбільш відомих баз даних є "База даних довгожителів" (The International Database on Longevity або IDL), яка збирає інформацію про найстаріших людей з усього світу. Ця база даних містить інформацію про понад 15 000 людей, які дожили до віку 105 років і старших, і вона постійно оновлюється з новими даними про довгожителів.

Іншою відомою базою даних є "Старіння й довголіття" (Aging and Longevity), що була створена у Японії та містить інформацію про понад 10 000 людей, які дожили до віку 100 років і більше. Ця база даних також містить інформацію про фактори, що впливають на старіння та довголіття, такі як генетика, харчування та здоров'я.

Крім того, існує база даних "Населення та старіння" (The Human Mortality Database), яка збирає інформацію про смертність та тривалість життя з різних країн світу. Ця база даних містить інформацію про більше ніж 40 країн та дозволяє проводити порівняльний аналіз старіння та тривалості життя в різних країнах.

Усі ці бази даних дозволяють проводити наукові дослідження про старіння та довголіття, що може допомогти в розумінні фа-

кторів, які впливають на тривалість життя та розробці стратегій для збереження здоров'я та підвищення якості життя людей на старість.

У цій роботі ми працюватимемо переважно із базою даних IDL (The International Database on Longevity), оскільки вона є однією з найбільших і найдостовірніших. База даних була створена в 1990 році Майклом Віторгським (Michael G. W. V. Lorigeril), головним редактором журналу "Журнал довголіття та старіння" (Journal of Longevity and Aging).

Для збору даних до бази IDL використовуються різні джерела, зокрема наукові статті, повідомлення в ЗМІ та інтерв'ю з довгожителами. Основні джерела даних - це державні статистичні служби та національні організації, які відстежують народження та смертність у своїх країнах.

Дані, зібрані для IDL, є досить достовірними, оскільки вони базуються на державних статистичних даних та інших офіційних джерелах, які перевіряються на наявність помилок та недостовірностей. Проте, як і у будь-якій базі даних, можливі неточності та недоліки, що пов'язані з відмінностями у форматі збору та обробки даних у різних країнах.

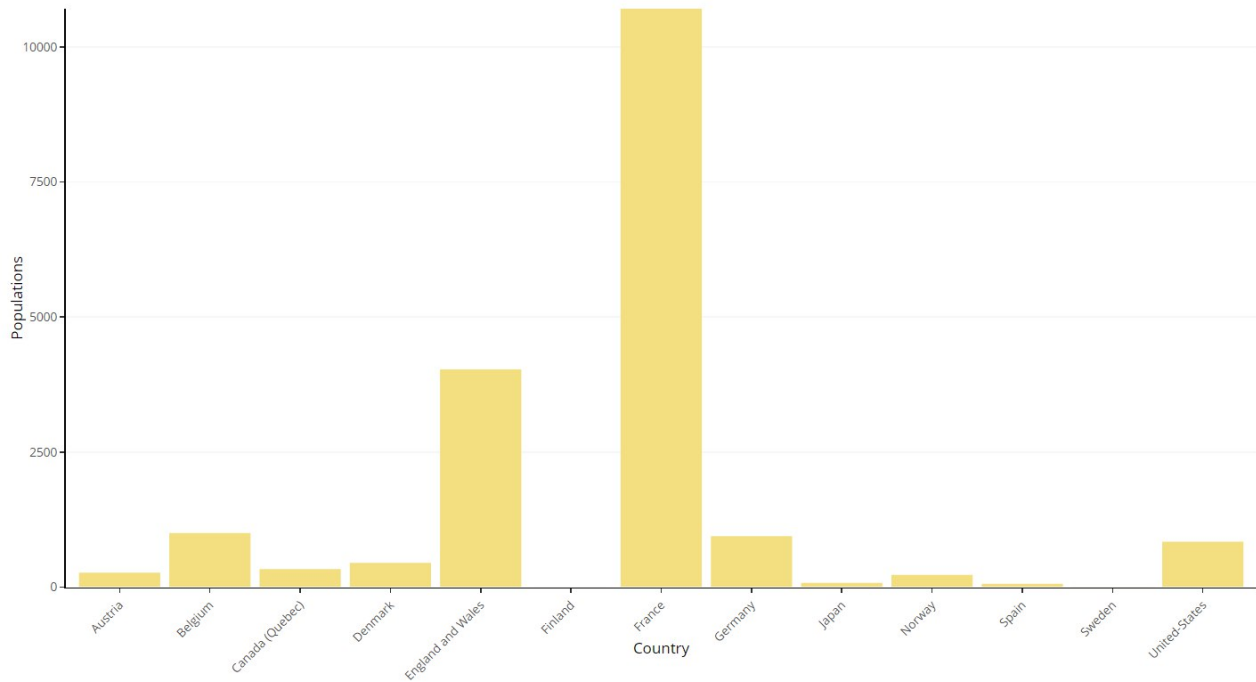
Більшість довгожителів, зареєстрованих в базі IDL, мешкають у Європі та Північній Америці. На яких континентах зосереджена більшість довгожителів може бути пов'язано з більш розвинутою системою статистики в країнах цих континентів. База даних IDL містить інформацію про людей, які живуть до 105 років та старших, іноді включаючи такі характеристики як стать, рівень освіти, заняття та стан здоров'я.

На даний момент, останнє оновлення бази даних IDL проводилось у 2020-2022 роках, особи, що прожили 105-109 років у ній називаються довгожителами, більше 110 років - супердовгожите-

лями. Проведемо оглядовий аналіз цієї бази даних.

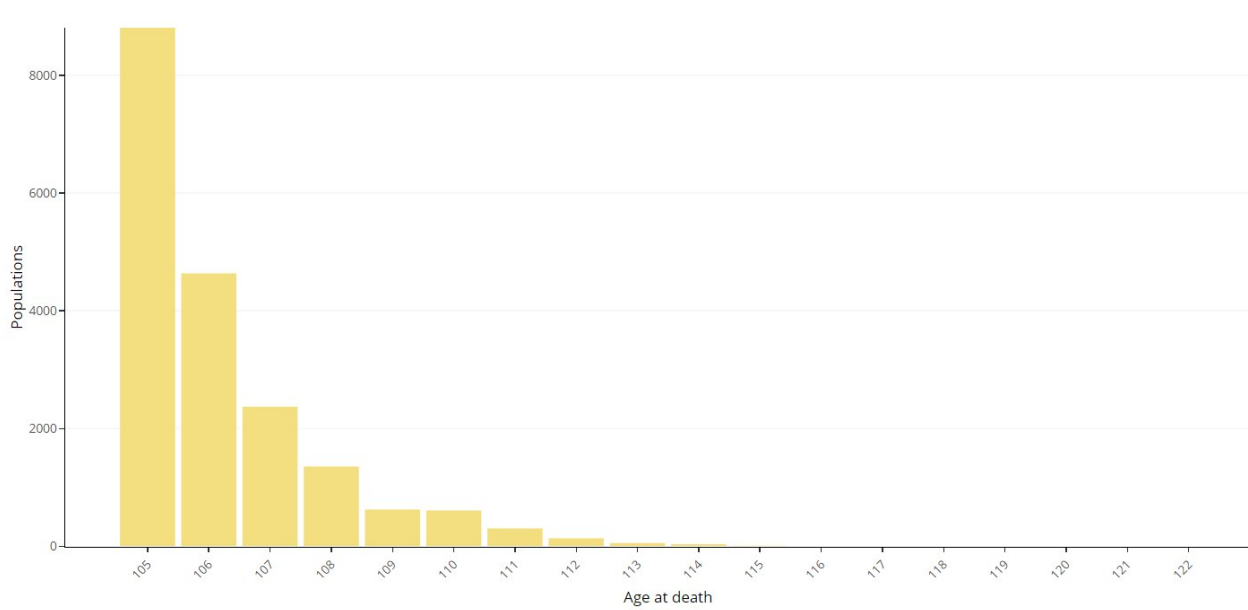
Загальна кількість довгожителів та супердовгожителів у ній складає 18 959, з них довгожителі (105-109 років) - 17 798 осіб, і супердовгожителі (110 років і більше) - 1 161 особа.

Створимо гістограму розподілу за країнами:



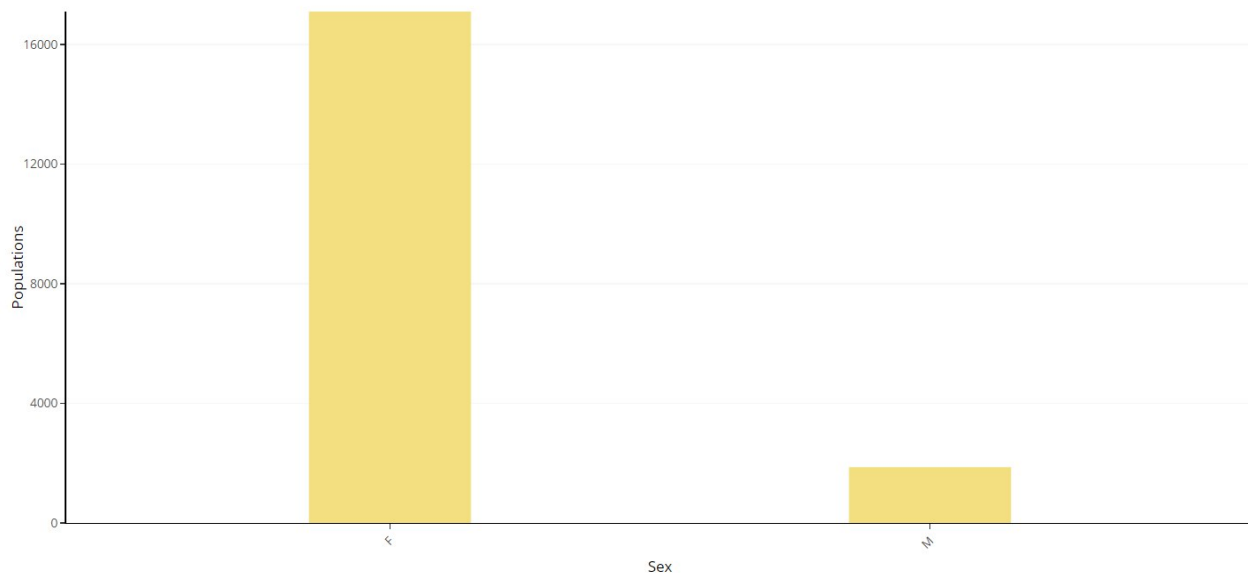
Отже, найбільше довгожителів в IDL із Франції, на другому місці - Велика Британія, на третьому - Німеччина.

Зобразимо також гістограму розподілу віку довгожителів:



Помітно, що кількість довгожителів, що досягли даного віку, спадає зі збільшенням віку, причому розподіл дещо нагадує експоненційний. Найбільший вік у базі даних - 122 роки.

Цікаво також зобразити розподіл довгожителів за статтю:



Бачимо, що кількість жінок-довгожительок значно переважає кількість чоловіків: 17 092 проти 1 867, що є цікавою ідеєю для

досліджень.

Також, нам буде цікаво дослідити дані для довгожителів після 90-95 років, яких немає в базі IDL, для цього використовуватимемо базу даних Dutch Statistical Office (C.V.S.), що містять дані про вік смерті довгожителів із Нідерландів 1877-1881 років народження.

До неї входять дані про смертність 20 081 особи, до яких входять 8280 чоловіків та 11 801 жінок.

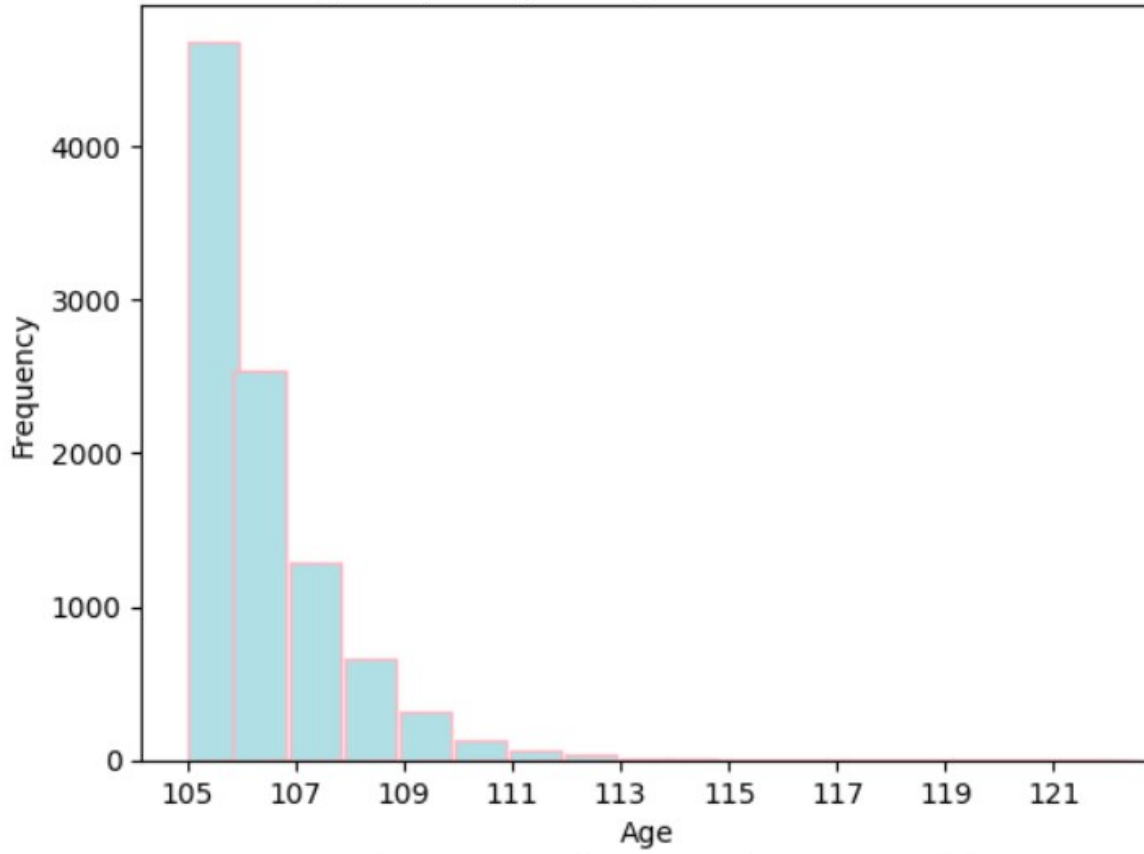
Візуалізація даних

Отже, використаємо базу даних довгожителів IDL, і будемо досліджувати розподіл даних окремо для різних країн, де кількість записів є достатньо великою. Топ-10 країн за кількістю записів відображена в наступній таблиці:

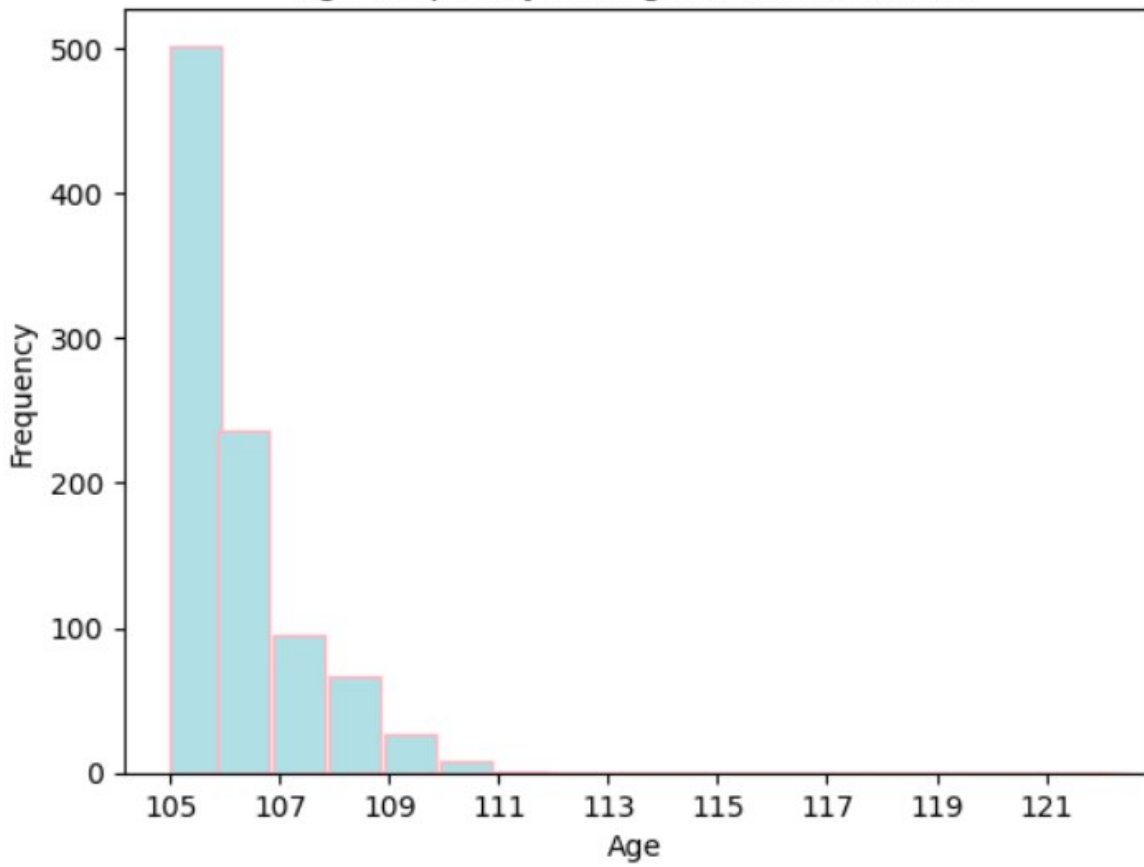
country	count
Франція	10662
Велика Британія	1139
Німеччина	955
США	797
Бельгія	575
Данія	453
Канада	342
Австрія	230
Норвегія	216
Японія	78

Оскільки, як ми побачили в попередньому розділі, кількість довгожителюк значно переважає кількість довгожителів, можна підозрювати, що і шуканий розподіл для різних статей може відрізнятись. Зобразимо гістограми розподілу для жінок і чоловіків із кількох країн окремо, а також гістограми для даних із бази C.V.S.(чоловіки та жінки із Нідерландів), після досягнення 95 років :

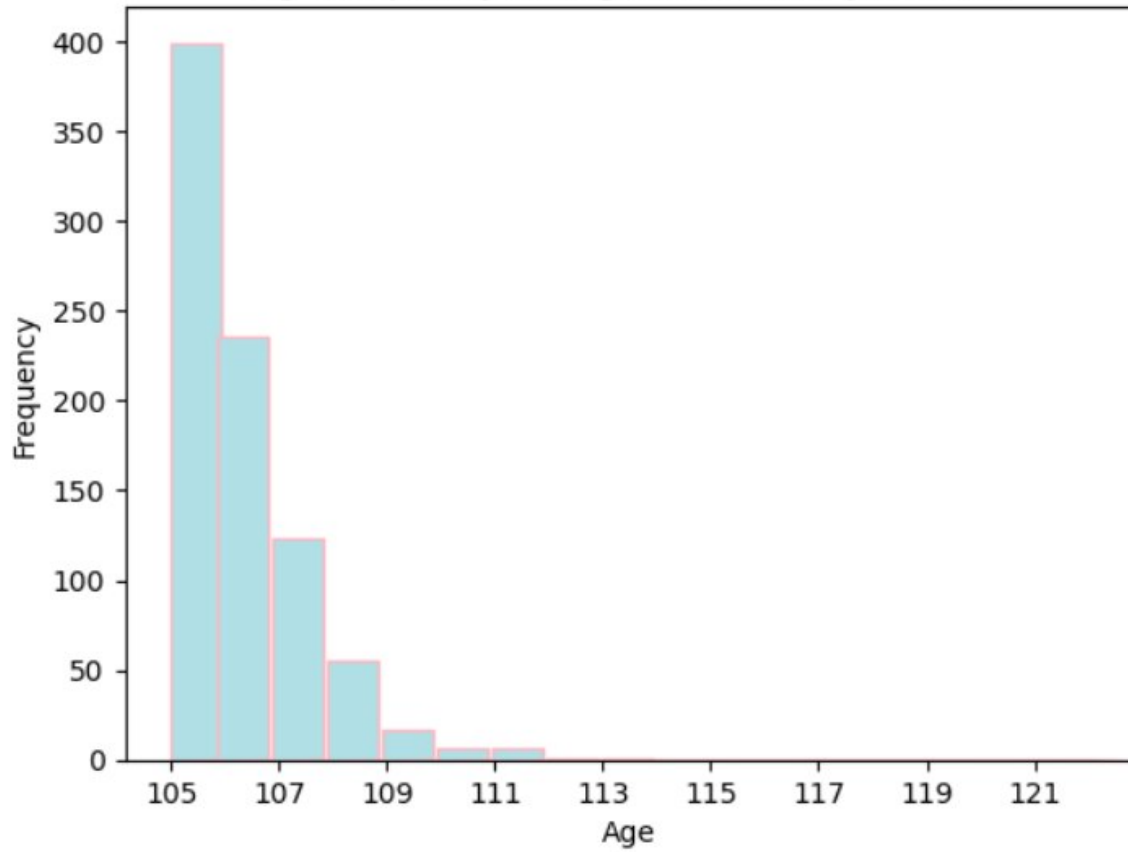
Age frequency Histogram (France female)



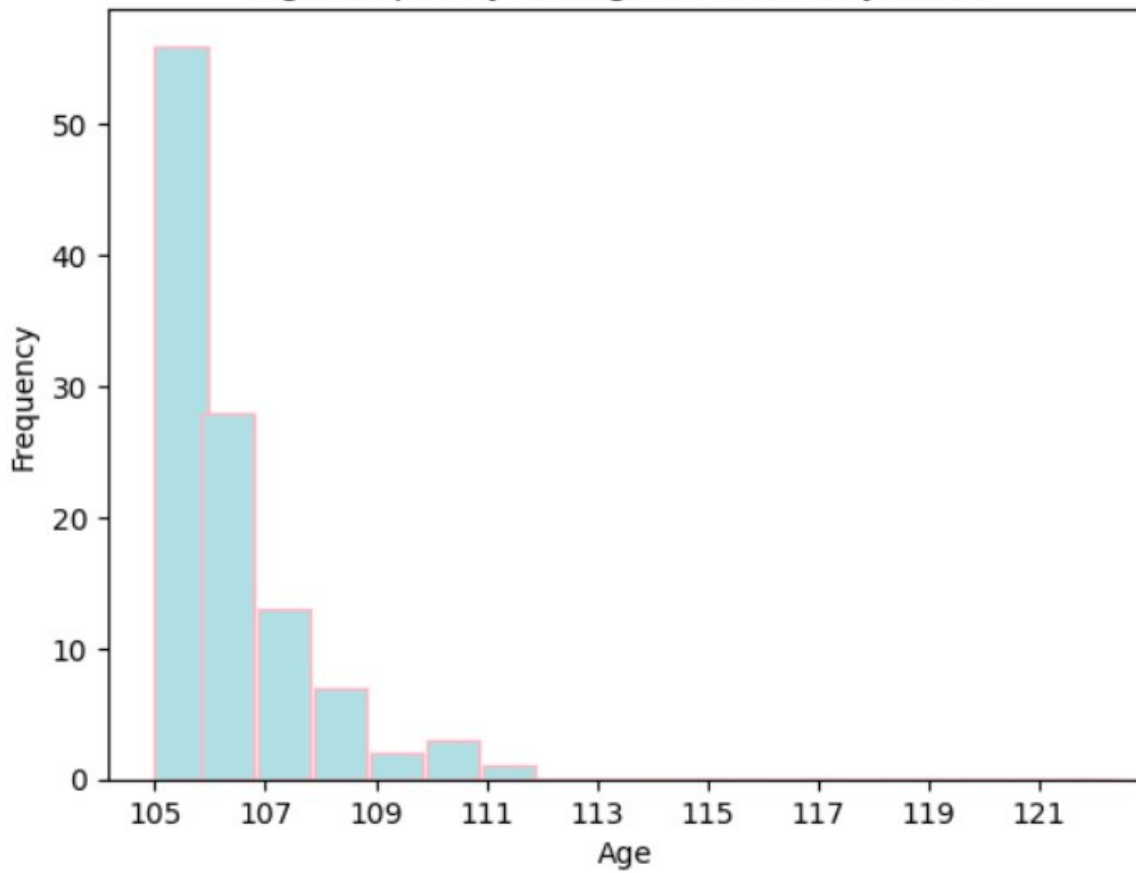
Age frequency Histogram (France male)



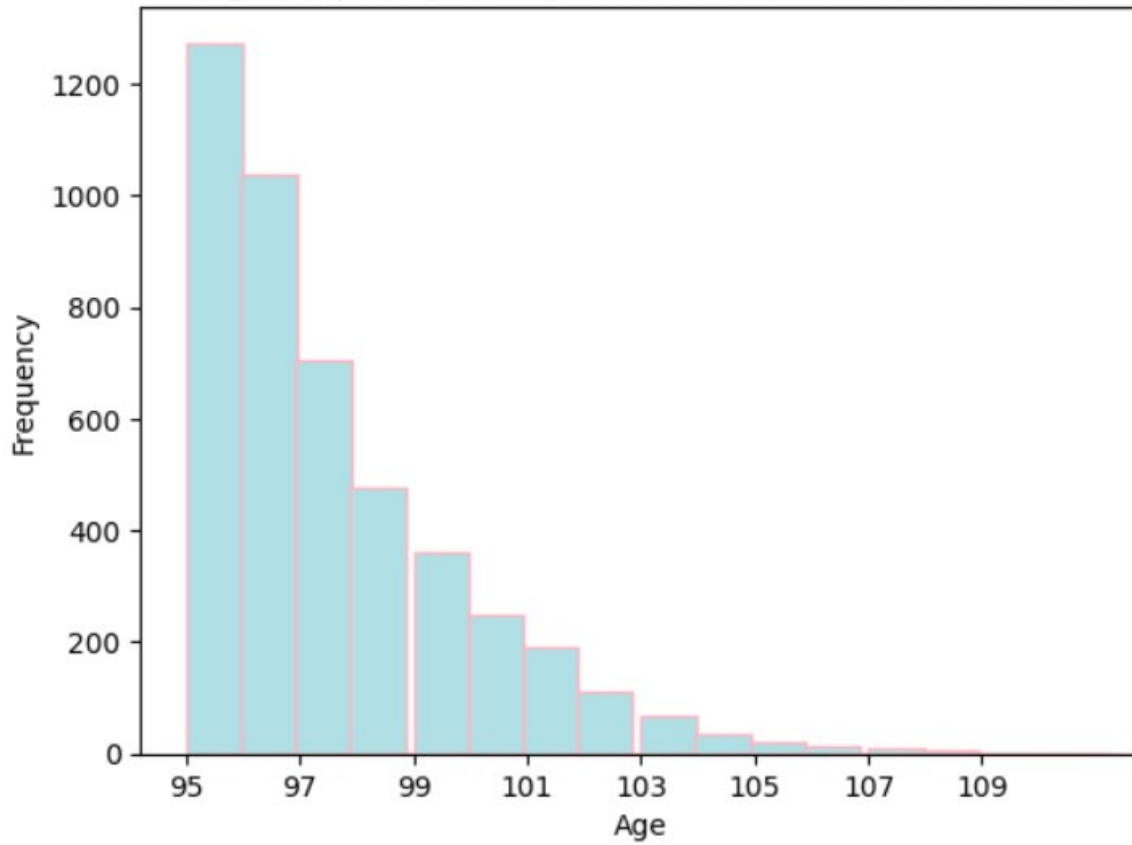
Age frequency Histogram (Germany female)



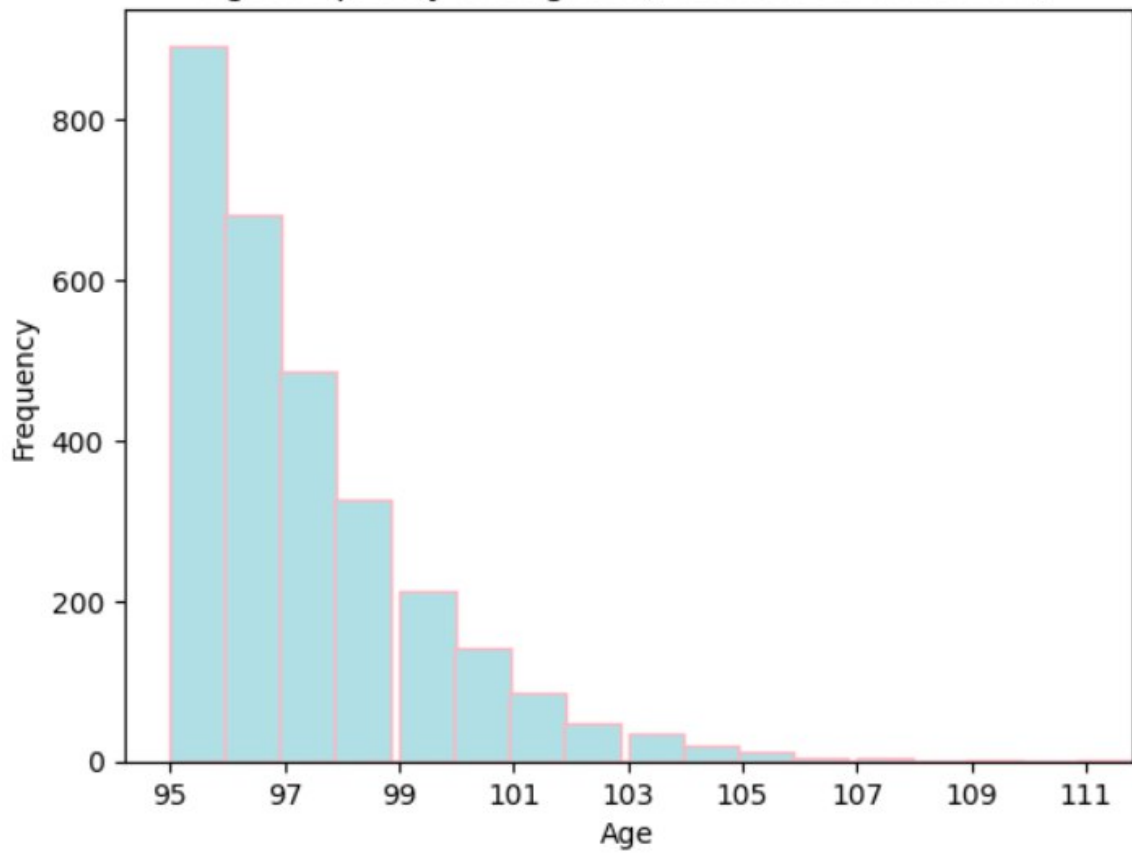
Age frequency Histogram (Germany male)



Age frequency Histogram (Netherlands female 95+)



Age frequency Histogram (Netherlands male 95+)



Код для проведення маніпуляцій з даними та побудови гістограм мовою Python можна переглянути в додатку Б.

З гістограм помітно деякі відмінності між розподілами віку жінок та чоловіків, а також між розподілами для даних із IDL (від 105 років) та C.V.S. (після 95 років), проте також помітно, що для всіх них є підстави перевірити згоду з експоненціальним розподілом.

Проблеми моделювання розподілу

Однією з ключових задач у статистичному моделюванні є відповідність даних певному розподілу, з якою ми переважно і працюватимемо у цій роботі. Для вирішення цієї задачі використовуються статистичні критерії, коротко пригадаємо суть найпоширеніших:

1) Критерій Колмогорова-Смірнова: цей критерій порівнює емпіричну функцію розподілу даних з теоретичною функцією розподілу. Він базується на максимальній відстані між цими двома функціями. Якщо ця відстань мала, то ми можемо стверджувати, що дані відповідають теоретичному розподілу.

2) Критерій χ^2 -квадрат: цей критерій порівнює фактичну частоту спостережень в кожному з k інтервалів з теоретичною частотою, яку можна отримати з моделі розподілу. Якщо отримане значення критерію χ^2 -квадрат є меншим за критичне значення з таблиць квантилів розподілу χ^2 -квадрат з $k-1$ ступенів вільності, то ми можемо прийняти гіпотезу про відповідність даних розподілу.

3) Критерій Андерсона-Дарлінга: цей критерій також порівнює емпіричну функцію розподілу даних з теоретичною функцією розподілу. Він базується на сумі вагованих квадратів відхилень між цими двома функціями. Якщо отримане значення критерію менше за критичне значення з таблиць квантилів Андерсона-Дарлінга, то ми можемо прийняти гіпотезу про відповідність даних розподілу.

У цій роботі ми працюватимемо переважно із критерієм χ^2 -квадрат. Проте, для використання статистичних критеріїв, маю-

чи набір даних, потрібно визначитись із типом розподілу, який будемо перевіряти, і його параметрами. І якщо з першим завданням справитись не так складно, то вирішення другого зовсім не є очевидним, і породжує проблему підбору параметрів.

Один з підходів до вирішення цієї задачі полягає у використанні статистичних методів оцінки параметрів, наприклад, методу максимальної правдоподібності, що вимагає визначення функції правдоподібності, яка описує ймовірність того, що наші дані були отримані з певного розподілу з деякими параметрами. Потім ми шукаємо значення параметрів, які максимізують цю функцію правдоподібності.

Однак, в статистичних методах оцінки параметрів часто необхідно виконувати операції диференціювання та інтегрування, що може бути складним в програмній реалізації, зокрема якщо ми маємо справу зі складними функціями.

Тому однією із цілей цієї роботи буде створення та використання чисельного алгоритму підбору оптимальних параметрів будь-якого розподілу для мінімізації χ^2 -квадрат статистики, що не використовуватиме складних обчислень з операціями диференціювання чи інтегрування, але працюватиме швидше ніж метод простого перебору, і буде достатньо точним.

Алгоритм оптимізації

Для початку введемо точне формулювання поставленої задачі:

нехай відомо, що до T років дожило n осіб, і після T років на інтервалах $(0, 1), (1, 2), \dots, (i - 1, i), \dots, (m - 1, m)$ померло $n_1, n_2, \dots, n_i, \dots, n_m$ осіб відповідно.

Запишемо статистику χ^2 як функцію від параметрів розподілу $\lambda_1, \lambda_2, \dots, \lambda_m$:

$$\chi^2 = \chi^2(\lambda_1, \lambda_2, \dots, \lambda_m) = \sum_{i=1}^m \frac{(n_i - np_i)^2}{np_i},$$

де $\sum_{i=1}^m n_i = n$,

$p_i = \psi(i, \lambda_1, \lambda_2, \dots, \lambda_m) - \psi(i - 1, \lambda_1, \lambda_2, \dots, \lambda_m)$ - ймовірності потрапляння в i -й інтервал згідно функції розподілу $\psi(\cdot)$,

$i = 1, 2, \dots, m$.

Будемо розв'язувати задачу мінімізації значення статистики хі-квадрат:

$$\chi^2(\lambda_1, \lambda_2, \dots, \lambda_m) \rightarrow \min_{\lambda_1, \lambda_2, \dots, \lambda_m},$$

при цьому $\lambda_1, \lambda_2, \dots, \lambda_m > 0$.

Якщо $\lambda_1^*, \lambda_2^*, \dots, \lambda_m^*$ і $\chi^2(\lambda_1^*, \lambda_2^*, \dots, \lambda_m^*)$ - оптимальний розв'язок задачі мінімізації, - $\chi^2(\lambda_1^*, \lambda_2^*, \dots, \lambda_m^*)$ порівнюється з критичним значенням хі-квадрат χ_{kr}^2 з обраним рівнем значущості, і при:

$\chi^2(\lambda_1^*, \lambda_2^*, \dots, \lambda_m^*) < \chi_{kr}^2$ - гіпотеза про вид розподілу приймається,

$\chi^2(\lambda_1^*, \lambda_2^*, \dots, \lambda_m^*) > \chi_{kr}^2$ - відхиляється.

Отже, фактично ми будемо шукати такі параметри $\lambda_1^*, \lambda_2^*, \dots, \lambda_m^*$ обраного розподілу, при яких значення хі-квадрат статистики на заданому наборі даних буде мінімальним.

Для цього введемо позначення:

$I_k = (b_k, c_k)$ - проміжки, в яких лежать значення λ_k , $b_k \geq 0$,
 $\lambda_k \in I_k$, $k = \overline{1, m}$,

де m - кількість параметрів розподілу.

Значення меж інтервалів I_k буде задавати користувач.

На кожному кроці алгоритму будемо обраховувати значення $\Delta_k = c_k - b_k$ - довжини інтервалів I_k ,

Починаючи з другого кроку, межі інтервалів і відповідні значення Δ_k будемо перераховувати згідно формул:

$$b_k = \lambda_k - \Delta_k/4$$

$$c_k = \lambda_k + \Delta_k/4,$$

І перевірятьимемо для кожного b_k , чи b_k не є меншим нуля (таке може відбуватись при виборі занадто великого початкового проміжку I_k), якщо так - b_k присвоюємо 0.

1) далі для кожного λ_k обрахуємо значення $\lambda_k^i = b_k + u_k^i \Delta_k$, де $i = \overline{1, N}$, $N = 10^p$, значення p можна підібрати емпірично, від нього залежатиме точність та швидкість збіжності алгоритму, u_k^i - випадкове значення, згенероване із рівномірного розподілу на проміжку $(0, 1)$,

2) для кожного набору λ_k обчислимо значення $\chi^2(\lambda_1^i, \lambda_2^i, \dots, \lambda_m^i)$,

3) знайдемо $\min_{1 \leq i \leq N} \chi^2(\lambda_1^i, \lambda_2^i, \dots, \lambda_m^i) = \chi^2(\lambda_1^*, \lambda_2^*, \dots, \lambda_m^*)$

4) Для першого кроку оптимальним значенням χ^2 і параметрів будемо вважати $\chi^2(\lambda_1^*, \lambda_2^*, \dots, \lambda_m^*)$ і $\lambda_1^*, \lambda_2^*, \dots, \lambda_m^*$, для наступних кроків оптимальними будемо вважати значення χ^2 і параметрів, обрахованих на даному кроці, якщо χ^2 менше за χ^2 , обраховане на попередньому кроці, інакше - оптимальні значення χ^2 і параметрів із попереднього кроку.

Умовою зупинки можна обрати досягнення обраної кількості

ітерацій, або досягнення умови $|\chi_{i*}^2 - \chi_{i-1*}^2| < \alpha$, де $\alpha > 0$ - деяке досить мале число.

Таким чином, фактично на кожному кроці алгоритму, завдяки множнику із рівномірного розподілу обиратиметься найоптимальніше значення статистики хі-квадрат для заданого проміжку, після чого проміжки звужуватимуться навколо попереднього оптимального значення параметрів, внаслідок чого, очевидно, в результаті отримаємо параметри, що максимально близькі до тих, за яких значення χ^2 буде мінімальним.

Алгоритм реалізовано у додатку А.

Розподіл Вейбула

Розподіл Вейбула є одним з найбільш поширених розподілів у статистиці, який застосовується для моделювання різних процесів в різних областях. Цей розподіл зазвичай використовується для аналізу даних залежності часу до відмови (або життєздатності) об'єкта від різних факторів, таких як напруга, температура, вологість тощо.

Розподіл Вейбулла широко використовується в наукових дослідженнях, наприклад, для аналізу даних про тривалість життя людей після досягнення певного віку, для прогнозування тривалості роботи електронних пристроїв, для аналізу тривалості лікування пацієнтів, та інших сфер життєдіяльності.

Одним з головних показників, що використовуються для характеристики розподілу Вейбула, є параметр форми, який визначає форму розподілу. Цей параметр може бути як додатним, так і від'ємним, в залежності від форми розподілу. Наприклад, для додатного параметра форми розподіл Вейбула має скісну форму, а для від'ємного - зворотно-скісну.

Функція щільності ймовірності розподілу Вейбула має наступний вигляд:

$$f(x) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}, & x \geq 0 \\ 0, & x < 0 \end{cases},$$

де λ - параметр масштабу, k - параметр форми.

Моменти розподілу Вейбула можна обчислити з використанням гамма-функції. Зокрема, математичне сподівання розподілу

Вейбула має наступний вигляд: $E(X) = \lambda\Gamma(1 + \frac{1}{k})$

Дисперсія розподілу Вейбула може бути обчислена за формулою:

$$D(X) = \lambda^2(\Gamma(1 + 2/k) - (\Gamma(1 + 1/k))^2)$$

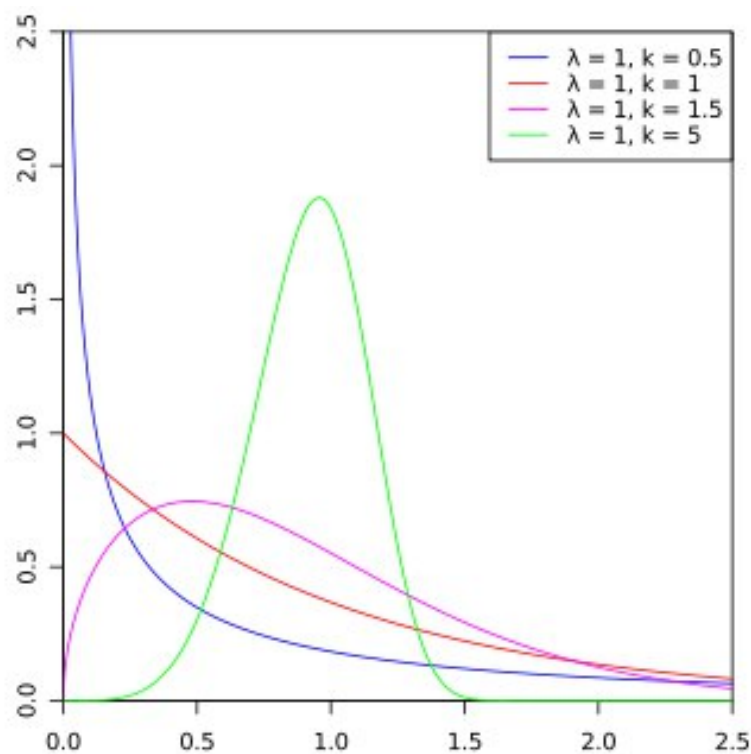
де λ і k - параметри розподілу, а $\Gamma(\cdot)$ - гамма-функція.

Функція розподілу визначається як ймовірність того, що випадкова величина X буде меншою за заданий поріг x :

$$F(x) = P(X \leq x) = 1 - e^{-(\frac{x}{\lambda})^k}$$

де μ - математичне сподівання.

Графік функції щільності розподілу Вейбула залежно від параметрів λ та k :



Нескладно також помітити, що при $k = 1$ розподіл Вейбула збігається з експоненційним розподілом, тобто експоненційний розподіл є частинним випадком розподілу Вейбула. Зв'язок

між розподілом Вейбула і експоненціальним розподілом важливий, оскільки якщо є підозра, що певний шуканий розподіл має більш складну структуру, ніж простий експоненціальний, то можна розглянути використання розподілу Вейбула з параметрами, які дають необхідну складність розподілу.

Розподіл Вейбула може використовуватись, зокрема, для опису розподілу залишкового часу життя після настання певного віку, оскільки він добре описує ситуації, коли імовірність відмови збільшується з часом. В цьому контексті, x буде залишковим часом життя після настання певного віку, λ - параметр масштабу, що визначає середнє значення залишкового часу життя, k - параметр форми, що визначає форму розподілу. Параметр k визначає, наскільки швидко розподіл зменшується зі збільшенням залишкового часу життя.

Розподіл Гомперца

Розподіл Гомперца є одним із статистичних розподілів, який використовується для моделювання залишкового часу життя людей та інших живих організмів. Цей розподіл названий на честь британського демографа та статистика Роберта Гомперца, який вперше запропонував використовувати цей розподіл у 1949 році для моделювання шансів на смерть у певному віці.

Ще однією цікавою особливістю розподілу Гомперца є те, що він може бути використаний для моделювання ризику в разі наявності двох або більше факторів, що впливають на ризик. У такому випадку, розподіл Гомперца можна розглядати як розширення експоненційного розподілу.

У розподілі Гомперца є два параметри: параметр форми та параметр масштабу. Параметр форми відповідає за форму розподілу, а параметр масштабу відповідає за середній залишковий час життя.

Функція розподілу має наступний вигляд:

$$F(x) = 1 - e^{-\frac{\beta}{\gamma}(e^{\gamma x} - 1)}$$

Функція щільності:

$$f(x) = \begin{cases} \beta e^{\beta/\gamma + \gamma x - \beta/\gamma e^{\gamma x}}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Значення параметрів $\beta, \gamma > 0$.

Проте, розподіл Гомперца зазвичай застосовується для аналізу даних про людей віком 50-60 років і старших. Зокрема, у

статті "Heterogeneity in old age mortality: an international comparison"(1998), Vaupel та його співавтори аналізували дані про залишковий час життя для людей віком 60 років та старших у 14 країнах світу. Вони використовували розподіл Гомперца для опису залежності між інтенсивністю смертності та віком, та зробили висновок, що розподіл Гомперца є ефективним інструментом для моделювання залишкового часу життя після віку 60 років. У статті "Modeling the Age Pattern of Mortality for Comparative Mortality Analysis"(2006), Герман Піццерле та його співавтори використовували розподіл Гомперца, щоб моделювати залишковий час життя для людей віком від 50 років та старших, враховуючи різні ефекти, такі як ефект статі, ефект календарного періоду та ефект країни. Також, стаття "Modelling the distribution of the length of life"(1992) Гейла МакКіннона та Роберта Лі наводить приклад використання розподілу Гомперца для аналізу залишкового часу життя для людей віком 65 років та старших в Сполучених Штатах.

Отже, розподіл Гомперца застосовується переважно для аналізу залишкового часу життя після віку 50-60 років та старших, і наразі, імовірно, не існує відомих досліджень, що підтвержували б доцільність використання розподілу Гомперца для опису даних вікових груп, що нас цікавлять - від 90 і більше років, тому ми спробуємо використати також цей розподіл для дослідження.

Дослідження розподілу

Отже, використаємо базу даних IDL, і для декількох країн для кожної статі створимо датафрейми для дослідження. Для кожного з них використаємо колонку 'agedays' - вік у днях, після досягнення якого наступила смерть. Код, що реалізує усі описані далі маніпуляції наведено у додатках А і Б.

Аналізуватимемо усіх довгожителів, які померли у віці більше 105 років, тому від даного значення віднімемо $365 * 105 + 25$, аби отримати значення залишкового часу життя, після досягнення 105 років (з врахуванням очікуваної кількості високосних років).

Отримані дані будемо групувати за деяким розміром інтервалу у днях, підбраному емпірично з урахуванням об'єму вибірки, аби отримати достатню кількість групованих інтервалів.

Проміжки, в які потрапляє менше ніж 10 значень будемо об'єднувати між собою (що необхідно для коректного використання критерію χ^2 -квадрат).

Таким чином, наприклад, для вибірки жінок-довгожительок із Франції, до якої входить 9727 значень, будемо групувати за інтервалами розміру 150 днів, і отримаємо:

interval	count
(0, 1)	2204
(1, 2)	1799
(2, 3)	1427
(3, 4)	1039
(4, 5)	823
(5, 6)	626
(6, 7)	464
(7, 8)	340
(8, 9)	282
(9, 10)	213
(10, 11)	143
(11, 12)	110
(12, 13)	76
(13, 14)	45
(14, 15)	42
(15, 16)	20
(16, 17)	25
(17, 18)	19
(18, 21)	17
(21, 43)	13

Згідно описаного в одному із попередніх розділів алгоритму оптимізації хі-квадрат статистики, перевіримо значення хі-квадрат статистики на цих даних для експоненційного розподілу, розподілу Вейбула та розподілу Гомперца.

Для експоненційного розподілу задамо інтервал пошуку параметра λ як проміжок $(0, 1)$, для розподілу Вейбула: $\lambda \in (0, 1)$, $\alpha \in (0, 2)$, для розподілу Гомперца: $\beta \in (0, 1)$, $\gamma \in (0, 1)$.

Отримані значення мінімізованої хі-квадрат статистики та відповідні параметри розподілів:

експоненційний: $\chi^2 = 59.0832027504306$, $\lambda = 0.2796774544015044$,
Вейбула: $\chi^2 = 18.645482053509642$, $\lambda = 0.2533831492792398$, $\alpha = 1.060919160767153$,

Гомперца: $\chi^2 = 17.580242839219227$, $\beta = 0.2614318762423342$, $\gamma = 0.0210347382913875$.

Кількість ступенів свободи обраховуємо як різницю кількості інтервалів, використаних для розбиття даних, та параметрів розподілу -1:

$$df_{exp} = 20 - 1 - 1 = 18,$$

$$df_{weib} = 20 - 2 - 1 = 17,$$

$$df_{gomr} = 20 - 2 - 1 = 17.$$

Згідно з таблицею критичних значень хі-квадрат, що розміщена в додатку В для рівня значущості 0.05:

гіпотеза про експоненційний розподіл відхиляється, оскільки є більшою за 30.14,

гіпотеза про розподіл Вейбула не відхиляється, оскільки є меншою за критичне значення - 28.87,

гіпотеза про розподіл Гомперца також не відхиляється, оскільки є меншою за критичне значення.

Перевіримо, наскільки збігатимуться ці результати із результатами для чоловіків-довгожителів із Франції аналогічно (розмір вибірки - 935, розмір інтервалів групування - 100, кількість інтервалів - 17):

експоненційний: $\chi^2 = 20.917508965422623$, $\lambda = 0.20620651131851092$,
Вейбула: $\chi^2 = 20.840954687192436$, $\lambda = 0.20310461139872982$, $\alpha = 1.0087907443501014$,

Гомперца: $\chi^2 = 19.8219235872399320$, $\beta = 0.19831284304271043$, $\gamma = 0.0119254321782155$.

Кількість ступенів свободи:

$$df_{exp} = 17 - 1 - 1 = 15,$$

$$df_{weib} = 17 - 2 - 1 = 14,$$

$$df_{gomp} = 17 - 2 - 1 = 14.$$

Згідно таблиці критичних значень, гіпотезу про вид розподілу не можна відхилити ні для експоненційного, ні для розподілу Вейбула, ні для розподілу Гомперца. Отже, ці дані, на відміну від попередніх, можуть бути розподілені експоненційно.

Проведемо експерименти і для інших країн із бази даних IDL аналогічно, результати підсумуємо у таблиці:

країна	стать	к-сть	розподіл	df	χ^2	$\chi^2_{critical}$	параметри
FR	ж	9727	exp	18	59.0832	28.87	0.2797
FR	ж	9727	weibull	17	18.6455	27.59	0.2534; 1.0609
FR	ж	9727	gompertz	17	17.5802	27.59	0.2614; 0.0210
DEU	ж	845	exp	14	15.5302	23.68	0.2307
DEU	ж	845	weibull	13	6.6067	22.36	0.1941; 1.1016
DEU	ж	845	gompertz	13	6.3577	22.36	0.2062; 0.0329
FR	ч	935	exp	15	20.9175	25.00	0.2062
FR	ч	935	weibull	14	20.8410	23.68	0.2031, 1.0088
FR	ч	935	gompertz	14	19.8219	23.68	0.1983, 0.0119
UK	ч	314	exp	9	11.5960	16.92	0.2892
UK	ч	314	weibull	8	11.5032	15.51	0.2818, 1.0173
UK	ч	314	gompertz	8	11.2432	15.51	0.2777, 0.0161
DEU	ч	110	exp	5	4.5542	11.07	0.2942
DEU	ч	110	weibull	4	4.0054	9.49	0.2658, 1.0769
DEU	ч	110	gompertz	4	4.1130	9.49	0.2723, 0.0371

Підсумовуючи, бачимо такі тенденції:

1) у розглянутих країнах для жінок мінімізовані χ^2 -квадрат статистики для розподілу Вейбула та розподілу Гомперца є у декілька разів меншими, ніж для експоненційного, і є меншими за критичне значення, на відміну від експоненційного, проте, для Німеччини, гіпотеза про експоненційний розподіл не відкидається, але зважаючи на різницю у χ^2 -квадрат статистиках, гіпотезу про експоненційний розподіл залишкового часу життя для жінок віку від 105 років можна відкинути, для розподілу ж Вейбула та розподілу Гомперца - прийняти, причому різниця у χ^2 -квадрат статистиках для розподілу Гомперца та розподілу Вейбула є незначною, проте величина є меншою для розподілу Гомперца;

2) для чоловіків ситуація значно відрізняється: різниця між χ^2 -квадрат статистиками для усіх трьох розподілів є дуже незначною, і зменшується зі зростанням кількості даних у вибірці. Окрім цього, зокрема, параметр α розподілу Вейбула є дуже близьким до одиниці (а як ми знаємо з попередніх розділів - при $\alpha = 1$ розподіл Вейбула співпадає з експоненційним). І згідно таблиці критичних значень, гіпотеза про вид розподілу не відкидається ні для розподілу Вейбула, ні для експоненційного, ні для розподілу Гомперца, проте можна стверджувати, що для чоловіків розподіл залишкового часу життя, зокрема, задовольняє експоненційний розподіл і експоненційний розподіл для них є найбільш характерним, на відміну від результатів для жінок;

3) при цьому якщо приймати за розподіл залишкового часу життя чоловіків експоненційний розподіл - параметр цього розподілу коливається від 0.2062 до 0.2942 для обраних даних, звідси, враховуючи розмір інтервалів групування, математичне сподівання, що обчислюється як λ^{-1} складає: 450 (Велика Британія), 485 (Франція), 543 (Німеччина) днів, або від 1.23 до 1.49 року, що від-

повідас очікуваному часу, прожитому після досягнення 105 років;

4) для жінок, аналогічно згідно розподілу Вейбула обчислимо математичне сподівання за формулою $1/\lambda(\Gamma(1+1/k))$, отримаємо 738 (Франція) та 596 (Німеччина) днів, або 1.63 - 2.02 роки, що зокрема, перевищує очікуваний залишковий час життя чоловіків.

Тепер розглянемо довгожителів, починаючи з досягнення 90 та 95 років, із бази даних C.V.S, тут розмір одного інтервалу групування - 1 рік. Результати:

вік	стать	к-сть	розподіл	df	χ^2	$\chi^2_{critical}$	параметри
від 95	ж	4559	exp	11	55.5003	19.68	0.3629
від 95	ж	4559	weibull	10	18.4114	18.31	0.3226; 1.1003
від 95	ж	4559	gompertz	10	14.4880	18.31	0.3267; 0.0551
від 95	ч	2947	exp	10	37.4972	18.31	0.4071
від 95	ч	2947	weibull	9	3.4274	16.92	0.3552; 1.1143
від 95	ч	2947	gompertz	9	4.4974	16.92	0.3620; 0.0576
від 90	ж	3943	exp	13	125.1683	22.36	0.2665
від 90	ж	3943	weibull	12	20.0104	21.03	0.2062; 1.1791
від 90	ж	3943	gompertz	12	12.3497	21.03	0.2217; 0.0710
від 90	ч	2736	exp	12	79.9819	21.03	0.2948
від 90	ч	2736	weibull	11	16.4015	19.68	0.2356; 1.1659
від 90	ч	2736	gompertz	11	12.0781	19.68	0.2501; 0.0678

Тут, як бачимо і для жінок, і для чоловіків гіпотеза про експоненційний розподіл однозначно відхиляється, а гіпотези про розподіл Вейбула та розподіл Гомперца - переважно приймаються, окрім вибірки жінок від 95 років - для них приймається тільки гіпотеза про розподіл Гомперца. Також, різниця між статистиками розподілу Гомперца та Вейбула стала більшою - для усіх вибірок, окрім чоловіків від 95 років статистика для розподі-

лу Гомперца є меншою, при однаковій кількості ступенів свободи.

З цього можемо зробити висновок, що для жінок віком від 95 і від 90 років найбільш характерним є розподіл Гомперца (від 90 - з параметрами 0.22, 0.07; від 95 - 0.33, 0.06), а для чоловіків від 95 років - розподіл Вейбула з параметрами 0.36, 1.11; від 90 - розподіл Гомперца з параметрами 0.25, 0.07.

При цьому математичне сподівання залишкового часу життя після досягнення 95 років складає:

2.99 років для жінок,

2.7 років для чоловіків.

Після 90 років:

4.58 років для жінок,

4.02 років для чоловіків.

Висновки

Отже, в даній роботі:

- було розглянуто залишкову тривалість життя як кількісну проблему, що може бути вирішена за допомогою статистичного аналізу даних,

- було розглянуто методи моделювання розподілу залишкового часу життя людини, реалізовано універсальний алгоритм підбору параметрів розподілу, що мінімізують значення χ^2 -квадрат статистики,

- було досліджено та підібрано розподіли та математичне сподівання очікуваного залишкового часу життя після досягнення віку x залежно від значення x у випадках 90, 95 та 105 років для різних країн, а також залежно від статі.

Підсумовуючи, можна зробити такі загальні висновки:

Для чоловіків смертність із віком при екстремальних значеннях зростає стрімкіше, оскільки починаючи від 105 років вік їх смерті задовольняє експоненційний розподіл, на відміну від жінок, для яких найбільш характерним є розподіл Гомперца, що може свідчити, зокрема про те, що вони мають складнішу закономірність тривалості життя.

Крім цього, для віку смерті жінок характерним є розподіл Гомперца і від 105, і від 95, і від 90 років, для чоловіків же - розподіл Гомперца є найбільш характерним лише від 90 років, а від 95 - розподіл Вейбула.

Зважаючи на це, а також на обраховані в роботі математичні сподівання прожитого віку для жінок та чоловіків для розподілів зі знайденими оптимальними параметрами, можна зробити висновок, що жінки в середньому живуть довше, стрімкість зростання їхньої смертності з віком є меншою, ніж у чоловіків. Ризик смерті для чоловіків збільшується швидше, ніж у жінок, починаючи принаймні від 90 років.

Література

[1] Rootzen H., Zholud D. Human life is unlimited – but short. *Extremes* 20, 713–728 (2017). <https://doi.org/10.1007/s10687-017-0305-5>.

[2] Aarssen K., Haan L. (1994). On the maximal life span of humans. *Mathematical Population Studies* 4, 259-281. 2.

[3] Gavrilov, L. A., and Gravitova, N. S. (1991) *The biology of life span, a quantitative approach*. Hardwood Academic Publishers, Chur.

[4] Vaupel, J. W. (2010). Biodemography of human ageing. *Nature* 464 (25), 536-542. 2, 11.

[5] Weon, B. M. and J. H. Je (2009). Theoretical estimation of maximum human lifespan. *Biogerontology* 10, 65-71. 2.

Додатки

Додаток А

```
import numpy as np
import pandas as pd

class ChisquareOptimizer:

    def calculate_chi_square_statistics(self, n_i, p_i, n):
        chi_square = 0
        for i in range(len(n_i)):
            i_part = ((n_i[i] - n*p_i[i])**2)/(n*p_i[i])
            chi_square += i_part
        return chi_square

    def weibull_distribution_function(self, x, params):
        f = 1 - np.exp(-params[0]*(x**params[1]))
        return f

    def exponential_distribution_function(self, x, params):
        f = 1 - np.exp(-params[0]*(x))
        return f

    def gompertz_distribution_function(self, x, params):
        f = 1 - np.exp(-(params[0]/params[1])*
            (np.exp(params[1]*x)-1))
        return f
```

```

def calculate_p_i(self, params, dct, distribution: str):
    intervals = list(dct.keys())
    p_i_list = []
    for interval in intervals:
        if distribution == 'exponential':
            p = self.exponential_distribution_function
                (interval[1], params) -
            self.exponential_distribution_function
                (interval[0], params)
        elif distribution == 'weibull':
            p = self.weibull_distribution_function
                (interval[1], params) -
            self.weibull_distribution_function
                (interval[0], params)
        elif distribution == 'gompertz':
            p = self.gompertz_distribution_function
                (interval[1], params) -
            self.gompertz_distribution_function
                (interval[0], params)
        p_i_list.append(p)
    return p_i_list

def generate_random_uniform(self, size=2):
    generated = np.random.uniform(0, 1, size)
    return list(generated)

def count_n(self, dct):
    n_i = list(dct.values())
    n = sum(n_i)
    return n, n_i

```

```

def minimize_chisquare(self, distribution, dct,
intervals, iter_num = 10000):
    n, n_i = self.count_n(dct)
    params_num = len(intervals)
    for j in range(1, 11):
        deltas = []
        for interval in intervals:
            deltas.append(interval[1]-interval[0])

    if j != 1:
        betas = opt_betas
        b = []
        c = []
        for i in range(len(intervals)):
            b.append(betas[i] - deltas[i]/4)
            c.append(betas[i] + deltas[i]/4)

        for i in range(params_num):
            if b[i] < 0:
                b[i] = 0

        intervals = []
        deltas = []
        for i in range(params_num):
            intervals.append([b[i], c[i]])
            deltas.append(c[i] - b[i])

    chi_squares = []
    betas_list = []

```

```

for i in range(1, iter_num):
    u = self.generate_random_uniform(params_num)
    betas = []
    for i in range(params_num):
        betas.append
            (intervals[i][0] + deltas[i]*u[i])
    p_i = self.calculate_p_i
        (betas, dct, distribution)
    chi_square =
self.calculate_chi_square_statistics
        (n_i, p_i, n)
    chi_squares.append(chi_square)
    betas_list.append(betas)

min_chi_square = min(chi_squares)
min_betas = betas_list[np.argmin(chi_squares)]
print(f"step_{j}: {min_chi_square}")

if j == 1:
    opt_chi_square = min_chi_square
    opt_betas = min_betas

elif j != 1:
    if min_chi_square < opt_chi_square:
        opt_chi_square = min_chi_square
        opt_betas = min_betas

print('Optimal chisquare statistics and betas:')
print(opt_chi_square, opt_betas)
return opt_chi_square, opt_betas

```

Додаток Б

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from optimizer.optimizer_class import ChisquareOptimizer

optimizer = ChisquareOptimizer()
df = pd.read_csv('idl_complete_death_105.csv')
df.head()

df['BIRTH_COUNTRY'].value_counts().head()

df_france = df[df['BIRTH_COUNTRY'] == 'FRA']
df_uk = df[df['BIRTH_COUNTRY'] == 'UK']
df_germany = df[df['BIRTH_COUNTRY'] == 'DEU']
df_usa = df[df['BIRTH_COUNTRY'] == 'USA']

lst=df_germany[df_germany['SEX']=='F']['AGEYEARS'].tolist()

%matplotlib inline

x = lst
plt.hist(x, bins=100, range = (105, 122), width = 0.98,
color = "powderblue", ec="lightpink")
plt.xticks([105, 107, 109, 111, 113, 115, 117, 119, 121])
plt.gca().set(title='Age frequency Histogram
(France female 105+)', ylabel='Frequency')
plt.xlabel("Age")
```

```

plt.ylabel("Frequency")

age_france_male = list(df_france[df_france['SEX'] == 'M']
['AGEDAYS'] - 365*105-25)
age_france_female = list(df_france[df_france['SEX'] == 'F']
['AGEDAYS'] - 365*105-25)

age_uk_male = list(df_uk[df_uk['SEX'] == 'M']
['AGEDAYS'] - 365*105-25)

age_germany_male = list(df_germany[df_germany['SEX'] == 'M']
['AGEDAYS'] - 365*105-25)
age_germany_female = list(df_germany[df_germany['SEX'] == 'F']
['AGEDAYS'] - 365*105-25)

age_usa_male = list(df_usa[df_usa['SEX'] == 'M']
['AGEDAYS'] - 365*105-25)
age_usa_female = list(df_usa[df_usa['SEX'] == 'F']
['AGEDAYS'] - 365*105-25)

def create_dict_data(age_df, interval_size):
    dct = {}
    max_value = max(age_df)
    num_intervals = max_value // interval_size + 1

    for i in range(num_intervals):
        interval_start = interval_size * i
        interval_end = interval_start + interval_size
        interval_end = min(max_value + 1, interval_end)
        dct[(i, i+1)] = sum([1 for elem in age_df

```

```
        if interval_start <= elem < interval_end])
dct = {k: v for k, v in dct.items() if v != 0}
return dct
```

```
france_female_dct = create_dict_data
(age_france_female, interval_size = 150)
france_male_dct = create_dict_data
(age_france_male, interval_size = 100)
germany_female_dct = create_dict_data
(age_germany_female, interval_size = 120)
germany_male_dct = create_dict_data
(age_germany_male, interval_size = 160)
uk_male_dct = create_dict_data
(age_uk_male, interval_size = 130)
```

```
cbs_male_95 = {(0,1): 892, (1,2): 682, (2,3): 485,
(3,4): 325, (4,5): 211, (5,6): 140, (6,7): 85,
(7,8): 47, (8,9): 36, (9,10): 19, (10,11): 13,
(11,17): 12}
```

```
cbs_female_95 = {(0,1): 1273, (1,2): 1038, (2,3): 705,
(3,4): 476, (4,5): 362, (5,6): 249, (6,7): 192,
(7,8): 110, (8,9): 66, (9,10): 35, (10,11): 22,
(11,12): 15, (12,14): 16}
```

```
cbs_female_90 = {(0,1): 772, (1,2): 705, (2,3): 587,
(3,4): 461, (4,5): 397, (5,6): 297, (6,7): 227,
(7,8): 146, (8,9): 111, (9,10): 81, (10,11): 59,
(11,12): 45, (12,13): 22, (13,15): 19, (15, 19): 14}
```

```
cbs_male_90 = { (0,1): 606, (1,2): 508, (2,3): 419,  
               (3,4): 350, (4,5): 249, (5,6): 187, (6,7): 143,  
               (7,8): 97, (8,9): 72, (9,10): 39, (10,11): 25,  
               (11,12): 15, (12,14): 13, (14,18): 13}
```

```
dct = germany_male_dct
```

```
optimized_chi_exponential, params =  
optimizer.minimize_chisquare  
( 'exponential', dct, [[0, 1]])  
optimized_chi_weibull, params =  
optimizer.minimize_chisquare  
( 'weibull', dct, [[0, 1], [0, 2]])  
optimized_chi_gompertz, params =  
optimizer.minimize_chisquare  
( 'gompertz', dct, [[0, 1], [0, 1]])
```

Додаток В

d.f. / p-value	0.1	0.05	0.01	0.001
1	2.71	3.84	6.63	10.83
2	4.61	5.99	9.21	13.82
3	6.25	7.81	11.34	16.27
4	7.78	9.49	13.28	18.47
5	9.24	11.07	15.09	20.52
6	10.64	12.59	16.81	22.46
7	12.02	14.07	18.48	24.32
8	13.36	15.51	20.09	26.12
9	14.68	16.92	21.67	27.88
10	15.99	18.31	23.21	29.59
11	17.28	19.68	24.72	31.26
12	18.55	21.03	26.22	32.91
13	19.81	22.36	27.69	34.53
14	21.06	23.68	29.14	36.12
15	22.31	25.00	30.58	37.70
16	23.54	26.30	32.00	39.25
17	24.77	27.59	33.41	40.79
18	25.99	28.87	34.81	42.31
19	27.20	30.14	36.19	43.82
20	28.41	31.41	37.57	45.31
21	29.62	32.67	38.93	46.80
22	30.81	33.92	40.29	48.27