

# Анализа на демографски животни табели

*Статистичко моделирање*

Симона Ристовска

Факултет за информатички науки и компјутерско инженерство, УКИМ

Ментор: проф. д-р Билјана Тојтовска Рибарски

23 јуни 2025 год.

# Содржина

1	Вовед	2
2	Опис на податоците	2
3	Шведска $e_0$ од 1751 до 2023	4
4	Смртност по возраст $m_x$ од 1751 до 2023	5
5	Преостанати години од животот $e_x$ за 10, 20, 40 и 65 години (1751–2023)	7
6	Криви на смртност $m_x$ за Шведска	8
7	Распределби на $m_x$ за 2020 година на возраст 40–90	9
8	Мерки за оценка на квалитетот на фитувањето	11
9	Криви на преживување $l_x$ за Шведска (1800–2020)	12
10	Тренд на очекуваната животна должина при раѓање $e_0$ (мажи vs. жени)	13
11	Разлика меѓу половите во животната должина при раѓање $e_0$ (жени – мажи)	14
12	Троделна хеатмапа на $\log(m_x)$ за мажи и жени и нивната разлика	15
13	Хазардна стапка кај мажи и жени: $\frac{m_x^{(M)}}{m_x^{(F)}}$ преку време	16
14	Идеи за понатамошна доработка	18
15	Заклучок	19

# 1 Вовед

Целта на овој проект е преку детална статистичка анализа и визуелизација да се прикаже развојот на смртноста и животната должина во една држава во одреден временски период. Преку користење на животни табели, ќе се анализираат различни демографски индикатори како очекуван животен век, стапка на смртност, веројатност за умирање и криви на преживување, при што ќе се добие подобра претстава за тоа како човечкиот животен век се менувал низ историјата, особено во контекст на епидемии, војни, и социјални и медицински реформи.

Како држава за анализа ја избрав Шведска. Главната причина за изборот е достапноста и конзистентноста на податоците: Human Mortality Database (HMD) обезбедува висококвалитетни и детални животни табели за Шведска уште од 1751 година па се до денес. Тоа ни овозможува да направиме долгорочна анализа на животниот век и да го споредиме неговиот развој во различни периоди, како и да ги анализираме разликите по возраст и пол.

Проектот е структуриран во неколку сегменти. Прво, детално се објаснуваат основните демографски поими и атрибути што се користат во животните табели, за да се осигура правилно толкување на резултатите. Потоа следуваат визуелизации кои прикажуваат:

- Трендови во очекуваната животна должина при раѓање ( $e_0$ ) низ годините.
- Промените во стапките на смртност по возраст ( $m_x$ ) со текот на времето.
- Разлики во преживувањето помеѓу мажи и жени...

Покрај дескриптивната анализа, проектот вклучува и применето моделирање преку фитување на функции за смртност (Гомпертз и Вејбул модели) на реални податоци за современа Шведска, со цел подобро да се разберат биолошките и статистичките механизми што стојат зад процесот на стареење.

Анализата има и едукативна и истражувачка вредност: преку визуелно и квантитативно прикажување на еволуцијата на демографските индикатори, се добиваат увид и свесност за тоа како здравствените и социоекономските услови влијаат врз животниот век. Ова не е само историска демографска анализа, туку и прилог кон разбирањето на човековиот опстанок во различни општества и времиња.

## 2 Опис на податоците

Животните табели претставуваат статистички инструменти кои ни овозможуваат да ја прикажеме смртноста и преживувањето во една популација преку сите возрасти. Тие служат како основа за формирање на различни демографски индикатори од кој во овој проект најкористени ќе бидат:

- **Преживувачка функција**  $\ell_x$  (број на поединки што преживеале до возраст  $x$ ),
- **Веројатност за умирање**  $q_x$  (веројатност дека индивидуата на возраст  $x$  ќе умре до возраст  $x + 1$ ),
- **Интензитет на смртност**  $m_x$  (централна стапка на смртност при возраст  $x$ ),

- **Очекувана должина на живот на една индивидуа**  $e_x$  (очекуваниот број години кои ќе ги живее индивидуата од возраст  $x$  натаму).

Колона	Објаснување
Year	Година за која е пресметана животната табела
Age	Возраст $x$ (во години)
mx	Централна стапка на смртност при возраст $x$
qx	Веројатност за умирање на возраст $x$
ax	Средна старост при умирање на возраст $x$
lx	Број (или пропорција) што преживеале до возраст $x$
dx	Број на поединци што умираат на возраст $x$ ( $d_x = l_x - l_{x+1}$ )
Lx	Вкупен број години што ги поминале сите поединци во возраста $x$
Tx	Вкупен број години што ќе ги живеат оние кои се на возраст $x$
ex	Очекувана должина на живот на возраст $x$ , односно $e_x = T_x/l_x$

Табела 1: Колони и објаснување на животните табели

За подобро да ги објасниме атрибутите ќе земеме една популација од 100 луѓе кои сите умираат во рок од 6 години (од 0-та година до крајот на 5-та година)

- Променливата на возраста  $x$  го претставува интервалот на возраста од  $x$  до  $x + 1$
- $l_x$  го претставува бројот на луѓе кои преживеале до  $x$ -годишна возраст
- $d_x$  бројот на луѓе кои починале помеѓу  $x$  и  $x + 1$  година,
- А  $q_x$  е стапката на смртност, т.е. процентот на луѓе кои починале помеѓу  $x$  и  $x + 1$  година,  $q_x = d_x/l_x$

	A	B	C	D
1	age	qx	lx	dx
2	0	0.280	100	28
3	1	0.125	72	9
4	2	0.111	63	7
5	3	0.143	56	8
6	4	0.313	48	15
7	5	1.000	33	33

Слика 1: Примерок од животна табела

## Очекуван животен век

**Очекуван животен век (Life expectancy):** Просечниот број на дополнителни години кои се очекува да живее лице на дадена возраст, врз основа на досегашните податоци за смртност.

Како можеме да го пресметаме животниот век од тие бројки? Постојат два еквивалентни начини за толкување на очекуваниот животен век:

- просечниот (просечен) број на години живеени во дадена популација
- просечната (средна) возраст на смртта

$L_x$  е вкупниот број на живеени години над возраста  $x$  или вкупниот број на години живеени во популацијата за луѓе на возраст меѓу  $x$  и  $x + 1$ .

Од  $L_x$  можеме да го изведеме  $T_x$ , вкупниот број на живеени години ( $T_x$ ) од возраста  $x$  до крајот, па едноставно збирот на сите  $L_x$  од возраста  $x$  до последната година во нашата табела.

Последниот чекор е да се земе  $T_x$  и да се подели со бројот на луѓе на почетокот на секој период,  $l_x$ . Ние го правиме тоа на секоја возраст, за да го добиеме **очекуваниот животен век на возраст  $x$**  (означен како  $e_x$ ).

	A	B	C	D	E	F	
1	age	qx	lx	dx	Lx	Tx	ex
2	0	0.280	100	28	86	322	3.2
3	1	0.125	72	9	67.5	236	3.3
4	2	0.111	63	7	59.5	168.5	2.7
5	3	0.143	56	8	52	109	1.9
6	4	0.313	48	15	40.5	57	1.2
7	5	1.000	33	33	16.5	16.5	0.5

Слика 2: Дополнет примерок од животна табела

$m_x$

- **возрасно-специфична стапка на смртност или централна стапка на смртност.**
- **колку често се случуваат смртни случаи за таа популација-година.**

Замислете популација од **100 000** новородени. На секој роденден  $x$ , дел од нив ќе починат во текот на таа година, а останатите се живи. Вредноста  $m_x$  ни кажува колку брзо луѓето умираат на возраст  $x$ .

Во годината од возраст  $x$  до  $x + 1$ , сите луѓе придонесуваат „години“ во одреден збир. Тој збир е  $L_x$ . Бројот на смртни случаи во истата година е  $d_x$ . Тогаш:

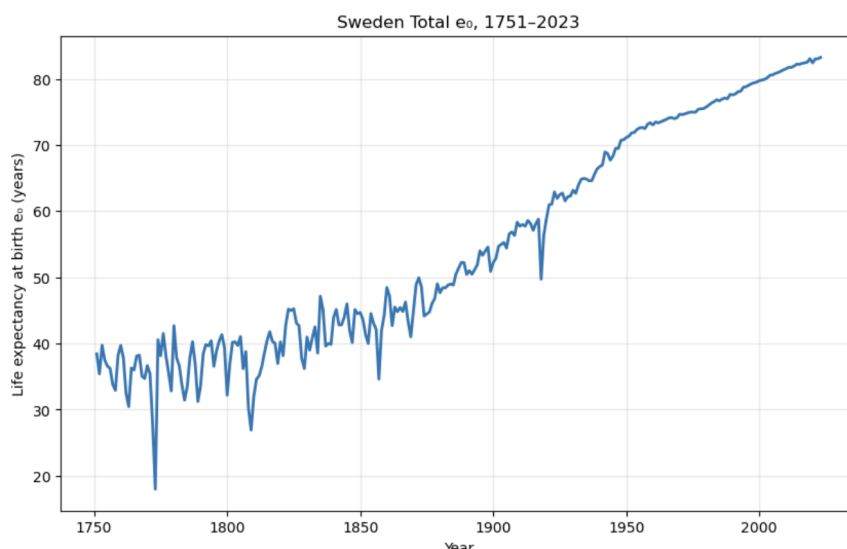
$$m_x = \frac{d_x}{L_x},$$

односно број на смртни случаи поделен со вкупните години живеење во тој возрасен интервал.

Тоа не е „веројатност“ за умирање (тоа би било  $q_x$ ), но многу се приближува кон  $q_x$  кога стапките се ниски. „Ако  $m_{80} = 0.05$ , тоа значи во просек 5 смртни случаи на 100 лица меѓу 80-годишници.“

### 3 Шведска $e_0$ од 1751 до 2023

**Прва и наједноставна визуелизација.** Во овој график ја прикажуваме вредноста на животната продолженост на новороденче ( $e_0$ ) за секоја година од 1751 до 2023, земајќи ги податоците од животни табели каде Age е 0. На хоризонталната



Слика 3: Очекувана должина на животот на едно новороденче ( $e_0$ ) во Шведска, 1751–2023.

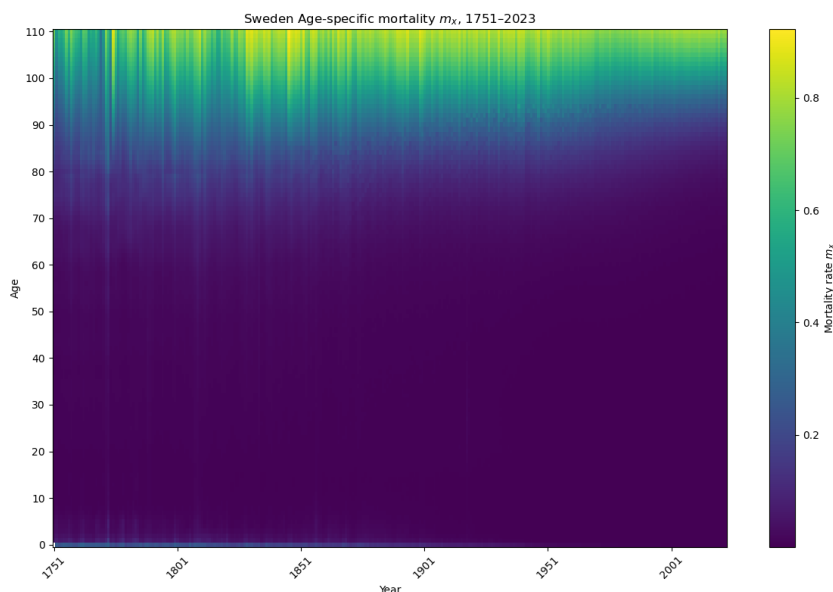
оска е годината ( $\text{Year}$ ), а по вертикалната е  $e_x$  (годишна продолженост на животот на новороденчето).

**Што се гледа на визуелизацијата.** Графикот е едноставна линија на животната продолженост во текот на 272 години. Хоризонталната оска (x-оска) се простира од 1751 до 2023, а вертикалната (y-оска) го прикажува очекуваниот број на години живот во моментот на раѓање. Во подалечните години околу 1750, вредноста на  $e_0$  се движи околу 35–40 години. Нацртана е една единствена сина линија која покажува дека вредностите благо варираат од година во година: имаме пониски точки (околу 20 години) за време на некои историски кризи (на пример, крајот на 18 век и Пандемијата на чуми), а потоа постепено се стабилизира околу 40–45 години во доцните 1800-ти. Падот во близина на 1918 година (околу 50 години на графикот) го означува влијанието на Шпанскиот грип. Потоа, по Втората светска војна, линијата рапидно се искачува кон вредности над 60 години. Од 1950-тите, кривата покажува континуиран пораст и доаѓа до околу 80–83 години до 2023.

**Интерпретација.** Првичните години сведочат за високата ранлива фаза во која секое епидемиско или воено нарушување веднаш се рефлектира во поголеми падови, додека современите периоди покажуваат стабилно и мирно зголемување благодарение на напредокот во санитарната инфраструктура, родителската здравствена нега и општата социјална заштита.

## 4 Смртност по возраст $m_x$ од 1751 до 2023

**Што е прикажано?** Во оваа визуелизација (т.н. heatmap) се прикажани вредности на централната стапка на смртност  $m_x$  за секоја година од 1751 до 2023 (хоризонтална оска) и за секоја возраст од 0 до 110 (вертикална оска). Податоците потекнуваат од животни табели, каде прво се филтрирани сите редови за кои  $\text{Age}$  е од 0 до 110. Секоја селектирана вредност  $m_x$  е прикажана со бои: потемни (виолетови) нијанси означуваат ниски стапки на смртност, додека потоплите (жолто-зелени)



Слика 4: Смртност по возраст  $m_x$  во Шведска, 1751–2023.

нијанси покажуваат повисоки стапки.

**Што се гледа во визуелизацијата.** Графикот ни прикажува како стапките на смртност варираше во текот на повеќе од 270 години за различни возрасти:

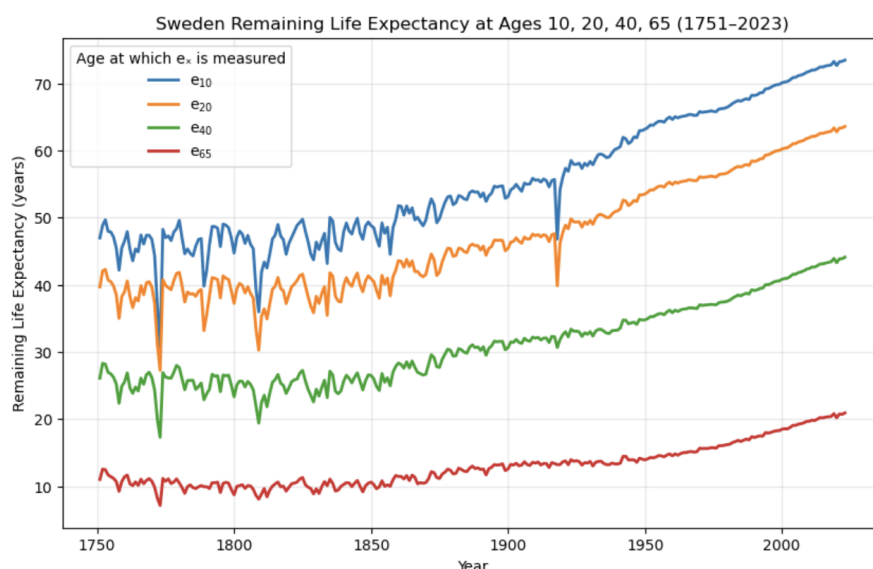
- **Смртност на новороденчиња (Age = 0):** Најдолниот ред (возраст = 0) е најинтензивен (потопли нијанси) во 18-тиот и 19-тиот век, што зборува за висока смртност на новороденчиња (види слика 5). Постепено, кон почетокот на 20-тиот век, преоѓа кон побледи бои (помали вредности), како резултат на подобрувања во јавното здравје и вакцинација.
- **Детска смртност (Age 1–10):** Подолу на вертикалната оска (возрасти 1–10) се гледаат слични, но не толку интензивни ленти – доцниот 19-ти век и почетокот на 20-тиот: постепен континуиран пад, но сепак извесни врвови (потопли ленти) за време на епидемии.
- **Младинска и возрасна смртност (Age 10–50):** Во текот на целиот 19-ти век вредностите се релативно ниски (движење во темносино-виолетова зона), но има периодични кратки зголемувања за време на војни (на пример, во 1808 и 1814 – Наполеоновите војни) и епидемии (ова се кратко видливи вертикални светли ленти).
- **Средновечна до постара смртност (Age 50–80):** Најнесомнено, со текот на годините, овој сегмент постепено се поместува кон побледи нијанси. Во средината на 19-тиот век сè уште има релативно темни тонови, но кон крајот на 20-тиот век (после Втората светска војна) почнува да биде побело-зелен (пониски стапки на смртност). Ова укажува на подолг животен век и подобрена грижа за хронични заболувања.
- **Постари возрасти (Age 80–110):** Горниот дел од heatmap-ата (Age 80) прикажува многу високи стапки на смртност (жолто-зелени тонови во раните години), а со напредокот во медицина и подобри социјални услуги, и тука постепено

се намалуваат вредностите. На крајот на 20-тиот и почетокот на 21-ви век, линијата кај  $\text{Age} > 90$  е побледа („помалку жолта“), што значи подобрување на условите за постарите лица.



Слика 5: Зголемен преглед на светлите нијанси во смртноста на новороденчињата.

## 5 Преостанати години од животот $e_x$ за 10, 20, 40 и 65 години (1751–2023)



Слика 6: Преостаната животна продолженост  $e_x$  за возрастите  $x = 10, 20, 40, 65$  во Шведска (1751–2023).

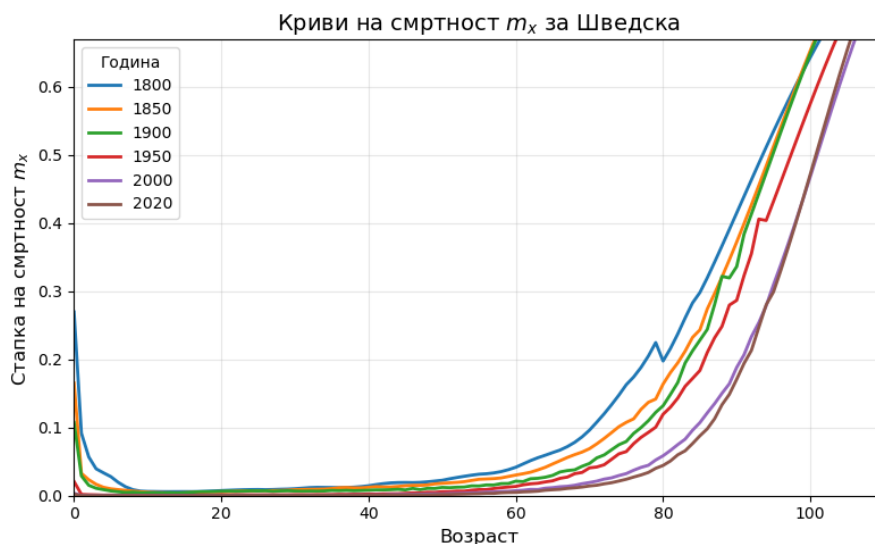
**Што прикажува графикот?** Четири линии покажуваат преостаната животна продолженост  $e_x$  за оние што веќе ја достигнале возраста  $x$  (10, 20, 40 и 65 години), пресметана за секоја година од 1751 до 2023.

- **1751–1850:** Силни годишни колебања поради епидемии и војни. Поголеми падови се јасни кај помладите ( $e_{10}$ ,  $e_{20}$ ), но влијае и кај  $e_{40}$  и  $e_{65}$ .
- **1850–1918:** Постепен пораст на сите  $e_x$  со пониски флукутации. Неколку кратки падови (најизразен околу 1918, „Spanish flu“).
- **1950–1980:** По Втората светска војна следува стабилен раст.  $e_{10}$  се искачува од 55 на 65 години,  $e_{20}$  од 47 на 55,  $e_{40}$  од 30 на 32–33 и  $e_{65}$  од 13 на 14.
- **1980–2023:** Континуиран раст со намалени годишни осцилации. До 2023,  $e_{10} \approx 75$ ,  $e_{20} \approx 64$ ,  $e_{40} \approx 44$ ,  $e_{65} \approx 21$ .



**Клучна порака.** Долготрајниот тренд покажува трансформација од големи варијации во 18–19 век (епидемии, војни) кон постојан и непречен раст по 1950-тите, што рефлектира напредок во јавното здравство, медицината и социјалните услови.

## 6 Криви на смртност $m_x$ за Шведска



Слика 7: Годишни криви на смртност  $m_x$  според возраст за Шведска во 1800, 1850, 1900, 1950, 2000 и 2020 година.

**Што прикажува графикот?** На овој линиски дијаграм се прикажани возраст-специфични стапки на смртност  $m_x$  (од 0 до 110 години) во шест фундаментални години (1800, 1850, 1900, 1950, 2000, 2020).

**Клучни забелешки:**

- **1800 и 1850:**

- Многу висока стапка кај доенчињата (околу 0–1 година) и стрмно опаѓање до скоро нула во возраста 5–10 години.
- Од 30–40 години па натаму,  $m_x$  се искачува експоненцијално (ефект на ослабени услови и епидемии).

- **1900 и 1950:**

- Драматичен пад на детската смртност во споредба со 1800/1850.
- Во зрелите години (30–60)  $m_x$  е повисока во 1900 отколку во 1850, но постепено опаѓа до 1950, особено кај возрастите над 60 (напредок во јавното здравје и медицина).

- **2000 и 2020:**

- Претежно „плосната“ крива до возраст 50–60 години (екстремно ниска детска и средновековна смртност).

- Најстрмен раст на  $m_x$  се мери после 70 години, но и оваа старосна стапка значително се поместува кон повисоки возрасти: во 1800 експоненцијалниот пораст започнува околу 50–60 години, а во 2020 дури над 70–75 години.
- Стапките за старите возрасти (80–100+) се значително пониски во 2020 отколку во сите претходни години, што ја илустрира „компресијата“ на стареењето и поголемиот опстанок до подоцнежна возраст.

**Клучна порака.** Во текот на 220 години, Шведска поминува од многу високи стапки на смртност за новороденчиња и млади лица кон масивно намалување на смртноста на сите возрасти до средна возраст. Од средината на XX век и особено до 2020, едвај има смртност под 50 години, додека релативната тежина на старите возрасти се преместува кон 80–100 и понатаму.

## 7 Распределби на $m_x$ за 2020 година на возраст 40–90



Слика 8: Фит на централната стапка на смртност  $m_x$  во 2020 година (возраст 40–90) со Гомпертз и Вејбул модели.

### Што прикажува графикот?

- Црните точки се оригиналните вредности на  $m_x$  за возрасниот интервал 40–90 во 2020.
- Сината линија е Гомпертз-фит ( $m_x = \alpha e^{\beta x}$ ) со параметри  $\alpha = 7.05 \times 10^{-6}$  и  $\beta = 0.110$ .
- Портокаловата испрекината линија е Вејбул-фит ( $m_x = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1}$ ) со параметри  $k = 11.77$  и  $\lambda = 88.19$ .

### Клучни забелешки:

- Во долниот дел (возраста 40–60), двата модели малку го следат  $m_x$ , бидејќи реалната смртност е многу ниска и константна.

- Во средните возрасти (60–75), Гомпертз ги следи  $m_x$  податоците релативно блиску, а Вејбул го прави истото само што се наоѓа малку подолу за разликата од Гомпертз кривата.
- Над 75 години, Гомпертз (сината крива) ја засилува експоненцијалата и ја следи стрмнината на  $m_x$ , но Вејбул (портокаловата крива) покажува уште поголем пораст кон крајот (85–90), при што двете распределби имаат слично однесување, но Вејбул малку подобро ги следи податоците во овој случај.
- Параметарот  $\beta = 0.110$  (Гомпертз) ја означува годишната стапка на „старење“ во смртност, додека  $k > 10$  (Вејбул) ја покажува релативно стрмната крива во високите возрасти.

**Како е извршено фитувањето?** Фитувањето е извршено преку нумеричка оптимизација на параметрите на двата модели врз основа на реалните вредности на  $m_x$ .

За Гомпертз моделот, се применува линеарна регресија врз логаритмираните вредности:

$$\ln(m_x) = \ln(\alpha) + \beta x$$

од каде што преку метода на најмали квадрати (least squares) се добиваат параметрите  $\alpha$  и  $\beta$ .

За Вејбул моделот, бидејќи изразот:

$$m_x = \frac{k}{\lambda} \left( \frac{x}{\lambda} \right)^{k-1}$$

не е линеарен во параметрите, фитувањето се врши преку нумеричка минимизација на вкупната квадратна грешка меѓу оригиналните и предвидените вредности на  $m_x$ . Се користи методата `scipy.optimize.minimize`, со почетни приближувања и ограничувања за да се обезбеди стабилна и реалистична пресметка на параметрите  $k$  и  $\lambda$ .

**Демографско значење на моделите:** Гомпертз моделот се користи широко во биостатистика и актуарство, бидејќи ја рефлектира идејата дека смртноста расте експоненцијално со стареењето. Параметарот  $\beta$  директно ја мери „брзината на стареење“. Колку е поголем  $\beta$ , толку побрзо расте ризикот од смрт со возраста.

Од друга страна, Вејбул моделот овозможува поголема флексибилност: за  $k > 1$ , ризикот расте со возраста (старење), за  $k = 1$  ризикот е константен (експоненцијална дистрибуција), а за  $k < 1$  ризикот паѓа со возраста (ретко применливо во човекова смртност). Во овој контекст, вредноста  $k \approx 11.8$  сугерира нагласен раст на ризикот кај постарите возрасти.

**Целта на ваквото фитирање:** Целта не е само да се добие добра апроксимација на  $m_x$ , туку и да се разбере структурата на смртноста во повозрасните групи и дали таа се однесува според очекуваните биолошки и статистички закони. Двата модели се согласуваат со реалните податоци, но Вејбул овозможува малку подобро доловување на стрмниот пораст во екстремните години (85–90), што е важен индикатор за компресија на смртноста – феномен во кој повеќето смртни случаи се концентрираат во покасни години од животот.

**Клучна порака.** Гомпертз-моделот добро го опишува експоненцијалниот пораст на  $m_x$  кај возрасните над 60, додека Вејбул покажува малку поизразена кривина кај најстарите групи. Во двата случаи, фит-параметрите укажуваат на многу низок

раст на  $m_x$  пред 60 години и стрмен пораст по 75–80 години, што ја демонстрира компресијата на смртноста кон подоцнежна старост.

## 8 Мерки за оценка на квалитетот на фитовањето

За да се процени колку добро моделите на Гомпертз и Вејбул ги следат реалните податоци за централната стапка на смртност  $m_x$  кај населението во Шведска во 2020 година (возраст 40–90), беа пресметани неколку стандардни метрики за квалитет на фитот: RMSE,  $R^2$  и AIC.

Модел	Број на параметри	RMSE	$R^2$	AIC
Gompertz	2	0.00709735	0.969988	-500.699
Weibull	2	0.00233717	0.996746	-613.999

Табела 2: Мерки за евалуација на Гомпертз и Вејбул моделите (возраст 40–90, Шведска, 2020).

### Објаснување на метриките

- **RMSE (Root Mean Squared Error):** ја мери просечната разлика помеѓу предвидените вредности од моделот и реалните податоци. Помала вредност значи дека моделот има подобро прилагодување на податоците.
- **$R^2$  (Коефициент на детерминација):** покажува колкав дел од варијацијата на податоците е објаснета преку моделот. Вредност блиска до 1 сугерира дека моделот добро го следи трендот на смртност.
- **AIC (Akaike Information Criterion):** оценува моделот преку балансирање на точноста и комплексноста. AIC казнува модели со повеќе параметри, па затоа помала вредност значи подобар компромис помеѓу прецизност и едноставност.

### Интерпретација на резултатите

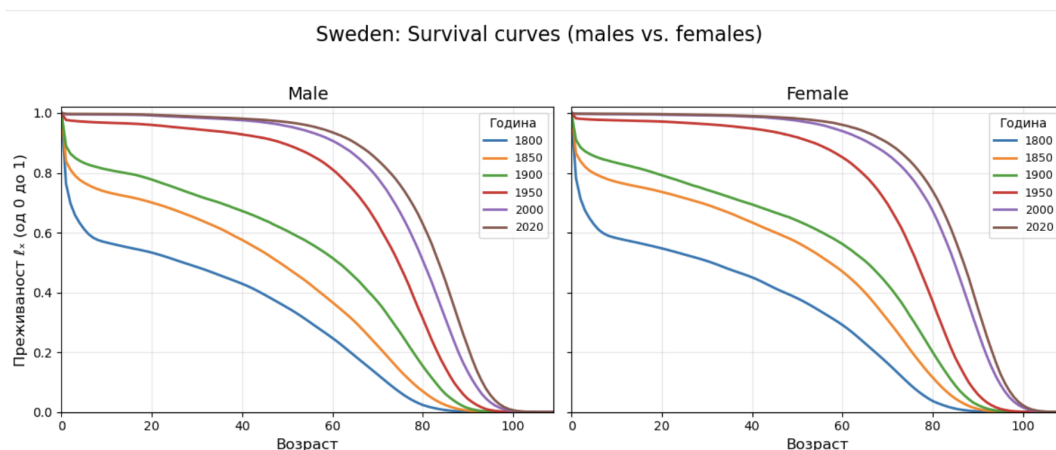
Според вредностите во Табела 2, може да се заклучи следново:

- **Вејбул моделот има значително помал RMSE (0.0023)** во споредба со Гомпертз моделот (0.0071). Тоа значи дека предвидувањата на Вејбул се поблиску до реалните вредности на  $m_x$  во речиси целиот интервал од 40 до 90 години.
- **Коефициентот на детерминација  $R^2$  кај Вејбул е 0.9967**, што е речиси совршено, наспроти 0.9699 за Гомпертз. Иако и двата модели добро го следат трендот, Вејбул ја објаснува речиси целата варијација на податоците.
- **Најсилен доказ доаѓа од AIC вредноста:** Вејбул има AIC од -613.999, наспроти -500.699 за Гомпертз. Со оглед на тоа што двата модели користат ист број параметри (2), разликата од над 100 поени во AIC укажува дека Вејбул моделот е значително подобар во балансот меѓу прецизност и едноставност.

**Заклучок:** Според сите три метрики, Вејбул моделот го дава најдобриот фит на реалните податоци за смртност во Шведска за 2020 година. Тој не само што прецизно го следи трендот и минимално отстапува од податоците, туку и според AIC е супериорен во смисла на оптимален компромис помеѓу точност и сложеност.

Овие резултати потврдуваат дека при анализа на повозрасна смртност (возраст 40–90), Вејбул моделот овозможува пофлексибилно и реалистично описување на кривата на ризикот.

## 9 Криви на преживување $l_x$ за Шведска (1800–2020)



Слика 9: Криви на преживување  $l_x$  (од 0 до 1) во распон на возрасти 0–105 за мажи (лево) и жени (десно) во Шведска за 1800, 1850, 1900, 1950, 2000 и 2020 година.

**Краток опис:** Прикажани се криви на преживување  $l_x$  за шест различни години и споредба помеѓу половите. Вредноста  $l_x$  ја претставува фракцијата од популацијата која преживува до дадена возраст  $x$ , започнувајќи од  $l_0 = 1.0$ .

**Што се гледа:**

- Општа траекторија:

- Во сите години  $l_x$  почнува од 1.0 и постепено опаѓа кон 0 на возраста 100–105 години.
- Најстрмниот пад се јавува во детството (0–5 години) во постарите години (1800–1900), додека во поновите години (1950–2020) кривите остануваат речиси хоризонтални до 60–70 години, што укажува на значително подобро детско и младинско преживување.

- Промени низ епохи (по години):

- **1800–1900:** Силно изразена смртност во детството и постојан пад на преживувањето со возраста. До 1900,  $l_5$  се подобрува на 0.85–0.90, во споредба со 0.7 во 1800.
- **1950:** Преоден период со значителен напредок — преживувањето во детството се стабилизира, а опаѓањето на  $l_x$  започнува во повозрасни години (60).

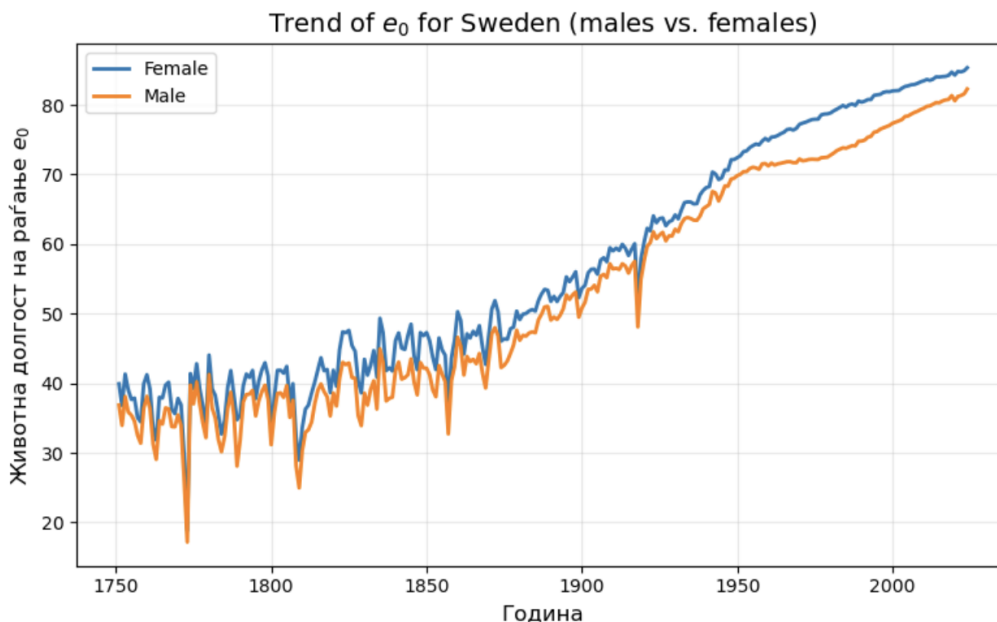
- **2000–2020:** Преживувањето е стабилно и високо до околу 70 години. Повеќето криви опаѓаат брзо дури по 80-та година, што покажува "компресија" на смртноста во повозрасни години и зголемена животна должост.

- **Разлики меѓу мажи и жени:**

- Жените имаат повисоки  $l_x$  криви во секоја година, особено изразено по 20-тата година.
- Во 2020, повеќе од 90% од жените преживуваат до 75 години, а кај мажите 85%.
- Разликата е најизразена по 80-та година — жените покажуваат побавно опаѓање и поголема веројатност за преживување во старост.

**Клучна порака:** Кривите ја документираат демографската транзиција на Шведска: од општество со висока детска смртност и краток животен век кон модерна популација со речиси универзално детско преживување и продолжен животен век. Жените секогаш покажуваат повисоко ниво на преживување, особено во доцните години, што укажува на биолошки и социјални фактори кои влијаат врз долгорочниот ризик од смртност.

## 10 Тренд на очекуваната животна должина при раѓање $e_0$ (мажи vs. жени)



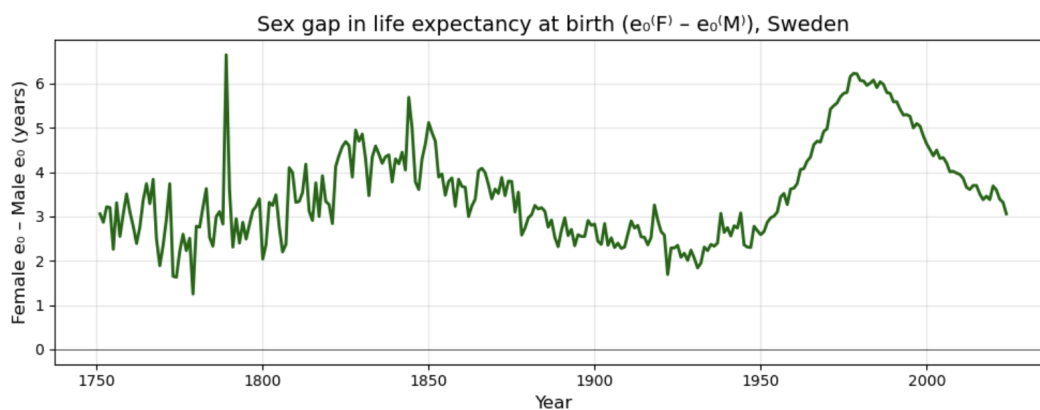
Слика 10: Тренд на  $e_0$  (години) во Шведска за мажи (портокалова) и жени (сина), 1751–2020.

**Краток опис:** Годишни вредности на животната должина при раѓање ( $e_0$ ) за мажи и жени во Шведска од 1751 до 2020 година.

**Клучни точки:**

- **Очигледна разлика меѓу половите:** Жените имаат повисоко  $e_0$  од мажите во целото време, од 1–2 години разлика во XVIII век до 4–5 години во средината на XX век, а потоа благо се стеснува до 3–4 години кон 2020.
- **Позитивен пораст преку два века:**
  - XVIII–XIX век: Постепен раст од: 36→48 (жени) и 34→45 (мажи) години, флуктуации при епидемии.
  - 1870–1930: Брзо зголемување од: 50→75 (жени) и 48→70 (мажи) години со модерни медицински реформи.
  - 1930–2020: Понатамошен пораст до приближно 85 години кај жените и 82 години кај мажите, благодарение на антибиотиците, вакцинациите и здравствената технологија.
- **Краткорочни падови во кризни години:** Ретки, но привремени опаѓања за 1–3 години околу Наполеонските војни (1808–14), шпанскиот грип (1918–19) и делумно во Втората светска војна, но по 1950 тие се речиси невидливи.
- **Сè поизразена стабилност по 1950:** Варијациите од година во година се минимизираат, што укажува на континуирана еволуција на јавното здравје и социјалните реформи.
- **Контракција на полова разлика по 1980:** Од средината на 1980-тите разликата меѓу половите благо се намалува, делумно поради подобрување на здравствените практики кај мажите (намалување на пушењето, нивна поголема грижа за кардиоваскуларни ризици).

## 11 Разлика меѓу половите во животната должина при раѓање $e_0$ (жени – мажи)



Слика 11: Разлика во  $e_0$  ( $e_0^F - e_0^M$ ) во Шведска, 1751–2020.

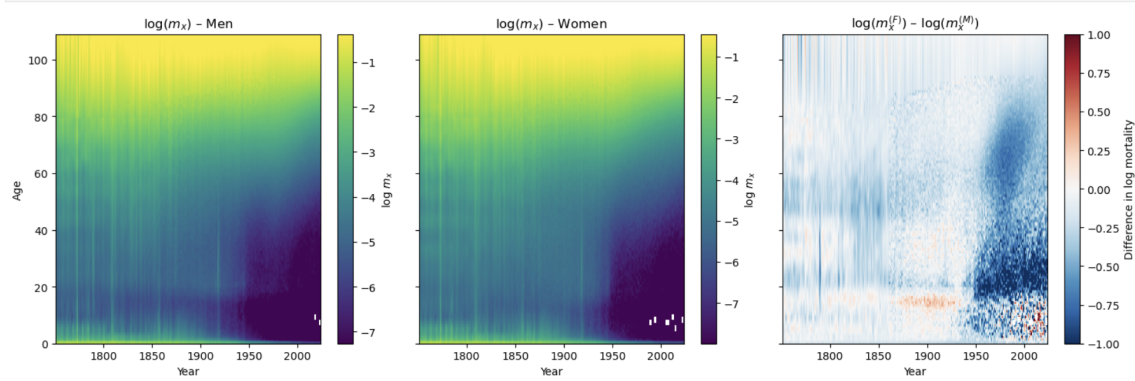
**Што е прикажано?** Годишната разлика во очекуваната животна должина при раѓање помеѓу жените ( $e_0^F$ ) и мажите ( $e_0^M$ ) во Шведска од 1751 до 2020 година. На вертикалната оска е  $\Delta e_0 = e_0^F - e_0^M$  во години.

**Клучни фази и интерпретации:**

- **XVIII–почеток на XIX век (1751–1820):**  $\Delta e_0$  варира околу 2–4 години, со благи флукутации поради епидемии и војни.
- **Средина на XIX век (1820–1870):** Јазот се намалува на околу 2–3 години поради силни демографски шокови (на пр. шпанскиот грип, воени конфликти).
- **Подем кон XX век (1870–1950):** Постепено зголемување од 3 до 5–6 години поради намалена детска смртност и воведување вакцинации.
- **Пик во средината на XX век (1950–1970):** Јазот достигнува максимум од околу 7 години, кога здравствените интервенции (антибиотиците, масовните имунизации) ги фаворизираат жените.
- **Контракција во доцниот XX и почетокот на XXI век (1970–2020):** Разликата постепено се стеснува од 7 на околу 3–4 години, поради подобрувања во машкото здравје и еднаков пристап до здравствени услуги.

**Значење:** Овој тренд ја илустрира постојаната предност на жените во очекуваната животна должина, но и динамиката на половиот јаз: од осцилации во XVIII–XIX век, преку драматичен раст и пик во средината на XX век, до значително стеснување кон модерното време.

## 12 Троделна хеатмапа на $\log(m_x)$ за мажи и жени и нивната разлика



Слика 12: Хеатмапи на  $\log(m_x)$  за мажи (лево), за жени (средно) и разлика  $\log(m_x^{(F)}) - \log(m_x^{(M)})$  (десно) во Шведска, 1751–2023, по возраст.

**Што е прикажано?** На левиот и средниот панел се претставени временско-возрасните распределби на логаритамот на стапката на смртност  $m_x$  за мажи, односно за жени. Во вертикална насока е возрастаот (од 0 до 110 години), а во хоризонтална насока е годината (од 1751 до 2023). Бојата ја покажува вредноста на  $\log(m_x)$  – од подлабоко виолетова (многу ниска смртност, на пример кај млади возрасни во XXI век) до светло жолта (висока смртност, типично во детството за XVIII век или кај старите возрасни).

На десниот панел е прикажана бојата која одговара на разликата

$$\Delta_{\log m}(x, t) = \log(m_x^{(F)}(t)) - \log(m_x^{(M)}(t)),$$



т.е. колку „поголем“ или „помал“ (негативен) е  $\log(m_x)$  за жените во однос на мажите, за секоја година  $t$  и возраст  $x$ .

**Како се читаат боите?**

• **Лев панел (мажи) и среден панел (жени):**

- Виолетово-сива боја околу  $-5$  на скалата (длабока темна нијанса) значи многу ниска смртност ( $m_x \approx e^{-5}$  или помалку).
- Жолто-сива боја околу  $-1$  значи релативно висока смртност ( $m_x \approx e^{-1} \approx 0.37$ ).
- На младите возрасти (0–5 години) во XVIII век гледаме жешки (жолти) „врнежи“, отсликувајќи високата детска смртност. Со текот на времето оваа област преминува во темни нијанси (пониска инфантилна смртност).
- На старите возрасти ( $>70$  години) откако ќе дојде XX и XXI век, бојата се движи од зелени нијанси ( $-2$ ) во виолетово ( $-4$ ) бидејќи смртноста на т.н. „старечки“ возрасни драматично се намалува.

• **Десен панел (разлика во лог смртност):**

- Топла (црвенкасто/портокалова) боја означува  $\Delta_{\log m}(x, t) > 0$ : т.е. дека  $\log(m_x^{(F)}) > \log(m_x^{(M)})$ , значи жените имаат повисока специфична смртност во таа комбинација возраст–година.
- Ладна (сина) боја означува  $\Delta_{\log m}(x, t) < 0$ : мажите имаат повисока специфична смртност.
- Бела нијанса или боја приближна до нула значи приближно иста лог-фаталност за двата пола.

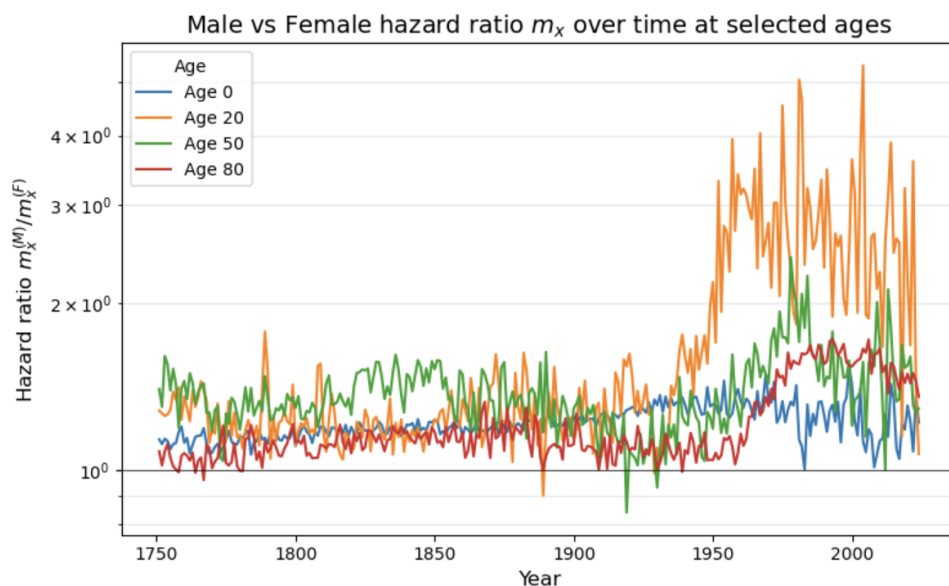
**Главна порака:** Хеатмапите покажуваат дека во периодот 1751–2023 драматично се намалува смртноста на сите возрасти за двата пола, но не сите возрасни групи и не за двата пола во исто темпо. Пред XIX век, детската смртност и мајчинската фаталност доминираат, а по 1900 машката смртност кај старешките возрасти останува повисока. По 1950 се појавуваат сини вертикални зони во десниот панел кај возрасти 40+, што укажува на тоа дека мажите имаат сè поголема специфична смртност во однос на жените. Напредокот во здравството, вакцинацијата, и контролата на хронични болести носат различни придобивки низ возрасните и половите групи, при што жените сè повеќе имаат предност во преживување во XX и XXI век.

## 13 Хазардна стапка кај мажи и жени: $\frac{m_x^{(M)}}{m_x^{(F)}}$ преку време

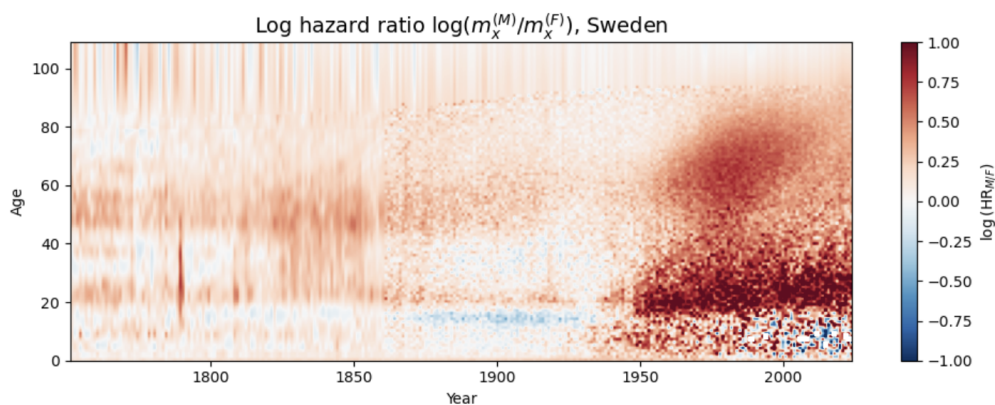
**Цел на анализата:** Овој дел има за цел да ги спореди стапките на смртност кај мажите и жените преку дефинирање на хазардниот однос

$$\text{HR}(x, t) = \frac{m_x^{(M)}(t)}{m_x^{(F)}(t)},$$

каде што  $m_x^{(M)}$  и  $m_x^{(F)}$  се годишните стапки на смртност за мажи и жени на возраст  $x$  и во година  $t$ .



(а) Хазардна стапка  $m_x^{(M)}/m_x^{(F)}$  за избрани возрасти (0, 20, 50, 80).



(б) Хеатмапа од логаритамот  $\log(m_x^{(M)}/m_x^{(F)})$  по возраста и годината во Шведска (1751–2023).

Слика 13: Машко-женски разлики во специфичната смртност во Шведска.

**Горен график (Figure 13a):** Го прикажува односот HR преку време за четири избрани возрасти: 0, 20, 50 и 80 години. Овој график овозможува да се забележат специфични трендови:

- **На возраст 0 (новороденчиња):** Рано во анализираниот период HR е блиску до 1, со умерени варијации, но не постои силна полова разлика.
- **На возраст 20:** По 1950 се јавува остар раст на односот – кај младите возрасни мажите стануваат значително поризични. Ова е поврзано со појава на социјални фактори како што се пушење, алкохол, сообраќајни несреќи и самоубиства.
- **На возраст 50:** Сличен раст во односот по 1950. Ова укажува на хронични болести како кардиоваскуларни проблеми кои повеќе ги погодуваат мажите.
- **На возраст 80:** Иако HR се зголемува, разликата е послаба – половиот јаз се намалува во подоцнежните години.

На графикот, вредност 1 на у-оската претставува ситуација кога мажите и жените имаат иста смртност. Сè над 1 означува поголема смртност кај мажите.

**Долен график (Figure 13b):** Оваа хеатмапа прикажува просторна (Age × Year) визуелизација на логаритмот на хазардниот однос:

$$\log \left( \frac{m_x^{(M)}}{m_x^{(F)}} \right),$$

што овозможува полесна интерпретација и центрирање околу нула (бела боја).

- **Црвенкасти/топли нијанси:** Мажите имаат значително поголема смртност ( $\log(\text{HR}) > 0$ ).
- **Сини/ладни нијанси:** Жените имаат поголема смртност.
- **Бели нијанси:** Стапките се слични.

Поголемиот дел од хеатмапата после 1950 е интензивно црвен за возрасти 20–70, што покажува трајна и зголемена смртност кај мажите. Особено остро е ова по 1970 година.

**Главна порака:** Иако во 18-тиот и 19-тиот век половата разлика во стапките на смртност е релативно стабилна и мала, со почеток на 20-тиот век – и особено по Втората светска војна – мажите покажуваат забележливо повисока смртност во многу возрасти. Ова е индикативно за влијание на модерниот животен стил и социјалните фактори, а не само биолошките разлики. Хазардниот однос е еден од најважните индикатори за анализа на половиот јаз во преживувањето.

## 14 Идеи за понатамошна доработка

Демографските табели овозможуваат анализа од различни аспекти и на различни нивоа на длабочина. Една можна насока е примената на кластерирање, каде што би се групирале животни траектории или индивиди според карактеристики како што се пол, долговечност или типична патека на промена на очекуваниот животен век – дали има нагли растови, континуиран стабилен пораст итн.

Дополнително, можат да се применат техники за предвидување на идните вредности на некои демографски индикатори, како што се  $e_0$  или  $q_x$ , со надгледувано учење – регресиони модели, дрва на одлука, стохастични шуми или градиентни бустинг методи. Особено интересна насока би била анализа со временски серии, користејќи модели како ARIMA, Prophet или длабоки рекурентни мрежи како LSTM, при што би се следел трендот на смртност или животен век низ децении.

Од аспект на машинско учење, постои можност да се истражи како autoencoder мрежи можат да се применат кај табелите на смртност, со цел да се добијат компактни латентни претстави на демографски структури. Подлабока анализа може да вклучи и variational autoencoders или баесови модели, со што ќе се добие генеративен пристап во симулација на нови животни табели. Во современ контекст, се поактуелна е темата за објасливост (explainability) кај ML-модели, особено кога се работи со чувствителни демографски податоци.

Овие насоки покажуваат дека демографската анализа не е ограничена само на описна статистика, туку претставува динамична и интердисциплинарна област која овозможува примена на напредни аналитички алатки и моделирање.

## 15 Заклучок

Во текот на изработката на овој проект, се соочив со повеќе аналитички и технички предизвици, кои беа составен дел од процесот на продлабочено разбирање на демографските животни табели. Еден од првите предизвици беше темелното запознавање со структурата на податоците – тие не беа само едноставни бројки, туку носеа во себе историски контекст, епидемии, социоекономски промени и здравствени иновации. Требаше време и обемно проучување за правилно да се интерпретираат концептите како  $m_x$ ,  $e_0$  и  $l_x$ , како и нивните логаритамски трансформации.

Дополнителен предизвик беше изборот на соодветен начин на визуелизација – како да се прикажат податоците на начин што ќе биде и информативен и естетски разбирлив. Одлучив да користам хеатмапи, хазардни односи и криви на преживување, бидејќи тие нудат прегледност преку време и простор (возраст), а истовремено овозможуваат споредба по пол и година.

Резултатите покажаа дека Шведска претрпела драматични демографски трансформации од 18-тиот до 21-виот век. Стапките на смртност значително се намалуваат низ вековите, особено кај новороденчињата и постарите лица, при што очекуваниот животен век значително се продолжува. Воедно, се потврдуваат и родовите разлики во смртноста – мажите речиси конзистентно имаат повисоки стапки на смртност, особено по средна возраст. Разликите стануваат сè поизразени во поновата историја, што упатува на социјални фактори кои влијаат различно на мажите и жените.

Иако проектот се базира на анализа на една земја, тој отвора пошироки прашања за тоа како политиките за здравствена заштита, социјалните услови и биолошките фактори влијаат врз долгоживоста и преживливоста. Сметам дека ваквите демографски пристапи се основа за подобро планирање на јавното здравство и разбирање на динамиката на населението низ времето.

## Литература

- [1] Ристовска, С. (2025). *Анализа на демографски животни табели: статистичко моделирање*. GitHub Notebook. <https://github.com/simonaristovska0/statistical-modeling>
- [2] *Human Mortality Database*, University of California, Berkeley (САД) и Max Planck Institute for Demographic Research (Германија). <https://www.mortality.org/>
- [3] Wilmoth, J. R., Andreev, K., Jdanov, D., Gleijer, D. A., Boe, C., Bubenheim, M., & Zich, L. (2014). *Methods protocol for the Human Mortality Database*, 2014 Edition. <https://www.mortality.org/Public/Docs/MethodProtocol.pdf>
- [4] Yashin, A. I., Begun, A. S., Boiko, P. E., Ukraintseva, S. V., & Oeppen, J. (1994). *The new trends in survival improvement require a revision of traditional gerontological concepts*. Mechanisms of Ageing and Development, 74(1-2), 1–12. <https://www.sciencedirect.com/science/article/abs/pii/0047637494900957>
- [5] Marthoz, F. (2021). *How to read and build a mortality table*. Medium. <https://fmarthoz.medium.com/how-to-read-and-build-a-mortality-table-a4eb5878abba>
- [6] Office for National Statistics (ONS), UK. *Guide to Calculating National Life Tables*. <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/methodologies/guidetocalculatingnationallifetables>