

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Уфимский государственный авиационный технический университет»
Факультет информатики и робототехники
Кафедра вычислительной математики и кибернетики

Математическое и программное обеспечение оценки и анализа показателей смертности населения России

Выполнил
студент гр. МО-206м
Вахитов А.Р.

Руководитель
к.ф.-м.н., доцент кафедры ВМиК
Прокудина Е.И.

Уфа 2021

Введение в предметную область

Модель смертности является одним из главных орудий в арсенале демографии, причём используется для анализа не только смертности, но и брачности, рождаемости, возрастной структуры населения, для разработки прогнозов численности и структуры населения. Методологические принципы той же модели находят применение и в других науках. Она применяется и в исследованиях миграционных процессов, и в социологии при изучении и прогнозировании социальной мобильности, и в экономике труда при изучении текучести рабочей силы, и в здравоохранении для прогнозирования заболеваемости. На ней базируются и финансовые расчёты по страхованию жизни.

Одними из основных показателей цивилизованности страны, как считают ООН и ВОЗ, являются уровень здоровья и продолжительность жизни её населения. По указу президента России В.В. Путина повышение ожидаемой продолжительности жизни населения является национальной целью России на 2030 год.

Изучение смертности по причинам смерти позволяет получить более полную картину демографической ситуации как в России в целом, так и на отдельных её территориях, определить факторы, влияя на которые можно увеличить продолжительность жизни населения.

Цель и задачи работы

Цель настоящей работы заключается в повышении качества оценки и анализа показателей смертности населения России на основе разработки соответствующего математического и программного обеспечения.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Провести аналитический обзор существующего программного обеспечения оценки и анализа показателей смертности.
2. Проанализировать математические модели и методы построения таблиц смертности, вычисления функциональных характеристик продолжительности жизни, преобразования кратких таблиц смертности в полные.
3. Адаптировать и модифицировать существующие методы построения таблиц смертности и преобразования кратких таблиц смертности в полные.
4. Разработать программное обеспечение оценки и анализа показателей смертности населения России.
5. Провести вычислительный эксперимент и проанализировать результаты.

Содержательная постановка задачи

Дано:

- таблица с коэффициентами смертности населения России;
- таблица со среднегодовой численностью населения России.

Требуется решить задачу оценки и анализа показателей смертности населения России. В частности, необходимо решить следующие подзадачи:

- построение таблиц смертности;
- вычисление функциональных характеристик продолжительности жизни;
- преобразование кратких таблиц смертности в полные.

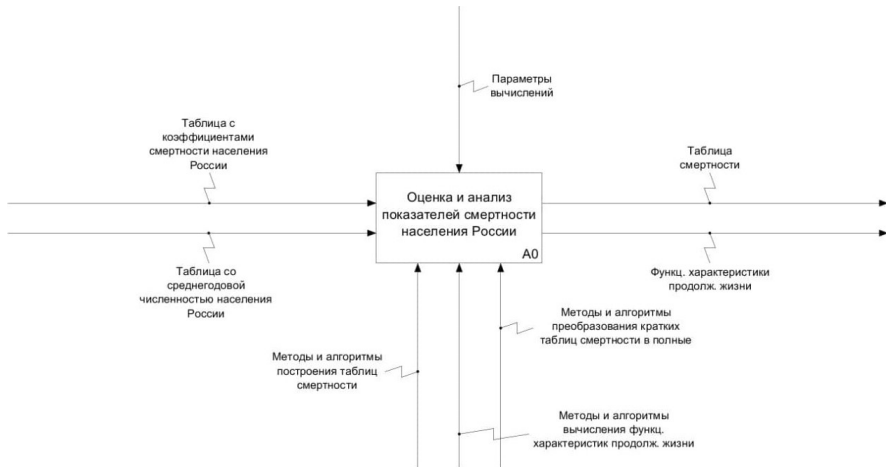
Обзор существующего программного обеспечения для решения поставленной задачи

К сравнению отобрано следующее программное обеспечение:

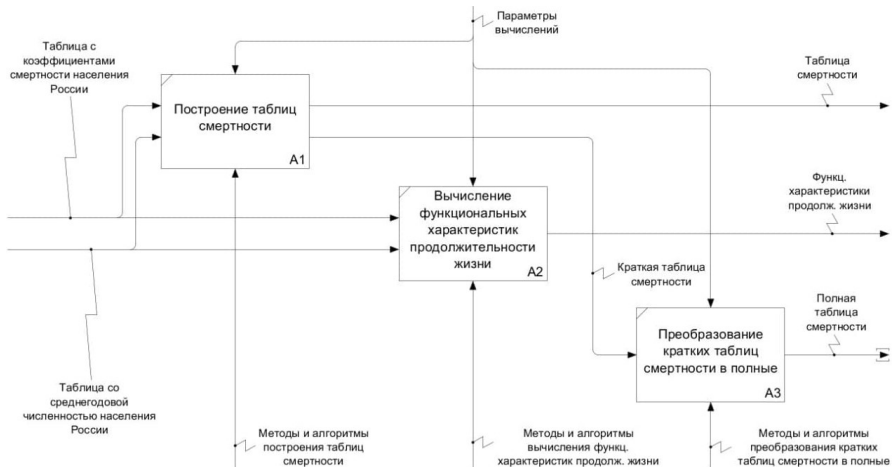
- Сайт «Справочник статистических показателей» Института демографии имени А.Г. Вишневского (Москва, Россия).
- Сайт «The Human Mortality Database» Института демографических исследований им. Макса Планка (Росток, Германия) и отделения демографии Калифорнийского университета (Беркли, США).
- Сайт «Оперативная статистика» Федеральной службы государственной статистики – Росстат (Москва, Россия).

	Справочник статистических показателей	The Human Mortality Database	Оперативная статистика
Данные по территориям России	+	–	+
Данные по категориям поселения	–	–	–
Данные по полу	+	+	–
Данные по причинам смерти	–	–	–
Построение таблиц смертности	+	+	–
Вычисление функциональных характеристик продолжительности жизни	–	–	+
Преобразование кратких таблиц смертности в полные	–	–	–
Визуализация и интерактивность	–	–	+

Формальная постановка задачи



Декомпозиция общей задачи



Задача построения таблиц смертности

Постановка задачи

Дано:

- таблица с коэффициентами смертности населения m_x по территориям, годам, категориям поселения, полу, возрастным группам и (при наличии) причинам смерти;
- таблица со среднегодовой численностью населения L_x по территориям, годам, категориям поселения, полу и возрастным группам;
- предельный возраст ω ;
- k -летние группы возраста.

Требуется построить таблицу смертности с показателями x , l_x , d_x , p_x , q_x , L'_x , e_x .

Для решения поставленной задачи используется демографический метод построения таблиц смертности.

Задача построения таблиц смертности

Решение

1. Выбрать параметры вычислений: территорию, год, категорию поселения, пол, причину смерти (при наличии).
2. Вычислить число умерших d_x по возрастным группам по формуле $d_x = m_x \times L_x$.
3. Вычислить число доживших l_x по возрастным группам по формуле $l_x = {}_k d_x + l_{x+k}$. Вычисления следует начинать с предельного возраста ω , для которой $l_\omega = 0$.
4. Нормировать значения числа доживших l_x так, чтобы значение l_0 стало равно 1.
5. Умножить нормированные значения числа доживших l_x на 10^5 .

Задача построения таблиц смертности

Решение

6. Вычислить вероятность дожить до точного возраста $x + k$ для тех, кто дожил до возраста x , обозначаемую ${}_k p_x$, по формуле ${}_k p_x = \frac{l_{x+k}}{l_x}$.
7. Вычислить вероятность умереть в интервале между точным возрастом x и возрастом $x + k$ для доживших до возраста x , обозначаемую ${}_k q_x$, по формуле ${}_k q_x = 1 - {}_k p_x$ или ${}_k q_x = \frac{l_x - l_{x+k}}{l_x}$.
8. Вычислить число человеко-лет, прожитых лицами одного поколения в интервале возраста от x до $x + k$, обозначаемое L'_x , по формуле $L'_x \approx \frac{l_x + l_{x+k}}{2}$.
9. Вычислить среднюю продолжительность предстоящей жизни e_x по возрастным группам по формуле $e_x = \frac{\sum_x^{\omega-k} L'_x}{l_x}$.
10. Объединить вычисленные показатели в таблицу.

Задача вычисления функциональных характеристик продолжительности жизни

Постановка задачи

Дано:

- таблица с коэффициентами смертности населения m_x по территориям, годам, категориям поселения, полу, возрастным группам и (при наличии) причинам смерти;
- таблица со среднегодовой численностью населения L_x по территориям, годам, категориям поселения, полу и возрастным группам;
- предельный возраст ω ;
- k -летние группы возраста.

Требуется построить графики функциональных характеристик продолжительности жизни новорождённого $F(x)$, $s(x)$, $f(x)$, $\mu(x)$ и функциональных характеристик остаточной продолжительности жизни $F_x(t)$, $s_x(t)$, $f_x(t)$, $\mu_x(t)$.

Для решения поставленной задачи используется метод вычисления статистических оценок функциональных характеристик продолжительности жизни.

Задача вычисления функциональных характеристик продолжительности жизни

Решение

1. Выбрать параметры вычислений: территорию, год, категорию поселения, пол, причину смерти (при наличии).
2. Вычислить число умерших d_x по возрастным группам по формуле
$$d_x = m_x \times L_x.$$
3. Вычислить число доживших l_x по возрастным группам по формуле
$$l_x = {}_k d_x + l_{x+k}.$$
 Вычисления следует начинать с предельного возраста ω , для которой $l_\omega = 0$.
4. Вычислить статистическую оценку функции распределения продолжительности жизни новорождённого $F(x)$ по формуле
$$F(x) = 1 - \frac{l_x}{l_0}.$$
5. Вычислить статистическую оценку функции выживания для новорождённого $s(x)$ по формуле
$$s(x) = \frac{l_x}{l_0}.$$
6. Вычислить статистическую оценку кривой смертей для новорождённого $f(x)$ по формуле
$$f(x) \approx \frac{d_x}{l_0}.$$
7. Вычислить статистическую оценку интенсивности смертности для новорождённого $\mu(x)$ по формуле
$$\mu(x) \approx \frac{d_x}{l_x}.$$

Задача вычисления функциональных характеристик продолжительности жизни

Решение

8. Вычислить статистическую оценку функции распределения остаточной продолжительности жизни человека в возрасте x , обозначаемую $F_x(t)$, по формуле $F_x(t) = 1 - \frac{l_{x+t}}{l_x}$.
9. Вычислить статистическую оценку функции выживания для человека в возрасте x , обозначаемую $s_x(t)$, по формуле $s_x(t) = \frac{l_{x+t}}{l_x}$.
10. Вычислить статистическую оценку кривой смертей для человека в возрасте x , обозначаемую $f_x(t)$, по формуле $f_x(t) \approx \frac{d_{x+t}}{l_x}$.
11. Вычислить статистическую оценку интенсивности смертности для человека в возрасте x , обозначаемую $\mu_x(t)$, формуле $\mu_x(t) \approx \frac{d_{x+t}}{l_{x+t}}$.
12. Построить графики вычисленных статистических оценок функциональных характеристик.

Задача преобразования кратких таблиц смертности в полные

Постановка задачи

Дано:

- краткая таблица смертности;
- k -летние группы возраста.

Требуется получить полную таблицу смертности.

Для решения поставленной задачи используется метод интерполяции значений числа доживших l_x краткой таблицы смертности по одному из трёх предположений: о равномерном распределении смертей, постоянства интенсивности смертности, Балдуччи.

Задача преобразования кратких таблиц смертности в полные

Решение

1. Выполнить интерполяцию значений числа доживших l_x краткой таблицы смертности по одному из предположений:
 - о равномерном распределении смертей по формуле $l_{x+t} = l_x + (l_{x+k} - l_x) \times \frac{t}{k}$, где $0 \leq t \leq k$;
 - постоянства интенсивности смертности по формуле $l_{x+t} = l_x \times \left(\frac{l_{x+k}}{l_x}\right)^{\frac{t}{k}}$, где $0 \leq t \leq k$;
 - Балдуччи по формуле $l_{x+t} = \frac{k \times l_{x+k}}{(k-t) \times \frac{l_{x+k}}{l_x} + t}$, где $0 \leq t \leq k$.
2. Построить полную таблицу смертности на основе интерполированных значений числа доживших l_{x+t} .

Структура программного обеспечения

АнализВыживаемости

Class

Fields

- списокДаннымиОСмертностиПоОднолетиюВозрастнымГруппам : List<СмертностьПоОднолетиюВозрастнымГруппам>
- списокДаннымиОСмертностиПоПятилетиюВозрастнымГруппам : List<СмертностьПоПятилетиюВозрастнымГруппам>
- списокДаннымиОЧисленностиНаселенияПоОднолетиюВозрастнымГруппам : List<ЧисленностьНаселенияПоОднолетиюВозрастнымГруппам>
- списокДаннымиОЧисленностиНаселенияПоПятилетиюВозрастнымГруппам : List<ЧисленностьНаселенияПоПятилетиюВозрастнымГруппам>
- ПредельнаяВозрастнаяГруппа : int

Properties

- СписокОдов : List<int>
- СписокКатегорийПоселения : List<КатегорияПоселения>
- СписокОднолетиюВозрастныхГрупп : List<int>
- СписокПолов : List<Пол>
- СписокПредположенийДляИнтерполяции : List<string>
- СписокПричинСмерти : List<ПричинаСмерти>
- СписокПятилетиюВозрастныхГрупп : List<int>
- СписокТерриторий : List<Территория>

Methods

- АнализВыживаемости()
- ВыполнитьИнтерполяциюПоПредположениюБалдуцци() : List<ЧислоЖивыхДляВозрастнойГруппы>
- ВыполнитьИнтерполяциюПоПредположениюРавномерногоРаспределенияСмертей() : List<ЧислоЖивыхДляВозрастнойГруппы>
- ВыполнитьИнтерполяциюПоПредположениюПостоянстваИнтенсивностиСмерти() : List<ЧислоЖивыхДляВозрастнойГруппы>
- ВыполнитьИнтерполяциюЧислаЖивыхПоВозрастнымГруппам() : List<ЧислоЖивыхДляВозрастнойГруппы>
- ВычислениеДанныхВозможно() : bool
- ВычислитьДанныеДляТаблицыСмертности() : List<СтрокаТаблицыСмертности>
- ВычислитьДанныеДляТаблицыСождаемойПродолжительностьюПредстоящейЖизни() : List<СтрокаТаблицыСождаемойПродолжительностьюПредстоящейЖизни>
- ВычислитьФункциональныеХарактеристикиОстаточнойПродолжительностиЖизни() : List<ФункциональныеХарактеристикиПродолжительностиЖизни>
- ВычислитьФункциональныеХарактеристикиПродолжительностиЖизниНоворожденного() : List<ФункциональныеХарактеристикиПродолжительностиЖизни>
- ЗаполнитьСписокВозрастныхГрупп() : void
- ЗаполнитьСписокОдов() : void
- ЗаполнитьСписокКатегорийПоселения() : void
- ЗаполнитьСписокПолов() : void
- ЗаполнитьСписокПредположенийДляИнтерполяции() : void
- НормализоватьЧислоЖивыхПоВозрастнымГруппам() : void
- ОбработатьВходныеДанные() : void
- ПодобратьПредположениеДляИнтерполяции() : РезультатыВычислительногоЭксперимента
- ПолучитьДанныеОСмертности() : Смертность
- ПолучитьДанныеОЧисленностиНаселения() : ЧисленностьНаселения
- СравнитьОпытнуюТаблицуСмертностиСЭталонной() : РезультатыВычислительногоЭксперимента

ЧислоУмершихДляВозрастнойГруппы

Class

ЧислоЖивыхДляВозрастнойГруппы

Class

СуммарноеСреднееЧислоЛетПрожитоеЧленамиСообщности

Class

ОбщееСреднееЧислоЛетКДожитию

Class

СтрокаТаблицыСмертности

Class

СтрокаТаблицыСождаемойПродолжительностьюПредстоящейЖизни

Class

ФункциональныеХарактеристикиПродолжительностиЖизни

Class

Демонстрация работы программного обеспечения

Построение таблицы смертности

Программное обеспечение оценки и анализа показателей смертности населения России

Тип возрастных групп

☐ Однолетние

☒ Пятилетние

☐ Однолетние интерполированные

Год: 2019

Территория: Республика Башкортостан

Категория поселения: Городское население

Пол: Мужчины

Причина смерти: Все причины смерти

Возрастная группа: 0

Предп. для интерполяции: 0 равномерном распределении смертей

Построить таблицу смертности

Построить таблицу ожидаемой продолжительности предстоящей жизни человека по причинам смерти

Построить графики функциональных характеристик продолжительности жизни новорожденного

Построить графики функциональных характеристик остаточной продолжительности жизни

Сравнить опытную таблицу смертности с эталонной (вычислительный эксперимент №1)

Подобрать предположение для интерполяции (вычислительный эксперимент №2)

Таблица смертности и графики успешно построены

Таблица смертности

x	lx исх.	lx	dx	Lx	Tx	qx	qx	ex
0	14173	100000	536	99732	6473168	0.99464	0.00536	64.73168
1	14097	99464	134	997588	6373436	0.99865	0.00135	64.07782
5	14078	99330	106	496385	5975848	0.99893	0.00107	60.16156
10	14063	99224	106	495855	5479463	0.99893	0.00107	55.22316
15	14048	99118	261	494938	4983608	0.99737	0.00263	50.27955
20	14011	98857	452	493155	4488670	0.99543	0.00457	45.40569
25	13947	98405	1051	489398	3995515	0.98932	0.01068	40.60276
30	13798	97354	2568	480350	3506117	0.97362	0.02638	36.0141
35	13434	94786	4099	463682	3025767	0.95676	0.04324	31.92209
40	12853	90687	4403	442428	2562085	0.95145	0.04855	28.25195
45	12229	86284	4495	420182	2119657	0.9479	0.0521	24.56605
50	11592	81789	6032	393865	1699475	0.92625	0.07375	20.77877
55	10737	75757	10273	353102	1305610	0.8644	0.1356	17.23418
60	9281	65484	13533	293588	952508	0.79334	0.20666	14.54566
65	7363	51951	14119	224458	658920	0.72822	0.27178	12.68349
70	5362	37332	9652	165030	434462	0.74487	0.25513	11.48398
75	3994	28180	8283	120192	269432	0.70607	0.29393	9.56111
80	2820	19897	10421	73432	149240	0.47625	0.52375	7.50063
85	1343	9476	9476	75808	75808	0	1	8
101	0	0	0	0	0	0	1	0

Таблица ожидаемой продолжительности предстоящей жизни человека по причинам смерти

Графики

Вероятн. дожития до след. возр. гр.

Оценка средн. продолж. жизни

Демонстрация работы программного обеспечения

Построение таблицы ожидаемой продолжительности предстоящей жизни человека по причинам смерти

Программное обеспечение оценки и анализа показателей смертности населения России

Тип возрастных групп

☐ Однолетние

☒ Пятилетние

☐ Однолетние интерполированные

Год: 2019

Территория: Российская Федерация после 17 марта 2014

Категория населения: Все население

Пол: Женщины

Причина смерти: Все причины смерти

Возрастная группа: 0

Предпол. для интерполяции: 0 равномерном распределении смертей

Построить таблицу смертности

Построить таблицу ожидаемой продолжительности предстоящей жизни человека по причинам смерти

Построить графики функциональных характеристик продолжительности жизни новорожденного

Построить графики функциональных характеристик остаточной продолжительности жизни

Сравнить опытную таблицу смертности с эталонной (вычислительный эксперимент NF1)

Подобрать предположение для интерполяции (вычислительный эксперимент NF2)

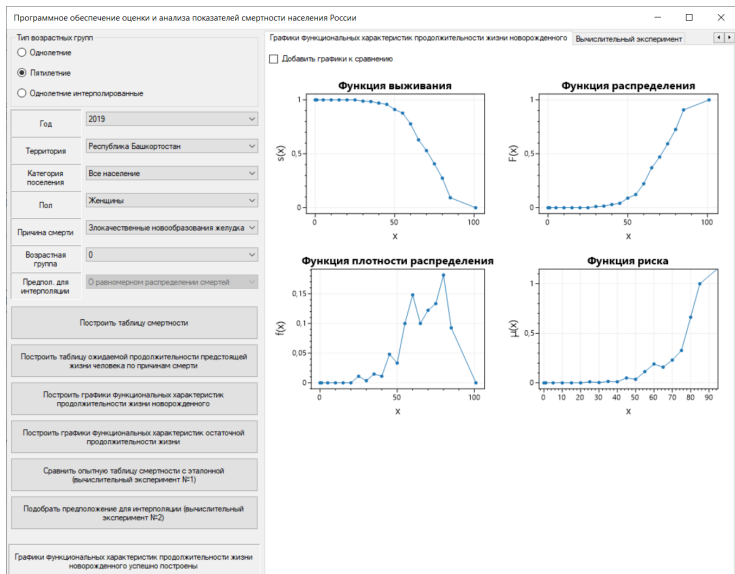
Таблица ожидаемой продолжительности предстоящей жизни человека по причинам смерти успешно построена

Таблица смертности | Таблица ожидаемой продолжительности предстоящей жизни человека по причинам смерти | Графики

Причина смерти	qx	ex
Гипертоническая болезнь с преимущественным поражением сердца	0	79.52747
Железочнокаменная болезнь (холецистит)	0	79.55678
Травмы конечностей	0	79.71911
Гипертоническая болезнь с преимущественным поражением почек	0	79.76262
Другие злокачественные новообразования кожи	0	79.81719
Инфаркт мозга	2E-05	79.89751
Инсульт, не уточненный как кровоизлияние или инфаркт	0	79.98626
Хроническая ишемическая болезнь сердца, неуточненная	0	80.30388
Сосудистые болезни кишечника	0	80.65856
Прочие формы хронической ишемической болезни сердца	0	80.88831
Болезнь Паркинсона и вторичный паркинсонизм	0.00055	81.08536
Холецистит	0	81.14101
Другие болезни костномышечной системы и соединительной ткани	0	81.19982
Последствия цереброваскулярных болезней	0	81.38087
Прочие цереброваскулярные болезни	0	81.44777
Атеросклеротическая болезнь сердца	0	81.82448
Болезнь Альцгеймера	0	81.92428
Эмфизема	0	82.20237
Острая ревматическая лихорадка	0	82.36113
Атеросклеротическая сердечно-сосудистая болезнь, так описанная	0	82.43079
Другие сальмонеллезные инфекции	0	82.5
Атеросклероз	0	82.72139
Гипертоническая энцефалопатия	0	83.06372
Прочие нарушения нервной системы	0.00132	83.48676
Цереброваскулярная болезнь, неуточненная	0	84.69808
Другие похозы	0	84.97564
Церебральный атеросклероз	0	85.12284
Другие уточненные поражения сосудов мозга	3E-05	85.28081
Пептическая язва, неуточненной локализации	0	85.49996
Дегенерация миокарда	0	86.45854
Старость	0	90.921

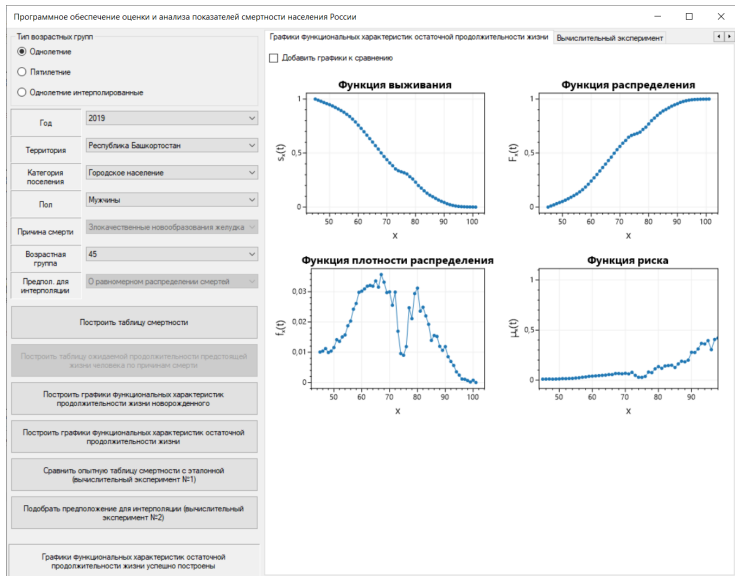
Демонстрация работы программного обеспечения

Построение графиков функциональных характеристик продолжительности жизни новорождённого



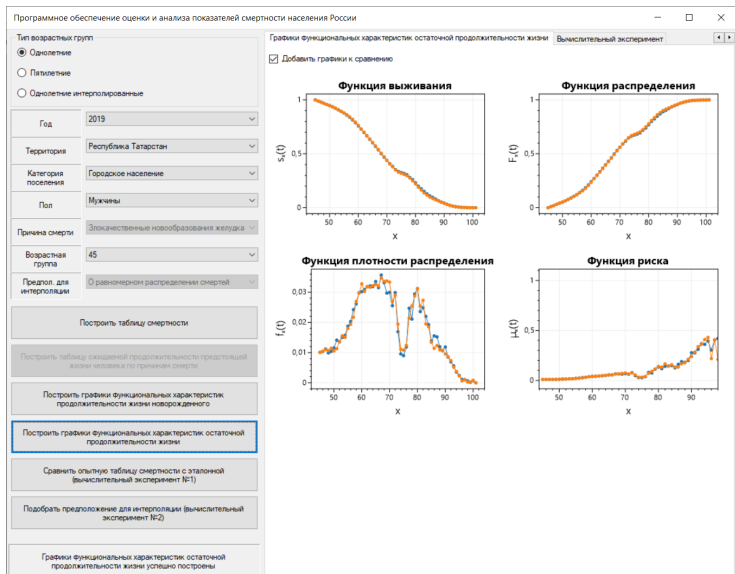
Демонстрация работы программного обеспечения

Построение графиков функциональных характеристик остаточной продолжительности жизни



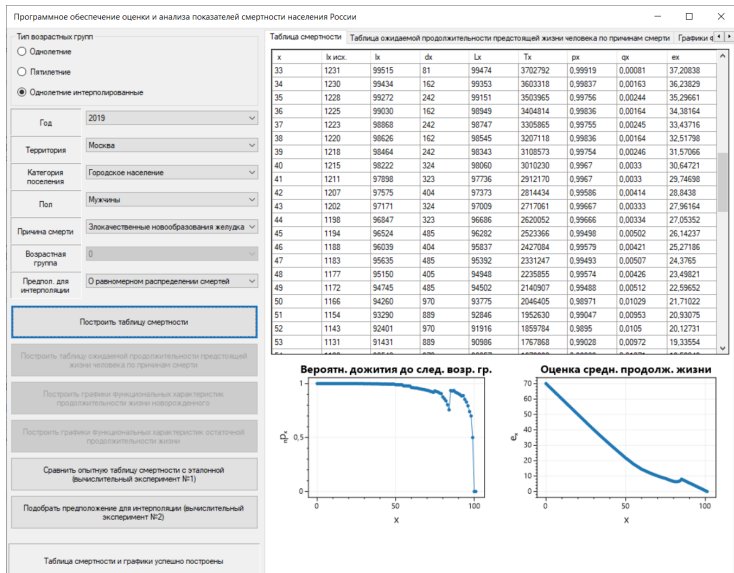
Демонстрация работы программного обеспечения

Построение графиков функциональных характеристик остаточной продолжительности жизни. Сравнение графиков для разных входных параметров



Демонстрация работы программного обеспечения

Преобразование краткой таблицы смертности в полную



Вычислительные эксперименты

Описание технических средств и входных данных

Вычислительные эксперименты проводились на ЭВМ со следующими характеристиками:

- операционная система Windows 10 Pro версии 2004 (64-битная архитектура);
- процессор Intel Core i7-3612QM с тактовой частотой 2.10 ГГц;
- оперативная память объёмом 12 ГБ;
- жёсткий диск объёмом 512 ГБ;
- видеодисплей с разрешением 1920 пикселей в ширину и 1080 пикселей в высоту.

Входные данные взяты из Российской базы данных по рождаемости и смертности (РосБРС) Центра демографических исследований Российской Экономической Школы (ЦДИ РЭШ). Входные данные включают в себя:

- перечень территорий России с кодами системы обозначения объектов административно-территориального деления (СОАТО);
- перечень причин смерти в соответствии с Краткой номенклатурой причин смерти;
- коэффициенты смертности по возрасту (однолетние группы), категории населения и полу в России и на её территориях за 1989–2019 гг.;
- коэффициенты смертности по причинам смерти, возрасту (пятилетние группы), категории населения и полу в России и на её территориях за 2015–2019 гг.;
- среднегодовое население по территориям России для расчёта коэффициентов смертности за 1989–2019 гг. по однолетним возрастным группам;
- среднегодовое население по территориям России для расчёта коэффициентов смертности за 2015–2019 гг. по пятилетним возрастным группам.

Вычислительный эксперимент №1

Описание

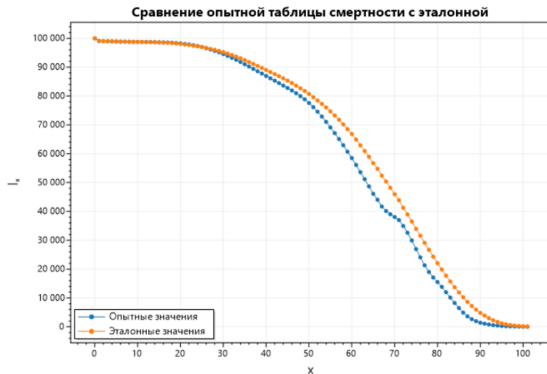
Входные параметры для построения опытной таблицы смертности:

- тип возрастных групп – однолетние;
- год – 2013;
- территория – Российская Федерация до 17 марта 2014;
- категория поселения – все население;
- пол – мужчины;
- причина смерти – все причины смерти;
- интерполяция – отсутствует.

В качестве критерия оценки точности принимается сумма расстояний между точками, обозначающими среднее число l_x членов совокупности, доживших до возрастной группы x , взятая для трёх возрастных промежутков: 6–17 лет, 30–55 лет и 56–80 лет.

Вычислительный эксперимент №1

Результаты



Результаты оценки точности:

- $A_{6-17} = 397$;
- $A_{30-55} = 63375$;
- $A_{56-80} = 203617$.

Вычислительный эксперимент №2

Описание

Входные параметры для построения реальной таблицы смертности:

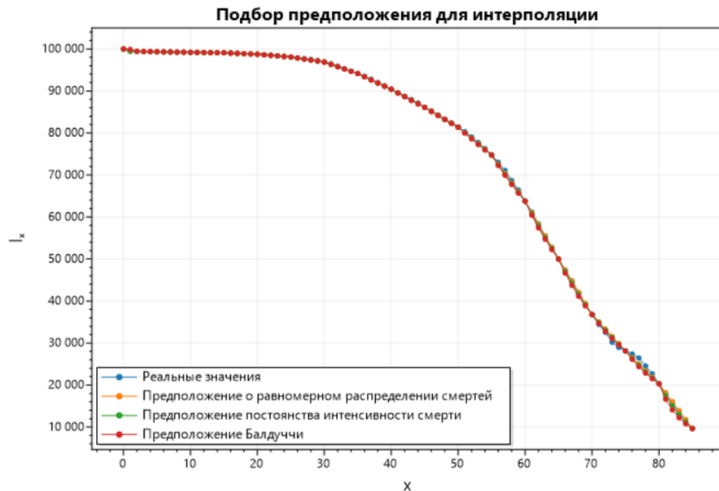
- тип возрастных групп – однолетние;
- год – 2019;
- территория – Республика Башкортостан;
- категория поселения – все население;
- пол – мужчины;
- причина смерти – все причины смерти;
- интерполяция – отсутствует.

Входные параметры для построения интерполированных таблиц смертности:

- тип возрастных групп – пятилетние;
- год – 2019;
- территория – Республика Башкортостан;
- категория поселения – все население;
- пол – мужчины;
- причина смерти – все причины смерти;
- интерполяция – по предположению о равномерном распределении смертей, или по предположению постоянства интенсивности смерти, или по предположению Балдуччи.

Вычислительный эксперимент №2

Результаты



Вычислительный эксперимент №2

Результаты

Результаты оценки точности интерполирования по предположению о равномерном распределении смертей:

- $A_{6-17} = 114$;
- $A_{30-55} = 2405$;
- $A_{56-80} = 10868$.

Результаты оценки точности интерполирования по предположению постоянства интенсивности смерти:

- $A_{6-17} = 114$;
- $A_{30-55} = 2909$;
- $A_{56-80} = 13660$.

Результаты оценки точности интерполирования по предположению Балдуччи:

- $A_{6-17} = 114$;
- $A_{30-55} = 3420$;
- $A_{56-80} = 17839$.

Заключение

В процессе выполнения выпускной квалификационной работы были решены следующие задачи:

1. Проведён аналитический обзор существующего программного обеспечения оценки и анализа показателей смертности.
2. Проанализированы математические модели и методы построения таблиц смертности, вычисления функциональных характеристик продолжительности жизни, преобразования кратких таблиц смертности в полные.
3. Адаптированы и модифицированы существующие методы построения таблиц смертности и преобразования кратких таблиц смертности в полные.
4. Разработано программное обеспечение оценки и анализа показателей смертности населения России.
5. Проведены вычислительные эксперименты.

Спасибо за внимание!