ZESZYTY NAUKOWE POLITECHNIKI ŚLĄSKIEJ 2000
Seria: GÓRNICTWO z. 246 Nr kol. 1480

Stanisław KOWALIK
Politechnika Śląska, Gliwice

SCHEMATY LOSOWANIA W BADANIACH STATYSTYCZNYCH STOSOWANYCH W OBSZARACH PROBLEMOWYCH BEZPIECZEŃSTWA PRACY

Streszczenie. W pracy przedstawiono różne schematy pobierania próby z populacji generalnej. Te schematy stosuje się w badaniach statystycznych. Przedstawiono też istotę metody reprezentacyjnej, która zajmuje się tymi zagadnieniami. Zwrócono uwagę na możliwość wykorzystania metody reprezentacyjnej w badaniach bezpieczeństwa. Omówiono następujące schematy losowania: losowanie nieograniczone indywidualne ze zwracaniem, losowanie nieograniczone indywidualne bez zwracania oraz losowania ograniczone warstwowe, systematyczne, zespołowe jednostopniowe i wielostopniowe.

SCHEMES OF SAMPLING IN STATISTICAL RESEARCHES IN AREAS OF SAFETY

Summary. In this work different schemes of sampling from general population are represented. These schemes are used in statistical researches. The nature of representative method which deals which these problems is also presented. It turns attention to possibility of utilisation of representative method in investigations of safety. The following schemes of sampling are discussed: unrestricted individual sampling without return, unrestricted individual sampling without return and, limited stratified sampling, systematical, joint single-stage and multistage.

1. Wstęp

Na wstępie zacytujemy M. Balińskiego, który już w 1834 r. napisał w książce „Statystyczne obrachowanie miasta Wilna" (Znicz, Wilno, 1834) następujące słowa: „We wszystkich podaniach statystycznych pierwszą miarą ich wartości jest prawda i dokładność,

a drugą, żeby były jak najnowsze" [8]. W badaniach statystycznych nie rozpatruje się całej zbiorowości generalnej, ale bierze się pewną liczbę elementów ze zbiorowości generalnej. Te elementy nazywa się próbą, która jest podstawą do uogólnień i wyciągania wniosków dotyczących całej populacji ze względu na rozpatrywaną cechę.

Przedmiotem tej pracy jest wybór odpowiedniej próby, która by reprezentowała całą zbiorowość generalną możliwie jak najdokładniej. Elementy do próby wybiera się w sposób losowy. W zależności od posiadanych informacji o całej populacji stosuje się odpowiednie schematy losowania. Schematem losowania nazywa się konkretny i spójny z punktu widzenia formalnego oraz merytorycznego sposób postępowania przy doborze próby z danej zbiorowości generalnej, który zapewnia spełnienie warunku zbiorowości [8].

Przy rozwiązywaniu problemu trzeba dokonać abstrakcyjnego modelu rzeczywistości, czyli zdefiniować zbiór danych mających reprezentować rzeczywistą sytuację. Wybór ten powinien być podporządkowany problemowi, który ma być rozwiązany. Potem następuje wybór tych informacji. Wybór reprezentacji danych jest często dość trudny i problem reprezentacji trzeba analizować na wielu poziomach szczegółowości [10].

Wyborem odpowiedniego schematu losowania zajmuje się metoda reprezentacyjna.

Praca niniejsza ma na celu zachęcić także do stosowania metody reprezentacyjnej do badań związanych z bezpieczeństwem. W tych badaniach prowadzi się różne statystyki i wykazy wypadkowości w różnych przekrojach, np. według miejsca zaistnienia (na kopalni, na oddziale, na powierzchni, na ścianie, na przodku), według przyczyn wypadku [7], ze względu na wiek, ze względu na stanowisko, czas itp. Analizuje się także warunki konieczne i wystarczające następstw zdarzeń niebezpiecznych [1]. Wykorzystując logikę i elementy teorii zdarzeń tworzy się łańcuchy warunków koniecznych i identyfikuje się zagrożenia wypadkowe [1]. Istnieje specjalna klasyfikacja wydarzeń powodujących wypadki [7]. W pracy [3] rozważa się teoretyczne podstawy określania miar stanu zagrożenia bezpieczeństwa w wyrobiskach górniczych. Tworzy się pewne modele informacyjno-prognostyczne dotyczące zagrożeń na kopalni. Rozumie się je jako pewną zweryfikowaną wiedzę teoretyczną odniesioną do konkretnego obiektu podziemnego [3]. W celu stworzenia modelu należy dysponować danymi w miarę reprezentatywnymi, aby utworzony teoretyczny model dobrze odzwierciedlał rzeczywistość. Z tego względu wydaje się celowe zastosowanie metody reprezentacyjnej także w badaniach bezpieczeństwa. Metoda reprezentacyjna daje naukowe podstawy do właściwego

doboru elementów do próby, na podstawie której tworzy się potem modele rzeczywistości i analizuje się je.

2. Istota metody reprezentacyjnej

Najpierw przedstawimy pokrótce istotę metody reprezentacyjnej. Metoda reprezentacyjna jest działem statystyki matematycznej, która zajmuje się schematami losowań próby z populacji generalnej. W metodzie reprezentacyjnej rozważa się konsekwencje użycia określonego schematu losowania próby na wartości stosowanych estymatorów. Przy wyborze schematu losowania próby kierujemy się przede wszystkim efektywnością estymatorów, a także kosztami badania i praktycznymi możliwościami przeprowadzenia losowania [2, 8, 9].

Metoda reprezentacyjna polega na losowym doborze próby ze skończonej populacji generalnej, opisie tej próby za pomocą charakterystyk statycznych, a następnie na uogólnieniu otrzymanych wyników na populację generalną. Populacja generalna musi być jednorodna, aby występujące w niej elementy podlegały działaniu tych samych przyczyn systematycznych. Zróżnicowanie elementów ma wynikać jedynie z przyczyn przypadkowych [2, 8, 9].

Zbiorowość próbna albo próbka lub próba jest to każdy niepusty podzbiór populacji generalnej. Na podstawie próbki formułuje się wnioski dotyczące postaci rozkładów prawdopodobieństwa lub parametrów zmiennych losowych. Od próby wymaga się, by była reprezentatywna, tj. aby z określoną dokładnością opisywała strukturę populacji generalnej. Reprezentatywność próby zależy od sposobu doboru próby i od jej liczebności. Po pobraniu próby, elementy próby są mierzone ze względu na badane cechy. Na podstawie tych pomiarów tworzy się tzw. statystyki. Są to pewne charakterystyki w próbie, takie jak wartość średnia, odchylenie standardowe, frakcja [2, 8, 9].

Następnie na podstawie statystyk tworzy się estymatory i ocenia odpowiednie nie znane parametry populacji generalnej. Ta ocena nazywa się szacowaniem albo estymacją parametrów.

Autor pracy [6] rozważa reprezentatywność próby w trzech aspektach:

1. Typologiczna reprezentatywność próby. Zmienne uwzględnione w typologicznej próbie bywają często zmiennymi wielowymiarowymi. W tym przypadku pobierana próba jest rezultatem kombinatorycznego skrzyżowania tych zmiennych, które chcemy, by znalazły

się w próbie. Abstrahujemy także od częstości występowania poszczególnych kombinacji w warunkach naturalnych.

2. Reprezentatywność ze względu na rozkłady określonych zmiennych. W tym przypadku proporcje w próbie powinny być zbliżone do rozkładów zmiennych.
3. Reprezentatywność pod względem statystycznych zależności. W tym przypadku znak i natężenie zależności w próbie są prawie takie same, jak znak i natężenie tej samej zależności w zbiorowości generalnej.

Zakres zastosowań metody reprezentacyjnej jest duży. Obejmuje on takie dziedziny, jak:

1. Badana rynku i sprzedaży, np.:
   - identyfikacja rynku,
   - badania postępowania konsumenta,
   - badania potencjału rynkowego,
   - badania udziału rynkowego,
   - analiza sprzedaży,
   - ustalanie kwot i obszarów sprzedaży,
   - testy rynkowe.
2. Badania przedsiębiorstwa i działalności gospodarczej, np.:
   - analiza trendów w dziedzinie gospodarowania,
   - analiza planów strategicznych,
   - analiza działalności wewnętrznej firmy.
3. Badania produktu, np.:
   - badania rozwoju nowego produktu,
   - badania opakowania,
   - badania konkurencyjności produktu,
   - testowanie istniejących produktów.
4. Badania reklamy, np.:
   - analiza przekazów reklamowych,
   - analiza mediów reklamy,
   - analiza wydatków na reklamę,
   - analiza efektywności reklamy,
   - badania motywacyjne,
   - badania działalności reklamowej konkurentów.

5. Badania cen, np.:
   - badania struktury cen w linii produktu,
   - analiza punktu krytycznego.
6. Badania dystrybucji, np.
   - badania kanałów dystrybucji,
   - badania lokalizacji magazynów.
7. Badania otoczenia działalności przedsiębiorstwa, np.:
   - analiza prawnych aspektów działań promocyjnych,
   - analiza aspektów ekologicznych w działalności firmy,
   - badania wartości społecznych w działalności gospodarczej [8].

Metoda reprezentacyjna jest wykorzystywana w tzw. badaniu opinii publicznej przez wyspecjalizowane przedsiębiorstwa, takie jak: Instytut Gallupa w Stanach Zjednoczonych, Instytut Doxa lub Censas we Włoszech, Instytut Badania Opinii Publicznej w Allensbach w Niemczech oraz mający ogólnoświatowy zasięg Instytut Louisa Harrisa [8].

3. Schematy losowania w badaniach statystycznych

Schematy losowania można ogólnie podzielić na dwie grupy:

1) losowanie nieograniczone,
2) losowanie ograniczone.

Do pierwszej grupy zaliczamy losowanie indywidualne ze zwracaniem (jest to tzw. próba prosta) oraz losowanie indywidualne bez zwracania.

Do drugiej grupy zaliczamy losowanie warstwowe, systematyczne, zespołowe jednostopniowe i wielostopniowe.

3.1. Losowanie nieograniczone

3.1.1. Losowanie nieograniczone ze zwracaniem

Najpierw objaśnimy, co to jest losowanie nieograniczone. Polega ono na losowaniu od razu całej próby z całej populacji. Nie zawsze jest ono najwygodniejsze w praktyce i najbardziej efektywne z punktu widzenie statystycznego.

Losowanie ze zwracaniem polega na tym, że po każdym ciągnieniu wylosowana jednostka wraca z powrotem do zbiorowości generalnej.

Przyjmujemy następujące oznaczenia:

N – liczebność zbiorowości generalnej,

n – liczebność próby losowej,

a_i – element zbiorowości generalnej (i=1,...,N).

Jednoczesny wybór n-elementowej próby z całej N-elementowej zbiorowości generalnej oznacza wybór jednej z możliwych wariacji z powtórzeniami zbioru N-elementowego. Liczba tych wariacji wynosi

$$W_n^N = N^n \quad (n<N). \tag{1}$$

Każda z możliwych wariacji ma to samo prawdopodobieństwo realizacji. W ten sposób wylosowaną próbę nazywa się próbą prostą. Prawdopodobieństwo dostania się do próby każdego elementu zbiorowości generalnej wyraża się w tym przypadku wzorem

$$P(a_i) = 1 - \left(\frac{N-1}{N}\right)^n = 1 - \left(1 - \frac{1}{N}\right)^n. \qquad (i=1,\ldots,N;\ 1<n<N) \tag{2}$$

W wyniku takiego losowania te same jednostki mogą się dostać do próby kilkakrotnie. Z tego względu liczbę różnych pod względem cechy jednostek nazywa się efektywną liczebnością próby (number of distinc elements).

3.1.2. Losowanie nieograniczone bez zwracania

Losowanie bez zwracania polega na tym, że po każdym ciągnieniu wylosowana jednostka nie wraca do zbiorowości generalnej. W tego rodzaju losowaniu każda z możliwych kombinacji bez powtórzeń elementów

$$K_n^N = \binom{N}{n} = \frac{N!}{n!(N-n)!} \quad (n<N). \tag{3}$$

Prawdopodobieństwo dostania się do próby, w tym przypadku, w pierwszym ciągnieniu do każdego elementu zbiorowości generalnej wynosi

$$P(a_i) = \frac{n}{N} \quad (i=1,\ldots,N). \tag{4}$$

3.2. Losowanie ograniczone

3.2.1. Losowanie warstwowe

Losowanie warstwowe polega na tym, że przed losowaniem dzielimy populację generalną na jednorodne z punktu widzenia badanej cechy części, inaczej zwane warstwami. Z każdej warstwy losuje się oddzielnie elementy do próby. Podział populacji na warstwy powinien być dokonany w ten sposób, że warstwy pomiędzy sobą mocno się różnią, a wewnątrz są jednorodne ze względu na badaną cechę. Wtedy losowanie warstwowe jest efektywniejsze od losowania nieograniczonego, tj. daje mniejsze błędy przeciętne szacowanych parametrów.

Wprowadzamy dodatkowe oznaczenia:

m – liczba warstw w zbiorowości generalnej i w próbie,

N_j – liczebność j-tej warstwy,

n_j – liczebność wylosowanej próby w j-tej warstwie.

Mamy więc

$$N = \sum_{j=1}^{m} N_j, \tag{5}$$

$$n = \sum_{j=1}^{m} n_j. \tag{6}$$

Wyróżnia się cztery rodzaje losowania warstwowego: losowanie równomierne (equal allocation), losowanie proporcjonalne (proportional allocation), alokacja przy minimum wariancji, inaczej zwana schematem Neymana (Neyman allocation), oraz losowanie optymalne, tzw. schemat optymalny (optimum allocation).

Pierwszy sposób to losowanie równomierne. Polega ono na tym, że próbę n-elementową rozdziela się równo dla każdej warstwy. Liczebność próbki w każdej warstwie jest taka sama.

$$n_j = \frac{n}{m} \quad (j=1,\ldots,m). \tag{7}$$

W losowaniu proporcjonalnym w każdej warstwie losuje się do próby liczbę elementów proporcjonalną do liczebności warstwy. W tym przypadku rozkład elementów w warstwach jest taki sam, jak w całej populacji

$$\frac{n_j}{N_j} = \frac{n}{N} = \text{constans}. \tag{8}$$

Z tego otrzymujemy, że liczebność próbki w każdej warstwie wynosi

$$n_j = \frac{N_j}{N} n \quad (j=1,\ldots,m). \tag{9}$$

W alokacji zwanej schematem Neymana liczebność próbki w warstwie określa się na podstawie wielkości warstwy i jej wariancji.

$$n_j = \frac{N_j s_j}{\sum_{k=1}^{m} N_k s_k} n \qquad (j=1,\ldots,m), \tag{10}$$

gdzie s_j^2 oznacza wariancję w j-tej warstwie.

Jeżeli w schemacie Neymana uwzględnia się też zróżnicowany koszt c_j dla każdej warstwy, to takie losowanie nazywa się optymalnym lub schematem optymalnym.

$$n_j = \frac{\frac{N_j s_j}{\sqrt{c_j}}}{\sum_{k=1}^{m} \frac{N_k s_k}{\sqrt{c_k}}} n. \quad (j=1,\ldots,m), \tag{11}$$

3.2.2. Losowanie systematyczne

Losowanie systematyczne polega na tym, że do próby wybiera się jednostki nie w sposób losowy, lecz co którąś, np. co dziesiątą lub co setną. Jedynie pierwszy element wybiera się losowo. Tak zwany odstęp losowania (sampling interval) określa się następująco

$$k = \frac{N}{n}. \tag{12}$$

Wzór (12) określający interwał k jest tylko przybliżeniem, ponieważ k ex definitione musi być liczbą całkowitą.

3.2.3. Losowanie zespołowe jednostopniowe

Losowanie zespołowe (cluster sampling) polega na tym, że najpierw łączy się jednostki w określone grupy czy zespoły tworzące pewną odrębną całość, a następnie losuje się te zespoły. Jest to losowanie zespołowe jednostopniowe.

3.2.4. Losowanie zespołowe wielostopniowe

Losowanie zespołowe wielostopniowe stosuje się do bardzo dużych i złożonych zbiorowości generalnych. Polega ono na tym, że najpierw łączy się jednostki w określone

zespoły. Te zespoły dzieli się na mniejsze podzespoły, a te z kolei na zespoły coraz to niższego stopnia. Najpierw losuje się duże zespoły pierwszego stopnia, następnie losuje się zespoły coraz to niższego stopnia. Próbę stanowią elementy wylosowane z zespołów najniższego stopnia.

Na zakończenie przytoczymy jeszcze przykłady zastosowania metody reprezentacyjnej do oceny bezpieczeństwa pracy w kopalniach węgla kamiennego „Rydułtowy" i „Chwałowice". Ta metoda była wykorzystana w programie MERIT-ROB, który zastosowano w tych kopalniach. Procedury programu MERIT – Management Evaluation Regarding Itemized Tendentious (Ocena zarządzania pracy przez określanie kierunków zmian) należą do zbioru technik organizatorskich opracowanych w USA w celu analizy niebezpiecznych warunków pracy i ograniczenia niebezpiecznych zdarzeń [4]. Procedura MERIT-ROB (Rankingowa ocena bezpieczeństwa pracy) jest programem oceniającym stan bezpieczeństwa, w którego realizacji ma szansę uczestniczyć każdy pracownik kopalni. Jednym z etapów tej procedury jest określenie reprezentatywnej grupy pracowników (próby z populacji generalnej). Na podstawie numeru ewidencyjnego wszystkim pracownikom przypisano kolejne liczby naturalne. Losowano próbę za pomocą komputera o liczebności w granicach 6% do 10% populacji generalnej [4, 5]. Wielkość prób i liczebność populacji generalnych przedstawia tablica 1.

Tablica 1

Losowany wybór próby w kopalniach

Kopalnia	Stan ewidencyjny	Liczba respondentów	Udział respondentów
KWK „Chwałowice"	4397	368	8.37%
KWK „Rydułtowy"	3440	297	8.63%

Źródło: literatura [5]

4. Zakończenie

W pracy przedstawiono istotę metody reprezentacyjnej oraz opisano różne sposoby losowania próby w badaniach statystycznych. Metoda reprezentacyjna daje naukowe podstawy do wyboru odpowiedniego schematu losowania. Metoda ta znalazła zastosowanie w różnych dziedzinach wymienionych w rozdziale 2. W geologii można zastosować losowanie

warstwowe np. do pobieranie próbek z różnych złóż. Praca ta ma też na celu zwrócenie uwagi na możliwość wykorzystania metody reprezentacyjnej do badań związanych z bezpieczeństwem. W pracy podano przykłady wykorzystania tej metody do oceny bezpieczeństwa pracy w kopalniach węgla kamiennego „Rydułtowy" i „Chwałowice". Podniesienie poziomu bezpieczeństwa jest ważnym zagadnieniem, szczególnie w przemyśle górniczym.

LITERATURA

1. Cichowski E.: Identyfikacja zagrożenia w górnictwie węgla kamiennego. Wydawnictwo Politechniki Śląskiej, Gliwice 1999.
2. Greń J.: Statystyka matematyczna. Modele i zadania. PWN, Warszawa 1982.
3. Krzemień S.: Teoretyczne podstawy określania miar stanu zagrożenia bezpieczeństwa w wyrobiskach górniczych. Zeszyty Naukowe Politechniki Śląskiej, ser. Górnictwo z 204, Gliwice 1992.
4. Krzemień S.: Program MERIT – rankingowa ocena bezpieczeństwa pracy w kopalni. Seminarium międzynarodowe: Doświadczenia polskie i amerykańskie w zarządzaniu bezpieczeństwem pracy w górnictwie. Rudy Raciborskie, 18-18. 10. 1996.
5. Krzemień S., Kucza J.: Zastosowanie programu MERIT – rankingowa procedura oceny bezpieczeństwa pracy w kopalniach węgla kamiennego. Seminarium międzynarodowe: Doświadczenia polskie i amerykańskie w zarządzaniu bezpieczeństwem pracy w górnictwie. Rudy Raciborskie, 18-18. 10. 1996.
6. Nowak S.: Metodologia badań społecznych. PWN, Warszawa 1985.
7. Parchański J., Chudek M.: Identyfikacja zagrożeń geotechnicznych na podstawie statystyki wypadkowej. Międzynarodowa Konferencja „IV Szkoła Geomechaniki", Gliwice-Ustroń, 18-22. 10. 1999.
8. Steczkowski J.: Metoda reprezentacyjna w badaniach zjawisk ekonomiczno-społecznych. PWN, Warszawa 1995.
9. Steczkowski J.: Reprezentacyjne badania jakości wyrobów. Drukarnia Wydawnicza PLATAN, Kraków 1993.

10. Wirth N.: Algorytmy + struktury danych = programy. WNT, Warszawa 2000.

Recenzent: Dr inż. Ryszard Winiarczyk

Abstract

In this work different schemes of sampling from general population are represented. These schemes are used in statistical researches. The nature of representative method which deals which these problems is also presented. It turns attention to possibility of utilisation of representative method in investigations of safety. The following schemes of sampling are discussed: unrestricted individual sampling without return, unrestricted individual sampling without return and, limited stratified sampling, systematic, joint single-stage and multistage. The paper describes also examples of practical application of these methods by the assessment of safety in two selected coal mines