УДК 519.87:316.458.6

Ю.І. Шевяков

Харківський національний університет Повітряних Сил імені Івана Кожедуба, Харків

МЕТОД ПРОГНОЗУВАННЯ КІЛЬКОСТІ ОБСЯГІВ МЕТРОЛОГІЧНОГО ОБСЛУГОВУВАННЯ ВІЙСЬКОВИХ ЗАСОБІВ ВИМІРЮВАЛЬНОЇ ТЕХНІКИ НА ОСНОВІ УДОСКОНАЛЕНИХ БАГАТОФАКТОРНИХ РЕГРЕСІЙНИХ МОДЕЛЕЙ

Запропонований метод прогнозування кількості обсягів метрологічного обслуговування військових засобів вимірювальної техніки зразків ОВТ, що використовує удосконалені багатофакторні регресійні моделі.

Ключові слова: *метрологічне обслуговування, регресійна модель, фактори, фіктивні змінні.*

Вступ

Постановка задачі. Для забезпечення своєчасного метрологічного обслуговування (МО) зразків озброєння та військової техніки (ОВТ) необхідно вдосконалити систему управління силами й засобами метрологічного забезпечення та мати можливість здійснювати прогнозування стану військових засобів вимірювальної техніки (ВЗВТ). Від точності та достовірності прогнозу залежить ефективність реалізації управлінських рішень як в частині оцінки потреби МО зразків ОВТ, так і в частині визначення кількості сил і засобів, які використовуються в метрологічних підрозділах Збройних Сил України, та їх розподілу. Це особливо важливо в сучасних умовах розвитку й реформування Збройних Сил України, які характеризуються появою новітніх зразків озброєння та військової техніки, суттєвим розвитком інформаційних технологій, збільшенням чисельності Збройних Сил України, що вимагає як розширення парку пересувних лабораторій вимірювальної техніки, так і суттєвої якісної заміни обладнання [1 – 3]. Таким чином, наукове обґрунтування методу прогнозування кількості ВЗВТ зразків ОВТ для метрологічного обслуговування шляхом розробки відповідних математичних моделей є актуальним науково-технічним завданням.

Аналіз літератури. В [2, 3] розглянуті питання організації виробничої діяльності військових метрологічних лабораторій в Міністерстві оборони України та Збройних Силах України. Питання прогнозування на основі регресійних моделей розглянуті в [4, 5]. Разом з тим, в цих роботах не враховувалися суттєві чинники, які впливають на якість прогнозу, такі як стан військової техніки, стан системи безпеки держави,військова ситуація тощо.

Метою статті є обґрунтування методу прогнозування кількості ВЗВТ зразків ОВТ на основі удосконалених багатофакторних регресійних моделей.

Основний матеріал

При прогнозуванні метрологічного обслуговування ВЗВТ зразків ОВТ будемо виходити з того, що відома необхідна статистична інформація о кількості замовлень для метрологічного обслуговування ВЗВТ різних типів в залежності від значень факторів, які впливають на цю кількість замовлень. Природно вважати, що у якості цих факторів виступають кількості зразків ОВТ, які оснащені відповідними типами ВЗВТ. В цьому випадку будемо використовувати статистичні методи прогнозування, які засновані на моделях множинного регресійного аналізу [4].

Багатофакторну регресійну модель прогнозування замовлень для метрологічного обслуговування ВЗВТ розглядаємого типу представлено у вигляді

$$Y = f(x_1, x_2, ..., x_m, \varepsilon), \qquad (1)$$

де Y – випадкова залежна змінна, яка описує кількість замовлень на метрологічне обслуговування ВЗВТ певного типу; $x_1, x_2, ..., x_m$ - фактори, або незалежні змінні, що впливають на змінення Y та описують кількості зразків ОВТ кожного виду; ε - випадкова складова, яка описує вплив неврахованих та випадкових факторів на змінення Y; m – кількість видів зразків ОВТ. Розглянемо спочатку лінійну багатофакторну модель регресійну модель

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_m x_m + \varepsilon, \qquad (2)$$

де $\beta_0, \beta_1, ..., \beta_m$ - параметри моделі, причому $\beta_u = 0$, якщо зразок ОВТ u-го виду не містить ВЗВТ типу, що розглядається. Позначимо i-те спостереження змінної Y через y_i, факторів через $x_{i1}, x_{i2}, ..., x_{im}$, випадкової складової - ε_i відповідно, де $i = \overline{1, n}$, а n – кількість спостережень.

Під час проведення аналізу за моделлю (2) будемо вважати, що виконані такі умови [4]:

1. В моделі (2) похибка ε_i (y_i) – випадкова, а фактори $x_{i1}, x_{i2}, ..., x_{im}$ – невипадкові величини.

2. Математичне сподівання похибки ε_i дорівнює нулю: $M[\varepsilon_i] = 0$; $i = \overline{1, n}$.

3. Дисперсія ε_i (y_i) постійна: $D[\varepsilon_i] = \sigma^2$.

4. ε_i та ε_j некорельовані: $M[\varepsilon_i \varepsilon_j] = 0$; $i \neq j$.

5. ε_i (y_i) є нормально розподілена.

Для оцінки параметрів регресійної моделі у випадку, якщо виконуються умови 1-4, будемо використовують метод найменших квадратів. Ці вимоги пов'язані з тим, що оцінки параметрів моделі повинні бути незміщеними, обґрунтованими та ефективними (теорема Гаусса-Маркова) [4]. Оцінкою цієї моделі за вибіркою є рівняння регресії:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + ... + b_m x_m, \qquad (3)$$

де $\hat{y}$ – оцінка математичного сподівання залежної змінної Y; b_i; $i = \overline{0,n}$ - оцінки коефіцієнтів β_i.

Згідно методу найменших квадратів невідомі параметри b_i; $i = \overline{0,n}$ визначаються таким чином, щоб сума квадратів $S(b_0, b_1, ..., b_m)$ відхилень значень y_i від значень $\hat{y}_i$ рівняння регресії (3) була мінімальною:

$$S(b_0,..,b_m) = \sum_{i=1}^{n}(y_i - b_0 - ... - b_m x_{im})^2 \to \min . \qquad (4)$$

У відповідності до (4)

$$\begin{cases} \dfrac{\partial S}{\partial b_0} = -2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_{i1} - ... - b_m x_{im}) = 0; \\ . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \\ \dfrac{\partial S}{\partial b_m} = -2\sum_{i=1}^{n} x_{ip}(y_i - b_0 - b_1 x_{i1} - ... - b_m x_{im}) = 0. \end{cases} \qquad (5)$$

Після розв'язання (4) отримаємо значення параметрів рівняння регресії, в матричній формі:

$$b = \left(X^T X\right)^{-1} X^T y, \qquad (6)$$

де $b = [b_0, b_1, b_2, ..., b_m]^T$ – вектор параметрів рівняння регресії; $y = [y_1, y_2, ..., y_n]^T$ - вектор значень залежної змінної, а матриця значень факторів

$$X = \begin{bmatrix} 1 & x_{11} & ... & x_{1m} \\ 1 & x_{21} & ... & x_{2m} \\ ... & ... & .. & ... \\ 1 & x_{n1} & ... & x_{nm} \end{bmatrix}.$$

З математичної статистики відомо, що будь-яка статистика має бути перевірена на значущість. Іншими словами, за допомогою спеціальних критеріїв необхідно встановити, чи зумовлене значення цієї функції лише похибками вимірювання, чи вона відображає якусь суттєву (значущу) інформацію.

Значущість рівняння регресії (3) у цілому оцінюється за допомогою F-критерія Фішера, а саме: фактичне значення F-статистики

$$F_{факт} = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m} \qquad (7)$$

порівнюють з табличним значенням розподілу Фішера з m і n-m-1 ступенями свободи при заданому рівні значущості α:

$$F_{табл} = F_{\alpha;m;n-m-1}, \qquad (8)$$

де коефіцієнт детермінації

$$R^2 = \sum_{i=1}^{n}\left(\hat{y}_i - \overline{y}\right)^2 \Big/ \sum_{i=1}^{n}(y_i - \overline{y})^2 . \qquad (9)$$

Якщо $F_{факт} > F_{\alpha;m;n-m-1}$, то рівняння регресії (3) признається статистично значущим на рівні значущості α, що свідчить про адекватність моделі. Якщо $F_{факт} < F_{\alpha;p;n-p-1}$, то признається статистична незначущість рівняння регресії (3) на рівні значущості α, тобто модель вважається неадекватною.

Якщо побудована модель (3) адекватна за F-критерієм, то її застосовують для прогнозування залежної змінної. Прогнозне значення $\hat{y}_п$ визначається шляхом підстановки у рівняння регресії (3) відповідних значень факторів $x_{п1}, x_{п2}, ..., x_{пm}$:

$$\hat{y}_п = \hat{y}(x_{п1}, ..., x_{пp}) = b_0 + b_1 x_{п1} + ... + b_p x_{пm}. \qquad (10)$$

Для оцінки якості підбору регресійної моделі до спостережень y_i (характеристики її прогностичної сили) будемо використовувати коефіцієнт детермінації (9), що характеризує, яка частка варіації залежної змінної обумовлена варіацією факторів [4]. Наведемо властивості коефіцієнта детермінації:

1. Коефіцієнт множинної детермінації приймає значення на відрізку [0;1], тобто $0 \le R^2 \le 1$. Чим ближче R^2 до одиниці, тим краще регресія апроксимує емпіричні дані.

2. Якщо $R^2 = 1$, між змінними y та x_1, x_2,…, x_m існує лінійна функціональна залежність.

3. Якщо $R^2 = 0$, то варіація залежної змінної повністю обумовлена впливом випадкових та неврахованих факторів.

На практиці для оцінки ступеню апроксимації спостережень рівнянням регресії використовують наступні емпіричні правила [6]:

1) $R^2 > 0{,}95$ - висока точність апроксимації;

2) $0{,}7 < R^2 < 0{,}95$ - задовільна апроксимація;

3) $R^2 < 0{,}6$ - незадовільна апроксимація.

Недоліком коефіцієнта множинної детермінації R^2 являється те, що він, взагалі, збільшується при додаванні нових факторів, хоча це не обов'язково означає поліпшення якості регресійної моделі. Врахувати дану особливість дозволяє скоригований коефіцієнт детермінації $\hat{R}^2$, який визначається зяк [4]:

$$\hat{R}^2 = 1 - \frac{n - 1}{n - m - 1}\left(1 - R^2\right). \qquad (11)$$

На відміну від R^2 скоригований $\hat{R}^2$ може зменшуватись при введенні у модель нових факторів, які не чинять істотного впливу на залежну змінну. Для оцінки якості рівняння регресії використовують також й інші показники ефективності: середню абсолютну похибку у відсотках (MAPE), середню відсоткову похибку (MPE). Середня абсолютна похибка у відсотках розраховується за формулою

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \Big/ y_i \cdot 100\% \qquad (12)$$

та характеризує наскільки великі похибки прогнозу у порівнянні з фактичними значеннями залежної змінної. Допустима межа значень MAPE – не більше 8-10%. Середня відсоткова похибка дорівнює

$$MPE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right) \Big/ y_i \cdot 100\% \qquad (13)$$

та визначає зміщеність (постійне зменшення або збільшення) побудованого прогнозу. Значення, які обчислюються з співвідношення (12), інтерпретуються таким чином. Якщо модель прогнозування є незміщеною, то показник $|MPE|$ буде мати значення, яке близьке до нуля, і не повинен перебільшувати 5%. Якщо у результаті розрахунків отримується велике позитивне (негативне) відсоткове значення, то побудована модель послідовно переоцінює (недооцінює) результат прогнозу.

Розглянута регресійна модель розв'язання задачі прогнозування потреби в метрологічному обслуговуванні ВЗВТ зразків ОВТ потребує корегування, пов'язаного з умовою проведення обслуговування: застосування в умовах мирного часу, проведення АТО або воєнного часу. Запропонуємо наступний метод корегування результатів аналізу, який дозволяє оцінювати вплив значень кількісних якісних факторів (мирний час, АТО, воєнний час) на залежну змінну за допомогою так званих фіктивних змінних [4]. У якості фіктивних змінних, будемо використовувати дихотомічні змінні, які приймають лише два значення: 0 або 1. У нашому випадку введемо у модель (2) дві фіктивні булеві змінні z_1 та z_2:

$$Y=\beta_0+\beta_1 x_1+\beta_2 x_2+...+\beta_m x_m+\alpha_1 z_1+\alpha_2 z_2+\varepsilon, \qquad (14)$$

де α_1, α_2 – коефіцієнти при фіктивних змінних.

Параметри запропонованої моделі будемо оцінювати за допомогою методу найменших квадратів:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + ... + b_m x_m + a_1 z_1 + a_2 z_2, \qquad (15)$$

де a_1, a_2 - оцінки параметрів α_1, α_2. У разі, якщо скоригований коефіцієнт детермінації для рівняння регресії (15) $\hat{R}^2_{xz}$ буде перебільшувати $\hat{R}^2_{x}$ – скоригований коефіцієнт детермінації для рівняння регресії (3), отримаємо підтвердження щодо доцільності застосування саме рівняння регресії (15) для прогнозування потреби ВЗВТ. Окрім лінійної моделі (2) можливо використання й нелінійних багатофакторних регресійних моделей, а саме квадратичної

$$\begin{aligned} Y=&\beta_0+\beta_1 x_1+\beta_2 x_2+...+\beta_m x_m+\beta_{11}x_1^2+\beta_{22}x_2^2+...+\\ &+\beta_{mm}x_m^2+\beta_{12}x_1 x_2+\beta_{13}x_1 x_3+...+\beta_{m-1,m}x_{m-1}x_m+\varepsilon, \end{aligned} \qquad (16)$$

степеневої

$$Y = \beta_0 x_1^{\beta_1} x_2^{\beta_2} \cdot ... \cdot x_m^{\beta_m} + \varepsilon, \qquad (17)$$

експоненціальної

$$Y = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_m x_m) + \varepsilon \qquad (18)$$

та інших моделей. Причому, обирається та з них, для якої скоригований коефіцієнт детермінації - найбільший, а середня абсолютна похибка у відсотках та середня відсоткова похибка - мінімальні.

Для оцінки параметрів β_i в рівняннях нелінійних моделей можна використовувати два підходи [4].

Перший підхід заснований на лінеаризації моделі й полягає в тому, що за допомогою належних перетворень вхідних змінних залежність, що досліджується, представляється у вигляді лінійного співвідношення між отриманими змінними. Для лінеаризації можуть використовуватися як моделі, які є нелінійними за факторами, але є лінійними за параметрами, так і моделі, які є нелінійними за параметрами. Так квадратична модель (16) є нелінійною за факторами, але лінійна за параметрами. Її можна за допомогою нових змінних

$$\xi_{11} = x_1^2;\ \xi_{22} = x_2^2;\ ...;\xi_{mm} = x_m^2;\ \xi_{12} = x_1 x_2;$$
$$\xi_{13} = x_1 x_3;...;\xi_{m-1,m} = x_{m-1}x_m$$

звести до такої лінійної моделі:

$$\begin{aligned} Y =& \beta_0 + \beta_1 x_1 + ... + \beta_m x_m + \beta_{11}\xi_{11} + ... +\\ &+\beta_{mm}\xi_{mm} + \beta_{12}\xi_{12} + ... + \beta_{m-1,m}\xi_{m-1,m} + \varepsilon, \end{aligned} \qquad (19)$$

параметри якої знаходяться звичайним методом найменших квадратів за формулою (6).

Проте, слід відмітити й недолік такої заміни змінних, пов'язаний з тим, що вектор оцінок параметрів (6) для регресійної моделі (16) отримується не з умови мінімізації суми квадратів відхилень для вхідних змінних-факторів, а з умови мінімізації суми квадратів відхилень для перетворених змінних для моделі (19), що не одне й теж.

Більш складною задачею є оцінювання параметрів у випадку нелінійної моделі за параметрами. Так моделі (17) та (18) неможливо звести до лінійних моделей. Але, якщо розглянуті моделі представити у мультиплікативній формі, а саме:

$$Y = \beta_0 x_1^{\beta_1} x_2^{\beta_2} \cdot ... \cdot x_m^{\beta_m} \cdot \varepsilon \qquad (20)$$

та $\quad Y = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_m x_m) \cdot \varepsilon, \quad (21)$

то вони зводяться до лінійних [4, 5] логарифмуванням обох частин рівнянь (20, 21)

$$\ell n Y=\beta_0+\beta_1 \ell n(x_1)+\beta_2 \ell n(x_2)+...+\beta_m \ell n(x_m)+\ell n(\varepsilon);$$

$$Z = B_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_m X_m + E, \qquad (22)$$

де $Z = \ell nY$; $B_0 = \ell n\beta_0$; $X_1 = \ell n(x_1)$; $X_2 = \ell n(x_2)$;…; $X_m = \ell n(x_m)$; E= $\ell n(\varepsilon)$, $\ell nY = \beta_0 + .. + \beta_m x_m + \ell n(\varepsilon)$;

$$Z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_m x_m + E, \qquad (23)$$

де Z = ℓnY ; E= $\ell n(\varepsilon)$.

Отже, для оцінки параметрів регресійних моделей (16) – (18) пропонується другий підхід, який полягає в застосуванні методів оптимізації, на основі вхідних змінних - факторів та залежної змінної, наприклад, модифікованого метода Ньютона, метода Маркварда, метода можливих напрямків [7]. Застосування відповідного програмного забезпечення ПЕВМ, таких як пакети прикладних програм Mathcard, Statgraphics, Statistica, Excel знімає всі обчислювальні проблеми, які раніше були перепоною для розв'язання нелінійних багатофакторних регресійних моделей.

Для нелінійної залежності скоригований коефіцієнт множинної детермінації визначається за тією ж формулою, що й для лінійної залежності

$$\hat{R}^2 = 1 - \frac{n-1}{n-m-1}\left(1 - R^2\right), \qquad (24)$$

де m – кількість параметрів при змінних x. Відмінність формул (11) та (24) складається в тому, що для лінійної залежності під m розуміється кількість факторів, проте для нелінійної залежності під m розуміється кількість параметрів при змінних x та їх перетвореннях ($x^2, \sqrt{x}, \ell nx$). Як і для лінійної регресії, коефіцієнт детермінації використовується й для оцінки значущості рівняння нелінійної регресії за F-критерієм Фішера.

Співвідношення для F-статистики у цьому випадку має такий вигляд:

$$F_{\text{факт}} = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m}, \qquad (25)$$

де m – кількість параметрів при змінних x. Відмінність формул (7) та (25) є аналогічною вказаної відмінності між (10) та (23).

Введемо в регресійні моделі (16) – (18) фіктивні змінні, що враховують якісний фактор:

$$Y = \beta_0 + \beta_1 x_1 + ... + \beta_m x_m + \beta_{11} x_1^2 + ... + \beta_{mm} x_m^2 + \\ + \beta_{12} x_1 x_2 + ... + \beta_{m-1,m} x_{m-1} x_m + \alpha_1 \varsigma_1 + \alpha_2 \varsigma_2 + \varepsilon; \qquad (26)$$

$$Y = \beta_0 x_1^{\beta_1} x_2^{\beta_2} \cdot ... \cdot x_m^{\beta_m} z \varphi_1^{\alpha_1} \varphi_2^{\alpha_2} + \varepsilon; \qquad (27)$$

$$Y = \exp(\beta_0 + ... + \beta_m x_m + \alpha_1 \psi_1 + \alpha_2 \psi_2) + \varepsilon, \qquad (28)$$

Висновки

1. В статті запропонований метод прогнозування кількості військових засобів вимірювальної техніки зразків ОВТ, заснований на використанні багаттофакторних регресійних моделей.

2. Удосконалені багатофакторні регресійні моделі прогнозування кількості ВЗВТ для метрологічного обслуговування, в яких крім кількісних факторів, що враховують кількість зразків ОВТ кожного виду, додатково введений якісний фактор, що враховує можливий стан військової ситуації (мирний час, АТО, воєнний час). Вплив якісного фактору враховується за рахунок використання фіктивних змінних.

3. Для отримання незміщених оцінок параметрів нелінійних регресійних моделей запропоновано на відміну від застосування методів лінеаризації моделей використовувати методи оптимізації із застосуванням програмного забезпечення ПЕВМ.

Список літератури

1. Кузнецов І.Б. Організація метрологічного забезпечення військ (сил). Ч. 1 / І.Б. Кузнецов, П.М. Яблонський. – К. : НУОУ, 2009. – 356 с.

2. Наказ заступника Міністра оборони з озброєння – начальника Озброєння ЗС України "Про затвердження Керівництва з організації та порядку експлуатації вимірювальної техніки у ЗС України" від 1.06.2001 № 79.

3. Наказ начальника Центрального управління метрології і стандартизації "Про затвердження Керівництва з організації виробничої діяльності військових метрологічних лабораторій в Міністерстві оборони України та Збройних Силах України" від 14.05.2007 № 2.

4. Кремер Н.Ш. Эконометрика / Н.Ш. Кремер, Б.А. Путко. – М.: ЮНИТИ-ДАНА, 2003. – 311 с.

5. Четыркин Е.М.. Статистические методы прогнозирования / Е.М. Четыркин – М.: Статистика, 1975. – 184 с.

6. Гельман В.Я. Решение математических задач средствами Excel / В.Я. Гельман. - М.;Питер, 2008. – 235 с.

7. Химмельблау Д. Прикладное нелинейное программироание / Д. Химмельблау. – М.: Мир, 1975. – 534 с.

Надійшла до редколегії 8.08.2016

Рецензент: д-р техн. наук, проф. В.Б. Кононов, Харківський національний університет Повітряних Сил імені І. Кожедуба, Харків.

МЕТОД ПРОГНОЗИРОВАНИЯ КОЛИЧЕСТВА ОБЪЕМОВ МЕТРОЛОГИЧЕСКОГО ОБСЛУЖИВАНИЯ ВОЕННЫХ СРЕДСТВ ИЗМЕРЕНИЙ НА ОСНОВЕ УСОВЕРШЕНСТВОВАННОЙ МНОГОФАКТОРНОЙ РЕГРЕССИОННОЙ МОДЕЛИ

Ю.И. Шевяков

В статье предложен метод прогнозирования количества объемов метрологического обслуживания военных средств измерительной техники образцов ВВТ, основанный на использовании усовершенствованных многофакторных регрессионных моделей.

Ключевые слова: *метрологическое обслуживание, регрессионная модель, факторы, фиктивные переменные.*

METHOD OF FORECASTING THE NUMBER OF VOLUMES METROLOGICAL SERVICE OF MEASURING INSTRUMENTS MILITARY BASED IMPROVED MULTIVARIATE REGRESSION MODELS

Yu.I. Sheviakov

In the article the method of forecasting the volume metrology military service of measuring instruments weapons and equipment based on the use of advanced multivariate regression models.

Keywords: *metrological service regression model, factors dummies.*