

УДК 614.7:(614.1+613.6)

МЕТОДИ МАТЕМАТИЧНОЇ СТАТИСТИКИ В ЗАДАЧАХ ОЦІНКИ ТА УПРАВЛІННЯ ЕКОЛОГІЧНИМИ РИЗИКАМИ

Шишкіна А.Ю., спонукач, Козловська Т.Ф., к.х.н., доц.

*Кременчуцький державний політехнічний університет імені Михайла Остроградського
39614, м. Кременчук, вул. Першотравнева, 20*

E-mail: ecol@polytech.poltava.ua

Рассмотрены и проанализированы методы математической статистики с целью выбора оптимального для анализа, характеристик и оценки канцерогенного воздействия химических поллютантов окружающей природной среды на состояние здоровья человека.

Ключевые слова: факторы риска, химические загрязнители, математическая статистика.

Is inspect and analyzed methods of mathematic statistic to offer a selections for analysis, characteristic and assess cancerogenic influence of chemic pollutants the neighbouring environment of the situation health beings

Keywords: factors of risk, chemical pollutions, mathematic statistic.

Вступ. Поняття ризику пов'язано з імовірнісною уявою про вплив чинників навколишнього середовища на здоров'я людини. Величину ризику не можна безпосередньо вимірити, а можливо лише з деяким ступенем надійності оцінити при застосуванні кількісних характеристик чинників ризику та показників здоров'я населення, яке знаходиться під впливом цих чинників.

Чинниками ризику вважаються чинники будь-якої природи – фізичні, метеорологічні, генетичні, соціальні, поведінки і т.і., які мають достовірний зв'язок з порушенням здоров'я.

Кількісна оцінка достовірності емпірично встановлених зв'язків здійснюється за допомогою методів математичної статистики, які дозволяють урахувати невизначеності, що пов'язані як з вимірами впливів, так й визначенням характеру взаємозв'язків між впливами та ефектами. В основі статистичних методів лежить математичний апарат теорії ймовірностей [1].

Аналіз попередніх досліджень. При оцінці медичних ризиків [2] статистичні методи застосовують для вирішення наступних класів задач: перевірка припущення про те, що даний чинник (безпосередньо або в комбінації з іншими чинниками) є чинником ризику; дослідження питання про те, які рівні чинника ризику відповідає різним ризикам (має різний вплив на здоров'я); отримання залежностей “доза-ефект”.

Мета роботи. Вибір оптимальної статистичної математичної моделі для аналізу й оцінки канцерогенного впливу хімічних забруднювачів компонентів довкілля як міри визначення канцерогенного ризику.

Матеріал і результати досліджень. Роль статистичних методів в соціально-гігієнічному моніторингу визначається можливостями застосовувати ці методи з метою стискання та збільшення оглядності інформації, яка реєструється в процесі моніторингу; аналізу взаємозв'язків між

чинниками зовнішнього середовища та показниками здоров'я; аналізу ефективності дій щодо управління здоров'ям населення.

З метою обробки отриманих даних застосовують комп'ютерні прикладні програми типу STATGRAPHICS, BMDP, SPSS, STATISTICA. Зокрема, в наших розрахунках ми застосовували програмний комплекс Excel-STATISTICA.

При виборі і розробці критеріїв й параметрів екологічних факторів та їхньої оцінки, а також оцінки показників зміни стану здоров'я людей під впливом навколишнього середовища неминує виникати наступні проблеми:

- 1) ступінь вірогідності метричного чи бального вираження показника;
- 2) його значимість (універсальність) у часі і просторі;
- 3) можливість екстраполяції і співставлення одержуваних значень показників;
- 4) ступінь (напрямок) зміни значимості показників при їхньому сполученні і можливості їхнього підсумовування;
- 5) можливості практичного одержання показників (матеріально-технічні, фінансові, кадрові й організаційні реалії).

Багато факторів ризику ще не вивчені (наприклад, гепатогенні зони). Для оцінки ряду факторів ризику не вистачає достатньої інформаційної бази й експериментальних даних. Нерідко самі поллютанти не є токсичними, а шкоду здоров'ю заподіюють продукти їхнього розпаду чи взаємодії (синтезу) з іншими речовинами. Рівні концентрації багатьох токсинів у навколишньому середовищі можуть бути обмірювані лише приблизно. Не на всі шкідливі речовини встановлені нормативи їхнього вмісту в навколишньому середовищі (наприклад, вміст кадмію в ґрунтах). Не встановлені припустимі дози (пороги) для канцерогенів, тератогенів, мутагенів і ін.

Надзвичайну важливість, зрозуміло, мають

методична єдність і використання атестованих способів, засобів і лабораторій для виміру оцінюваних екологічних факторів, а також більш широке використання їхнього біотестування.

Постійно ведеться пошук лабораторних методів визначення стану організму, що могли б виявити інтегральні показники здоров'я і його зміни під впливом факторів навколишнього середовища (загальна фізіологічна реактивність, імунологічна реактивність, сполучність периферичної крові, різні ферментні системи, внутрішньоклітинний метаболізм, неспецифічна резистентність і т.д.). Однак методичні засоби рішення цієї задачі ще залишаються недостатньо розробленими, особливо в частині, що стосується оцінки позитивних впливів і комплексного впливу.

Не менші труднощі виникають і при виборі інтегральних критеріїв оцінки регіонів з погляду впливу на здоров'я населення, тобто медико-географічної (медико-екологічної) діагностики техногенно навантажених територій. Для оцінки здоров'я популяцій за рубезем запропонований ряд індексів здоров'я, п'ять з яких перелічуються нижче [2, 3]:

1) показник популяційного здоров'я, що включає наступні компоненти: дитяча смертність, загальні показники смертності і захворюваності, частота показника поширеності окремих хвороб, показник видужання після захворювань, показник видужання після нещасливих випадків, очікувана тривалість життя.

2) Спосіб оцінки індексу здоров'я, що включає дві шкали — шкалу погіршення здоров'я і шкалу недієздатності, які варто розглядати в єдності.

3) Пропорційний показник смертності (ППС), що обчислюється як відношення 50 % смертності в популяції людей старше 50 років до річного показника загальної смертності.

4) Корнельський медичний індекс (КМИ), заснований на запитальнику, що включає 194 компонента (144 — фізичних, 50 — психічних).

5) Метод, запропонований лабораторією вивчення людини, і включає 46 пунктів опитування, що стосується фізичного стану, фізичного розвитку, опірності організму, психічного стану і здатності до соціальної адаптації.

Оцінка стану здоров'я населення на шляху досягнення мети, проголошеної ВІЗ — "здоров'я для всіх", включає дані про смертність, захворюваність, непрацездатність через хворобу, показники соціального і психічного благополуччя, фізіологічні оцінки розвитку, демографічні змінні величини; виділяються показники смертності від окремих причин, дитяча і материнська смертність, середня тривалість майбутнього життя для окремих віків, роки непрожитого життя.

Заслуговує на увагу досвід вивчення (з використанням ЕОМ і математичного моделювання) впливу екологічних факторів на захворюваність населення, проведеного з 1980 р. у Пермському політехнічному інституті, що здійснюється для

чотирьох територіальних рівнів систематизації, узагальнення і глибини пророблення питань (область, місто, район, житловий квартал). У результаті проведеного дослідження отримані:

- кількісна оцінка часткового внеску екологічних факторів у формування захворюваності і смертності населення;

- ранжування пріоритетів як по нозологічним формам, так і по територіях області;

- кількісні критерії екологічного навантаження територій (припустима, помірна, висока, надзвичайно висока) по захворюваності населення (11 показників) і медико-демографічним показникам (народжуваність, загальна смертність, передчасна смертність чоловіків у працездатному віці);

- величини ймовірності екологічної (по окремих факторах) аерогенної і водного регіонального припустимого навантаження (РДН), безпечного за критеріями здоров'я населення з обліком місцевих соціальних умов.

Усе викладене дає підставу зробити наступні висновки:

- багато факторів ризику виникнення "середовищних" преморбідних станів, розладів здоров'я і захворювань не встановлені і не вивчені;

- не визначені уніфіковані інтегральні критерії оцінки стану здоров'я населення;

- не розроблені регіональні медико-екологічні нормативи;

- відсутня апробована система критеріїв і оцінки медико-екологічної ситуації на територіях країни;

- пріоритетне значення має проведення наукових досліджень і розробка практичних рекомендацій медико-екологічного характеру, аналогічних відомим

Для досліджень медико-екологічного спрямування характерним з точки зору теорії ймовірностей є випадковість, яка призводить до результату, що неможливо передбачити заздалегідь, навіть знаючі умови проведення дослідів. Прикладом такої випадковості є медичне обстеження, що в нашому випадку можна вважати однозначним. При цьому не вдається надійно передбачити наявність або відсутність певних захворювань. Результат випадкового дослідження називається елементарною подією.

Повна система елементарних подій – це такий набір елементарних випадків, один з яких обов'язково відбудеться при будь-якому дослідженні із заданим комплексом умов. Іншими словами це називають простором елементарних подій. Випадки, які включають усі можливі елементарні події, тобто які відбуваються обов'язково, називаються достовірними, або повними. Випадок, який не містить жодного елементарного, тобто того, що не відбудеться, зветься пустим, або неможливим.

Слід зазначити, що поняття ймовірності є базовим для кількісного опису ризиків при вивченні впливу чинників навколишнього природного середовища на здоров'я. Ймовірність події – це

міра, яка визначає шанс цієї події при випробуваннях порівняно з іншими випадками. Ризик визначається як ймовірність несприятливих ефектів для здоров'я людини або стану навколишнього середовища [3]. Формально-математично ймовірність визначається наступним чином: для заданого простору елементарних подій Ω ймовірністю називається функція $P(A)$, визначена для будь-якої події A та яка задовольняє наступним умовам (аксіомам):

- 1) ймовірність будь-якої події позитивна – $P(A) \geq 0$ для будь-якого A ;
- 2) ймовірність достовірної (повної) події дорівнює $1 - P(\Omega) = 1$;
- 3) для будь-якої системи неперехрестних подій $\{A_1, A_2, \dots\}$ ймовірність їх поєднання дорівнює сумі ймовірностей цих подій:

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots \quad (1)$$

Із зазначеного випливає, що ймовірність будь-якої події знаходиться в інтервалі від 0 до 1.

При аналізі ризиків необхідно встановити залежності між чинниками ризику, які досліджуються, та показниками здоров'я. При цьому необхідно враховувати можливі залежності між рівнями токсичної речовини в атмосферному повітрі, воді та ґрунтах, що обумовлені загальними джерелами забруднення.

Події вважаються незалежними, якщо ймовірність їх перехрещення дорівнює добутку їх окремих ймовірностей:

$$P(AB) = P(A) * P(B). \quad (2)$$

Із цього виразу можна зробити висновок, що для незалежних подій умовна ймовірність події A за умови B дорівнює безумовно ймовірності A :

$$P(A/B) = \frac{P(AB)}{P(B)} = \frac{P(A) * P(B)}{P(B)} = P(A). \quad (3)$$

Таким же чином умовна ймовірність B за умови A дорівнює умовній ймовірності B .

При дослідженні ризиків слід мати на увазі, що вимірювання як впливів, так й їх наслідків завжди включає деякий елемент невизначеності. У зв'язку з цим чинники ризику розглядаються як випадкові величини.

Випадковою величиною (ξ) називається будь-яка функція, яка задана на чисельності елементарних подій, тобто функція, яка за відповідними правилами приводить у відповідність елементарній події деяке число. Зокрема, якщо випробування полягають у вимірюваннях будь-якої кількісної характеристики чиннику ризику або здоров'я, то випадковою величиною є чисельність значень, що реєструються при кожному випробуванні. Ці значення вважаються вимірюваннями.

Невизначеність при змінами випадкових

величин має декілька джерел:

- неоднорідність простору елементарних подій (при дослідженні показників здоров'я – біологічна неоднорідність, яка притаманна популяції, що досліджується), тобто наявність подій, для яких випадкова величина приймає різні значення;

- випадкові помилки спостереження: помилкова класифікація хворої людини як здорової або навпаки, внаслідок недостатньої надійності діагностичних процедур похибок, пов'язані з обмеженою чутливістю приладів і органів чуттів, за допомогою яких реєструються показники приборів;

- систематичні похибки спостережень (зсувів), обумовлені невірною калібровкою вимірвальних приладів або деякими дефектами органів чуттів людини, яка здійснює вимірювання;

- залежність вивчаємої випадкової величини від інших випадкових величин (наприклад, захворюваності від рівня забруднення води).

Два перших джерела варіабельності забезпечують випадкові, а два останніх – закономірні, або систематичні, зміни випадкової величини.

Важливим аспектом оцінки ризиків є отримання характеристик "доза-ефект", тобто взаємозв'язків між випадковими величинами, відповідними чинниками ризику та показниками здоров'я. Але здійснення таких досліджень повинно проводитись за рахунок:

- обліку інформації про поведінку випадкової величини на значній кількості елементарних подій;

- мінімізації, наскільки це можливо, випадкових помилок спостереження, а також значної кількості спостережень для того, щоб помилки, які кривдять істинне значення реєструємих випадкових величин у бік завищення або заниження, взаємно компенсували один одного;

- виключення систематичних помилок спостереження, але, якщо це неможливо, використання при прийнятті рішень замість безпосередньо виключаємих величин таких функцій від них, які дозволяли б компенсувати

- зсуви (наприклад, різності – для випадку, коли помилки вимірів пов'язані із зсувом начала відліку).

Чинники ризику та показники здоров'я підрозділяються на 2 класу – дискретних і безперервних випадкових величин, - залежно від того, якою є кількість допустимих значень цих величин.

Випадкова величина ξ називається дискретною, якщо вона приймає кінцеву або лічену чисельність різних значень. Між будь-якими двома значеннями дискретної величини може знаходитись лише кінцева кількість її значень. Прикладами дискретних величин є:

- кількість випадків захворювань у певній групі населення за 1 рік;

- вік у повних роках;

- ступінь тяжкості захворювань та пошкоджень [3].

Розподіл дискретної випадкової величини ξ повністю визначено, якщо для кожного її значення x_i указана імовірність його прояву (p_i):

$$p_i = P\{\xi = x_i\}. \quad (4)$$

Функція, яка встановлює взаємозв'язок між величинами x_i та p_i , називається законом розподілу або імовірнісною функцією дискретної випадкової величини. Ця функція може бути заданою у вигляді таблиці або аналітичної залежності:

$$x = \frac{x_1 | x_2 | \dots | x_n}{p_1 | p_2 | \dots | p_n}. \quad (5)$$

Сума імовірностей p_i за всіма значеннями індексу i дорівнює 1, а для будь-якого значення x значення функції розподілу $F(x)$ дорівнює сумі імовірностей p_i при всіх $x_i \leq x$. Таким чином, $F(x)$ є ступеневою функцією, яка зберігає постійне значення на будь-якому інтервалі, що не містить точок x_i , в а кожній точці x_i присутній скачок на величину p_i .

При оцінці ризиків найбільш часто застосовують наступні види розподілів дискретних випадкових величин:

- дискретний рівномірний розподіл;
- розподіл Бернуллі;
- розподіл Пуассона.

У першому випадку для випадкових величин, розподілених за дискретним рівномірним законом, кожне з n можливих значень x_1, \dots, x_n приймається з імовірністю $1/n$. Такі випадкові величини застосовують як моделі подій з рівноймовірносним закінченням. Наприклад, у виробничих технологічних процесах залежно від контакту з агентом ризику розподіл здійснюється за 4 групами:

- 1 – контакт з агентом ризику виключений;
- 2 – епізодичні нетривалі контакти;
- 3 – систематичні нетривалі контакти;
- 4 – постійний контакт з агентом ризику протягом усього робочого часу.

Тоді розподіл персоналу за рівнями даного чиннику є рівномірним.

У випадку розподілу Бернуллі описуються випадкові величини, які приймають два значення – 0 та 1 – з імовірностями, відповідно, $q = 1 - p$ і p . Останній є параметром розподілу. Розподіл Бернуллі може бути застосований для описання індивідуального ризику (наприклад, якщо величина p дорівнює імовірності смерті від раку внаслідок впливу канцерогену, то вона є характеристикою канцерогенного ризику).

При розподілі Пуассона випадкова величина приймає значення 0, 1, 2, ... (будь-яке негативне число) з імовірностями:

$$p_i = \frac{\lambda^i e^{-\lambda}}{i!}, \quad (6)$$

де λ – параметр розподілу Пуассона. Найбільш доцільним є його застосування для

оцінки ризику окремих взаємно незалежних ефектів чинників ризику.

Безперервними називаються величини, які можуть приймати будь-які значення на деякому інтервалі та міститися нескінченна кількість їх значень. До безперервних чинників ризику належать концентрація, накопичена доза, захворюваність, смертність для популяцій, а також будь-які фізіологічні параметри для індивідів.

Для безперервної випадкової величини ξ функція розподілу $F(x)$ безперервна та, крім того, існує безперервна ненегативна функція $f(x)$, називається функцією щільності імовірності (а також функцією щільності або щільністю розподілу).

Цю величину можна застосовувати як функцію розподілу та щільності розподілу при оцінці ризику: на вісі абсцис позначають рівні чинники ризику, тоді функцію $F(x)$ можна інтерпретувати як імовірність несприятливого чиннику ризику до рівня x .

Функція розподілу безперервної випадкової величини набуває вигляду:

$$F(x_o) = P\{\xi \leq x_o\}, \quad (7)$$

де $P\{\xi \leq x_o\}$ – щільність розподілу безперервної випадкової величини.

Нижче наведені функції розподілу випадкових величин, найбільш часто застосовувані при описанні чинники ризику та показників здоров'я, а також для побудови статистичних загальних правил [3].

1) Нормальний (гаусів) розподіл. Практичне значення цього розподілу полягає в тім, що достатня кількість показників здоров'я на популяційному рівні (в тому числі, антропологічні та інші фізіологічні характеристики, а також показники захворюваності) підкоряються нормальному або приблизно нормальному закону розподілу. Крім того, в теорії імовірностей доказується так звана центральна лімітна теорема, відповідно до якої нормований розподіл суми незалежних випадкових величин, жодна з яких не домінує над іншими, зводиться к нормальному розподілу при збільшенні чисельності слагаємих. Ця теорема дає основу вважати, що розподіл випадкових помилок спостереження є нормальним.

Для нормальних випадкових величин розроблено широкий спектр статистичних методів, тому на практиці приймається завжди нормальний розподіл випадкової величини, коли не має явних підстав відкидати таке положення.

Для нормального розподілу функція щільності має вигляд:

$$f(x) = \frac{1}{s \sqrt{2\pi}} e^{-(x-a)^2 / (2s^2)}. \quad (8)$$

Чим більша величина σ , тим більше відхилення випадкової величини відносно її середнього значення. В точках $x = a \pm \sigma$ функція щільності має перетин, тобто змінює свій характер з випуклого на вогнутий. На практиці часто застосовують наступну

властивість нормального розподілу: незалежно від значень параметрів a і σ площа під графіком щільності розподілу складає:

- 0,68 для інтервалу $a \pm \sigma$;
- 0,95 для інтервалу $a \pm 1,96\sigma$;
- 0,99 для інтервалу $a \pm 2,58\sigma$;
- 0,9974 для інтервалу $a \pm 3\sigma$.

Основою для застосування нормального розподілу до реальних випадкових величин є той факт, що імовірність попадання значень нормальної випадкової величини за межі досить вузького інтервалу $a \pm 3\sigma$ складає менш, ніж 0,3%. Вираз $\xi \sim N(a, \sigma^2)$ використовується як позначення того, що випадкова величина ξ має нормальний розподіл з параметрами a і σ .

Якщо $\xi \sim N(a, \sigma^2)$, то для будь-яких коефіцієнтів A, B випадкова величина, отримана добутком ξ на константу B та додаванням константи A , також є нормальною з параметрами:

$$A+B\xi \sim N(A+Ba, B^2\sigma^2). \quad (9)$$

Тому $(\xi - a)/\sigma \sim N(0,1)$. Такий розподіл називається стандартним нормальним. Щільність стандартного нормального розподілу виражається формулою:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \quad (10)$$

2) Логнормальний розподіл [3]. Безперервна випадкова величина ξ зветься розподіленою за логнормальним законом з параметрами a та σ^2 , якщо випадкова величина $\eta = \ln \xi$ розподілена нормально з параметрами a та σ^2 ($\eta \sim N(a, \sigma^2)$).

Логнормальні випадкові величини можуть приймати тільки позитивні значення. На практиці логнормальний розподіл типовий для таких показників, як концентрація різних забруднювачів в атмосферному повітрі та поверхневих природних водах, інколи, підземних. Тому, логнормальний розподіл застосовують при оцінці ризиків, обумовлених забрудненням атмосферного повітря та природних вод, а в системах соціально-гігієнічного моніторингу частіш за все застосовують не абсолютні величини концентрацій забруднювачів, а патрульні логарифми від них. У деяких випадках такий підхід є ефективним й при моніторингу забруднень ґрунтів та оцінці пов'язаних з ним ризиків.

1) Експоненціальний розподіл – це розподіл, функція щільності якого дорівнює:

$$f(x) = 0 \text{ для всіх } x < 0, \\ f(x) = \lambda e^{-\lambda x} \text{ за умови, що } x \geq 0, \\ \text{де } \lambda - \text{позитивний параметр.}$$

Експоненціальний розподіл добре описує розподіл смертності залежно від віку для дитячого населення. У цьому випадку функція розподілу $F(x)$ визначає ризик смерті у віці x або раніше.

Розподіл Вейбула, Гомперця, Гомперця-Мейкема. Розподіл Вейбула виведений з аналізу тривалості безвідмовної роботи для багатокомпонентних технічних систем, але й багатьох випадках воно ефективно описує залежність ризику смерті від віку дорослої людини.

Функція розподілу та функція щільності Вейбула мають вигляд:

$$F(x) = 1 - e^{-(Ix)^n}, \quad (11)$$

$$f(x) = In e^{-(Ix)^{n-1}}. \quad (12)$$

Більш точний опис розподілу ризику смерті для дорослих дають інші види розподілів, які широко застосовуються в демографії:

- двопараметричний розподіл Гомперця з функцією щільності розподілу чинників:

$$f(x) = Re^{ax} \exp\left\{-\frac{R}{a}(e^{ax} - 1)\right\}; \quad (13)$$

- три параметричний розподіл Гомперця-Мейкема:

$$f(x) = (A + Re^{ax}) \times \\ \times \exp\left\{-Ax - \frac{R}{a}(e^{ax} - 1)\right\}, \quad (14)$$

де A, R, a - позитивні коефіцієнти.

Крім того, для різнопланового типу оцінок чинників екологічного ризику, зокрема, канцерогенного ризику, часто застосовують:

- розподіл χ^2 (хі-квадрат), де кожна з незалежних випадкових величин має стандартний нормально-імовірний розподіл, визначається як сума квадратів $\xi_i - \eta \sim \chi$;

- розподіл Ст'юдента (t-розподіл). Він є найбільш популярним при застосуванні методу вибірок;

- розподіл Фішера (F-розподіл). Цей розподіл базується на методі аналізу ризиків при різних рівнях чинників ризику у випадку, коли чинник ризику є дискретною величиною, а показник здоров'я, наприклад, є безперервною величиною.

Висновки. Таким чином, з всього переліку існуючих математичних підходів до визначення ступеня ризику найбільш придатними є математичні вирази (8), (13) і (14), оскільки вони дозволяють оцінювати одночасно як одиничні, так і комбіновані впливи різноманітних хімічних речовин компонентів навколишнього природного середовища, зокрема, атмосферного повітря та природних поверхневих та підземних водоносних горизонтів, в часі та просторі.

ЛІТЕРАТУРА

1. Качинський А.Б., Хміль Г.А. Екологічна безпека України: аналіз, оцінка та державна політика. – К.: НІСД, 1997, - 127 с.
2. Бредфорд Хилл А. Основы медицинской статистики. – М.: Медгиз, 1988. – 306 с.
3. Большаков А.М., Крутько В.Н., Пуцилло Е.В. Оценка и управление рисками влияния окружающей среды на здоровье населения. – М.: Эдиториал УРСС, 1999. – 254 с.

Стаття надійшла 20.03.2007 р.
Рекомендовано до друку к.т.н., доц.
Бахаревим В.С.