

Математичне моделювання прогнозу смертності чоловіків працездатного віку

Донецький національний медичний університет ім М.Горького

Луганський державний медичний університет

З використанням методів апроксимації емпіричних даних з наступною екстраполяцією невідомої функції тригонометричним рядом Фур'є розроблено модель математичного прогнозування показників смертності чоловіків працездатного віку (ЧПВ). Апроксимацію та екстраполяцію здійснювали на часовому інтервалі (0,T), де T=10 (кількість років, за які зібрані статистичні дані смертності ЧПВ області) тригонометричним многочленом N-го порядку функцією $y = f(x)$, для якої відомі m (поточне значення кількості років) її значень ($m=10$), $y_k = f(x_k)$, де $X_k = \frac{kT}{m}$ ($k=0,1,2,3\dots m-1$),

де Y_k – емпіричні значення функції смертності (статистичні дані кількості смертей), тобто смертність в k-ий рік дослідження ($k=0,1,2,3\dots m-1$)

k – порядковий номер досліджуваного року,

(0,T) – досліджуваний інтервал

Тригонометричний многочлен має наступний вид (1) :

$$Q_N(x) = \frac{a_0}{2} + \sum_{n=1}^N (a_n \cos(n \frac{2\pi}{T}x) + b_n \sin(n \frac{2\pi}{T}x)) , (0 \leq x \leq T), \quad (1)$$

де n – номер відповідного члена (коефіцієнта) ряду ($a_0, a_1, a_2, a_3, b_1, b_2, b_3$),

N – порядок многочлена (в випадку наших розрахунків N=3, що достатньо для отримання достатньої точності апроксимації)

Коефіцієнти тригонометричного ряду a_n, b_n визначаються наступними співвідношеннями (2):

$$\begin{aligned} a_n &= \frac{2}{T} \int_0^T f(x) \cos(n \frac{2\pi}{T}x) dx \\ b_n &= \frac{2}{T} \int_0^T f(x) \sin(n \frac{2\pi}{T}x) dx \end{aligned} \quad (2)$$

Так як в вищезазначених формулах нам невідомий аналітичний вид функції смертності, але відомі емпіричні значення функції з метою обчислення коефіцієнтів, застосовували в співвідношеннях (2) формулу прямокутників (4.3) для обчислювання інтегралів за значеннями підінтегральних виразів в точках $x_k = \frac{kT}{m}$ ($k = 0,1,2,\dots, m-1$) та отримали формули для обчислення коефіцієнтів a_n, b_n за якими будуть отримані коефіцієнти ряду ($a_0, a_1, a_2, a_3, b_1, b_2, b_3$):

$$a_n = \frac{2}{m} \sum_{k=0}^{m-1} y_k \cos\left(n \frac{2\pi k}{m}\right)$$

$$b_n = \frac{2}{m} \sum_{k=0}^{m-1} y_k \sin\left(n \frac{2\pi k}{m}\right), n = 0, 1, 2, \dots, N$$
(3)

Відомо, що на відміну від методу найменших квадратів, тригонометричний ряд дає найкраще наближення до функції $f(x)$, іншими словами, коефіцієнти (4.3) мінімізують суму квадратів відхилень (4):

$$S_N^2 = \sum_{k=0}^{m-1} [Q_N(x_k) - y_k]^2.$$
(4),

де $Q_N(x_k)$ – значення (кількість смертей) многочлена (4.1) в точках X_k

Y_k – емпіричне значення функції смертності (статистика смертності за кожний рік).

Таким чином многочлен (1), коефіцієнти якого для кожної з моделей знаходимо за формулами (3) слугує рішенням поставленої задачі математичного моделювання і прогнозування динаміки смертності ЧПВ.

Отримана математична модель прогнозування смертності була адаптована та апробована для кожної вибірки (за причиною смерті, віковою групою, містом мешкання) в залежності від врахованих ознак для прогнозування, для чого були обчислені їх власні коефіцієнти тригонометричного многочлена (1) за формулами (3). У розрахунках обмежилися сьома першими членами ряду ($a_0, a_1, a_2, a_3, b_1, b_2, b_3$), що є достатнім для досягнення необхідної точності моделей.

Для оцінки інформативності та значущості моделі здійснено дисперсійний аналіз, в результаті якого були розраховані:

– коефіцієнт детермінації (5)

$$R^2 = \frac{SS_R}{SS},$$
(5)

де SS_R – сума квадратів відхилень, розрахованих значень \hat{y}_i (абсолютна кількість смертей за i -ий рік) від середнього \bar{y} (середнє значення кількості смертей за досліджуваний період, математичне очікування).

SS – сума квадратів відхилень значень y_i , що спостерігалися від середнього \bar{y} . Була використана наступна градація інформативності отриманих моделей: задовільна – $0,3 < R^2 < 0,5$; достатня – $0,5 < R^2 < 0,7$; гарна – $0,7 < R^2 < 0,9$; оптимальна – $R^2 < 0,9$.

– F – критерій Фішера (6)

$$F = \frac{S_R^2}{S_0^2},$$
(6)

де S_R^2 – дисперсія відхилень \hat{y}_i від середнього \bar{y} ; S_0^2 – дисперсія відхилень y_i від \hat{y}_i .

Також здійснено аналіз помилки апроксимації як середнього відхилення модельних значень від експериментальних значень, так і відхилень модельних значень від середнього із застосуванням коефіцієнта кореляції Пірсона, який розраховували за формулою:

$$r_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}} \quad (7)$$

В цілому, модель вважають інформативною при $R^2 > 0,5$ і значущою, достовірною при рівні значущості за F – критерієм $p \leq 0,05$, або коли F – критерій Фішера табличний більший за F – критерій Фішера розрахунковий (модельний) для даного числа ступенів свободи. Результати математичного моделювання показали задовільні параметри інформативності, значущості та достовірності побудованих моделей.

Важливо, що створена модель реагує на вхідні дані, тобто є динамічною, а не статичною, її точність пропорційна довжині часового ряду. З подальшим набором статистики смертності значення коефіцієнтів многочлена будуть уточнюватися, тим самим будуть уточнюватися і прогнозні показники смертності ЧПВ.

На підставі отриманих математичних прогностичних моделей був створений розрахунковий модуль на основі пакету програми Microsoft Excel з прогнозування загальної та повікової смертності ЧПВ Луганської області та м. Луганська в залежності від причин смерті та місця проживання і отримані прогностичні коефіцієнти смертності на період до 2030 р., які будуть уточнюватися з отриманням статистичних даних про смерть ЧПВ за кожний послідуєчий рік.

Дана методика прогнозування заснована на статистичних даних про кількість випадків смерті, віці і місці проживання, могла би давати близьку до абсолютної точність прогнозу смертності при наявності даних про керуючі ознаки (шкідливі звички, наявність хронічних захворювань, перебування під диспансерним наглядом, спадковість, матеріальне становище, тощо) і обов'язковому їх обліку при прогнозуванні.

За прогнозом у найближчі роки при відсутності адекватних і ефективних заходів профілактики Луганську область очікує зростання смертності ЧПВ від основних причин смерті у практично всіх вікових групах, особливо це стосується хвороб травлення, ХСК, зовнішніх причин смерті, новоутворень. В загальній структурі смертності очікується зростання смертності ЧПВ від інфекційних хвороб до 91,82 на 100 тис. ЧПВ області в 2015 р. (в 1,4 рази), від новоутворень в 1,1 рази (до 119,3 в 2016-2017 рр. відповідно), від хвороб нервової системи до 35,2 на 100 тис. ЧПВ області в 2015 р. (в 1,6 рази), від ХСК в 2015-2018 рр. в 1,3 рази до 356,5 на 100 тис. ЧПВ області, від хвороб дихальної системи (68,45 відповідно) в 2013-2015 рр. в 1,8 рази, хвороб травлення в 2016-2017 рр. до 150,7 на 100 тис. ЧПВ (в 1,4 рази), від зовнішніх причин в 1,6 рази до 335,6 в період 2012-2018 рр..

Прогностично встановлено, що в кожній віковій групі рівень смертності ЧПВ та її структура мають свої особливості.