ZESZYTY PROBLEMOWE
POSTĘPÓW NAUK ROLNICZYCH
Zeszyt 76 1967

WARIANCJA PARAMETRÓW TORFOWISKA I JEJ WPŁYW NA USTALENIE LICZBY PUNKTÓW SONDACYJNYCH W BADANIACH KATEGORII C_1 i C_2

ВАРИАНЦИЯ ПАРАМЕТРОВ ТОРФЯНЫХ ЗАЛЕЖЕЙ И ЕЁ ВЛИЯНИЕ НА ОПРЕДЕЛЕНИЕ ЧИСЛА ЗОНДИРОВАННЫХ ПУНКТОВ В ГОРИЗОНТАХ C_1 И C_2

VARIATION OF PEAT BOGS PARAMETERS AND THEIR EFFECT ON THE ESTABLISHMENT OF THE SONDS NUMBER BY THE INVESTIGATION IN C_1 AND C_2 CATEGORY

RUDOLF HOHENBERG

1. WSTĘP

W badaniach polowych kategorii C_1 i C_2 najczęściej stosowana jest metoda punktów losowych [1] zwana przez torfoznawców metodą punktów rozproszonych. Zakłada się, że rozpoznawany obszar został wylosowany z rodziny torfowisk, a badany parametr na przykład miąższość złoża w danym punkcie pola jest zmienną losową. Rodzinę zmiennych losowych przyporządkowanych punktom płaszczyzny nazywamy polem stochastycznym [2]. Stąd modelem matematycznym torfowiska jest pole stochastyczne. Jest ono stacjonarne i izotropowe jeśli zmienne losowe mają wspólną średnią i wspólną zmienność (wariację) oraz współczynnik korelacji zmiennych losowych jest funkcją odległości pomiędzy przyjętymi punktami pobierania prób [3]. Za estymator badanego parametru przyjmuje się średnią poczynionych obserwacji w różnych punktach obszaru złoża w celu ustalenia najkorzystniejszego rozmieszczenia sondowań przy żądanej dokładności w szacowaniu badanego parametru.

Wiadomo, że niezbędna liczebność próby losowej, odpowiadająca wymaganej dokładności jest wprost proporcjonalna do wariancji badanego parametru, a odwrotnie proporcjonalna do kwadratu maksymalnego odchylenia średniej próby od średniej populacji generalnej. Trudność polega na tym, że wariancja badanego parametru torfowiska nie jest znana, gdyż badacz dysponuje jedynie wielkością pola obszaru i zarysem jego linii brzegowej. Stąd na ogół liczba przyjętych punktów sondacyjnych na torfowisku jest zaniżona lub zbyt duża. Na przykład przy badaniu torfowisk zalegających dolinę rzeki Biebrzy dokonano na dwóch obszarach o wspólnej powierzchni 700 ha, 1875 wierceń głębokościowych w celu ustalenia miąższości i zasobów złóż torfowych tego obszaru. Przyjmując zaś za model matematyczny torfowiska pole stochastyczne, okazuje się, że ilość przyjętych punktów na wyczucie dla jednego torfowiska była zbyt liczna, dla drugiego natomiast zaniżona, choć w sumie wykonano o 78 wierceń więcej aniżeli wynikałoby z przyjętego modelu.

Na liczność punktów sondacyjnych ma wpływ zasadniczy wariancja badanego parametru w populacji generalnej obserwowanej cechy złoża, co wyraża się związkiem

$$n = f(\sigma^2)$$

Inne zmienne, jak δ i t we wzorze (3), o którym później będzie mowa, dotyczą jedynie z góry założonej dokładności i wiarygodności wyników. To skłoniło mnie do próby opracowania metody dla określenia choćby przybliżonej wariancji cech charakteryzujących torfowiska, wykorzystując liczny materiał statystyczny z badań torfoznawczych [5] do [26]. Biorąc pod uwagę dane z ponad 400 torfowisk rozsianych na obszarze Polski zbadałem rozkład cech badanych i ich wariancję w próbach losowych.

2. OCENA WARIANCJI I JEJ MATEMATYCZNE UZASADNIENIE

Zajmę się pomiarami głębokościowymi torfowiska, parametru odznaczającego się największą zmiennością, największą wariancją. Oznaczmy przez x wartość pomiaru głębokości torfowiska na danym obszarze w punkcie P. Liczbę x można traktować jako wartość zaobserwowaną zmiennej losowej x. Taka interpretacja wartości x wzbudzić może pewne wątpliwości, gdyż teren nie jest czymś losowym, a wartość głębokości w już ustalonym punkcie jest określona jednoznacznie. Tak jest w istocie, lecz przed wyborem punktu P wartość x jest nieznana i w tym sensie jest losową.

Zmienną X uważać więc można za zmienną losową, przypadkową, tylko w sensie rozmaitości położeń punktu P w rozważanym terenie. Z określenia zmiennej losowej wynika, że jest ona zmienną typu ciągłego, której wartości spełniają nierówność:

$$x_{\min} \leqslant X \leqslant x_{\max}$$

gdzie $x_{\min}$ i $x_{\max}$ oznaczają odpowiednio wartość najmniejszą i największą głębokości torfowiska. Wartość $\overline{x}$ i σ^2 zmiennej X oblicza się ze znanych zależności:

$$\overline{x} = \int_{-\infty}^{\infty} x f(x)\, dx \tag{1}$$

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \overline{x})^2 f(x)\, dx \tag{2}$$

przy czym $f(x)$ oznacza funkcję gęstości zmiennej losowej X. Funkcja gęstości $f(x)$ głębokości torfowiska w punktach pomiarowych praktycznie rzecz biorąc nie jest znana, a w następstwie tego również nie jest znana wariancja σ^2 głębokości.

Jeżeli przyjmiemy, że średnia arytmetyczna może być miarą statystyczną cechy badanej w skończonym zbiorze złożonym z N elementów, to przy pomocy próby losowej o objętości $n < N$, wartość średniej arytmetycznej badanej cechy odchylać się będzie w jedną lub drugą stronę od średniej populacji generalnej. Maksymalne odchylenia jak również wiarygodność wyników muszą być rzecz zrozumiała naprzód przyjęte i są możliwe do ustalenia dzięki fachowemu przygotowaniu badacza. Niezbędną liczebność próby losowej, odpowiadającej wymaganej dokładności i prawdopodobieństwu można obliczyć na podstawie znanego wzoru (3):

$$n = \frac{t^2 \cdot \sigma^2}{\delta^2} \tag{3}$$

gdzie δ oznacza maksymalne odchylenie średniej próby od średniej populacji generalnej, t jest zmienną „testu Studenta", σ^2 oznacza wariancję badanej cechy w populacji generalnej.

Ocena wariancji badanej cechy nie zawsze jest możliwa w stadium przygotowawczym do badawczej próby losowej, w którym musi być ustalona objętość tej próby. Wiadomo jedynie, że liczebność próby zależy od wariancji cechy w badanej populacji generalnej, a więc jest ona funkcją σ^2:

$$n = f(\sigma^2)$$

Ponieważ nie jest możliwe określenie σ^2, gdyż wariancja populacji generalnej zazwyczaj nie jest znana, można ją z konieczności „szacować od góry" przy pomocy s^2 takiej, że:

$$s^2 \geqslant \sigma^2 \tag{4}$$

Gdy uda się ją znaleźć, to wzór (3) przyjmie postać (5):

$$n_s = \frac{t^2 \cdot s^2}{\delta^2} \tag{5}$$

przy czym $n_s \geqslant n$.

Objętość próby losowej n_s będzie odpowiadała stawianym warunkom, gdyż już z założenia objętość n jest wystarczająca. Określamy wartość s^2 własnością majoranty σ^2.

Można takie „odgórne szacowanie" s^2 przeprowadzić, gdy na podstawie poprzednich badań istnieje informacja o dyspersji badanej cechy w populacji generalnej lub gdy przeprowadzimy pewne specjalne badania w celu jej chociaż przybliżonego ustalenia.

W zagadnieniu określonym tytułem nie mamy żadnego statystycznego rozpoznania wariancji badanej cechy w populacji generalnej, ale zakładamy formę jej rozkładu.

W pierwszym przybliżeniu założymy rozkład dwupunktowy cechy badanej. Przyjmujemy średnią wartość cechy x oraz k_1 i k_2 jej najmniejszą i największą wartość w populacji generalnej

$$k_1 \leqslant x_{\min}$$

$$k^2 \geqslant x_{\max}.$$

Jakikolwiek by był faktyczny rozkład wartości badanej cechy dokoła $\overline{x}$ w przedziale $[k_1, k_2]$, nieznana wartość σ^2 jest na pewno mniejsza bądź też co najwyżej równa s^2, gdyż wariancja dwupunktowego rozkładu jest majorantą każdego rozkładu w przedziale $[k_1, k_2] \cdot s^2$ dwupunktowego rozkładu cechy może być obliczona z równania (6).

$$s^2 = (\overline{x} - k_1) \cdot (k_2 - \overline{x}) \tag{6}$$

Oprócz wielkości k_1 i k_2 zależy s^2 jeszcze od położenia średniej wartości badanej cechy wewnątrz przedziału $[k_1, k_2]$ i przyjmuje największą wartość gdy $\overline{x} = \frac{1}{2}(k_1 + k_2)$, to jest znajduje się pośrodku przedziału i posiada równą ilość elementów o wartościach cechy k_1 i k_2, wtedy

$$s^2 = \left(\frac{k_2 - k_1}{2}\right)^2 \tag{7}$$

Jeśli można by było przyjąć jakąś mniejszą wartość jako majorantę σ^2 innego rozkładu, to zmniejszy się niezbędna objętość próby losowej, a przez to nakład pracy i kosztów.

Cel, któremu podjęte rozważanie ma służyć nie wymaga jak stwierdza praktyka oceniania niewiadomej σ^2 przy pomocy stosunkowo dużej wariancji dwupunktowego rozkładu. Okazuje się, że w praktyce można przyjąć w przedziałach $[k_1, \overline{x}]$ i $[\overline{x}, k_2]$ rozkład prostokątny badanej cechy, gdyż częstotliwość wartości cechy w obu przedziałach jest raczej stała. Przyjmując więc częstotliwość w pierwszym przedziale jako f_1, a w drugim oznaczając ją przez f_2, funkcję gęstości w obu przedziałach określa równanie (8):

$$f(x)=\begin{cases} f_1 \text{ dla } k_1 \leqslant x \leqslant \bar{x} \\ f^2 \text{ dla } \bar{x} \leqslant x \leqslant k_2 \end{cases} \tag{8}$$

Stałe f_1 i f_2 są określone przez położenie x i wielkości przedziału $[k_1, k_2]$ i można je obliczyć, gdyż na podstawie teorii rozkładu zmiennej losowej musi zachodzić (9), pole pod krzywą rozkładu jest równe jedności.

$$f_1(\bar{x}-k_1)+f_2(k_2-\bar{x})=1 \tag{9}$$

Na podstawie (1) w rozważanym przypadku mamy:

$$\int_{-\infty}^{\infty} \mathrm{x}\, f(x)\, dx=\int_{k_1}^{\bar{x}} \mathrm{x} f_1\, dx+\int_{\bar{x}}^{k_2} \mathrm{x}\, f_2\, dx \tag{10}$$

Zaznaczone wyżej całkowanie daje

$$1/2\, f_1\cdot(\bar{x}^2-k_1)+\ 1/2\, f_2\cdot(k_2-\bar{x}^2)=\bar{x} \tag{11}$$

Wzory (9) i (11) tworzą układ równań, z którego po rozwiązaniu otrzymujemy:

$$\begin{aligned} f_1 &= \frac{k_2-\bar{x}}{(\bar{x}-k_1)\cdot(k_2-k_1)} \\ f_2 &= \frac{\bar{x}-k_1}{(k_2-\bar{x})\cdot(k_2-k_1)} \end{aligned} \tag{12}$$

Na podstawie (9) i (12) oraz uwzględniając znany wzór (2) na wariancję zmiennej x w populacji generalnej, możemy obliczyć majorantę wariancji badanej cechy:

$$s^2=\int_{k_1}^{\bar{x}}(x-\bar{x})^2\cdot f_1 dx+\int_{\bar{x}}^{k_2}(x-\bar{x})^2\cdot f_2 dx \tag{13}$$

Po rozwiązaniu całek (13) otrzymujemy dla dowolnej średniej

$$s^2=1/3\,(k_2-\bar{x})\cdot(\bar{x}-k_1) \tag{14}$$

3. WNIOSKI I ZASTOSOWANIA

Wzór (14) ułatwia „odgórne szacowanie" wariancji s^2 jako majoranty w przypadku gdy badacz przygotowujący próbę losową (posiada w przedmiocie badania potrzebną wiedzę i rzeczowe podejście do zagadnienia) może założyć rozkład prostokątny badanej cechy w przedziałach $[k_1, \bar{x}]$ i $[k_2, \bar{x}]$. Łatwo mu to przyjdzie, gdy istnieją rzeczowe uzasadnione powody do przypuszczenia, że krańcowe wartości badanej cechy k_1 i k_2 nie występują w sposób uprzywilejowany (nie ma skupień na końcach przedziałów), a wartość przeciętna w stosunku do wartości sąsiednich nie jest rażąco zaniżona.

Ocena zmienności na podstawie wzoru (14) powinna być więc poprzedzona przez:

a) Ustalenie wartości k_1 i k_2 przez podanie wartości badanej cechy, które nie będą przekroczone w dół i w zwyż.

b) Ponieważ $\bar{x}$ nie jest do dyspozycji, gdyż jest celem badania, należy przyjąć roboczą wartość $\bar{x}$ przy założeniu największej zmienności S^2, to znaczy najbliżej wartości $\bar{x} = {}^1/_2 (k_1 + k_2)$ (patrz dyskusja wartości s^2, wzór 6).

Rozważania teoretyczne w zastosowaniu do praktycznej oceny wariancji parametru torfowiska poparte dużym materiałem statystycznym z badań w kategorii C_1 i C_2 potwierdzają możliwość przyjęcia w wyżej przytoczonym sensie rozkładu prostokątnego i określenia wielkości k_1 i k_2 oraz średniej wartości $\bar{x}$ przy planowaniu liczebności próby losowej za pomocą wzoru (5) i wspomagającego wzoru (14). Okazało się bowiem, iż małe wartości cechy badanej i duże jej wartości tworzą dwa rozkłady prostokątne, w których k_1 jest wartością najmniejszą, a k_2 wartością największą parametru badanego torfowiska.

Jest to pewne uproszczenie zagadnienia, ale uzasadnione i dające możność założenia błędu szacowania, co w innych przypadkach jest problematyczne.

W tabeli 1 obliczono liczność próby przy założeniu $t = 0{,}95$ i $\delta = (5$ cm, 7 cm i 10 cm) dla wspomnianych wyżej dwu torfowisk w basenie rzeki Biebrzy. Porównując wyniki pomiarów można było stwierdzić, że we wspomnianych badaniach liczność próby przyjęta na wyczucie, dla pierwszego torfowiska odpowiada $\delta = 7$ cm, a dla drugiego torfowiska $\delta = 12$ cm.

Jak łatwo zauważyć wzór (5) podaje zależność hiperboliczną n od δ. Tabela 1 jest również wyrazem tej zależności i jak wynika z przebiegu tej zależności $\delta = 10$ cm jest progiem, który nie należy obniżać, gdyż wprawdzie zmniejsza się nieco liczebność próby losowej, ale zbyt szybko rośnie maksymalne odchylenie średniej empirycznej od wartości średniej w populacji generalnej.

Tabela 1

Liczebność niezbędnej próby losowej przy różnych wartościach δ

Torfowisko	k_1	k_2	X	δ		
				5 cm	7 cm	10 cm
				n	n	n
I	20	370	135	1384	706	347
II	30	620	285	4358	2262	1091

LITERATURA

1. M a k s i m o w Al.: Torf i jego użytkowanie w rolnictwie. Państw. Wydawn. Rolne i Leśne. Warszawa 1965 r.
2. Z u b r z y c k i St.: Zastosowania Matematyki nr 3. Wrocław 1957 r.
3. Przegląd Geologiczny nr 9 1959, s. 415—419.
4. C i c h y T., H o h e n b e r g R., P r a ż a k M.: Zeszyty Naukowe WSR nr 9, 1958 — s. 37—50.
5. T o ł p a St. i współpracownicy: Dokumentacja geologiczna torfowisk rejonu jeziora Białokońskie-Koninek, Niemierzewo — 1962.
6. Dokumentacja geologiczna torfowisk rejonu Klepicz — Tarnówko 1962.
7. Dokumentacja geologiczna torfowisk rejonu Powidz — Wylatkowo — Anastazewo, 1961 r.
8. Dokumentacja geologiczna torfowisk rejonu Wronczyn — Pobiedziska, 1961.
9. Dokumentacja geologiczna torfowisk rejonu Grzmiąca — Wielawino, 1962.
10. Dokumentacja geologiczna torfowisk rejonu Kowalki — Sucha Szczecinecka, 1962.
11. Dokumentacja geologiczna torfowisk rejonu Wszembórz — Pyzdry, 1961.
12. T o ł p a St. i współpracownicy: Dokumentacja geologiczna torfowisk rejonu Przesieki — Żelichowo, 1962.
13. Dokumentacja geologiczna torfowisk rejonu jeziora Lednica, 1961.
14. Dokumentacja geologiczna torfowisk rejonu Gozdowo — Samorzewo, 1961.
15. Dokumentacja geologiczna torfowisk rejonu Racot — Turów — Dalewo, 1962.
16. Dokumentacja geologiczna torfowisk rejonu Kunowo, 1962.
17. Dokumentacja geologiczna torfowisk rejonu Swarzędz — Główna — Uzarzewo, 1961.
18. Dokumentacja geologiczna torfowisk rejonu Sroczyn — Turostowo, 1961.
19. Dokumentacja geologiczna torfowisk rejonu jezioro Strzyżewskie — jezioro Zioło, 1961.
20. Dokumentacja geologiczna torfowisk rejonu Kwilcz, 1962.
21. Dokumentacja geologiczna torfowisk rejonu Mieścisko — Gołaszewo, 1961.
22. Dokumentacja geologiczna torfowisk rejonu Smiłowo — Żelgniewo, 1962.
23. Dokumentacja geologiczna torfowisk rejonu Żelazkowo — Jelito, 1961.
24. Dokumentacja geologiczna torfowisk rejonu Rychlik, 1962.
25. Dokumentacja geologiczna torfowisk rejonu Lutol Mokry — Jasieniec, 1961.
26. Dokumentacja geologiczna torfowisk rejonu Wolsztyn, 1961.

РЕЗЮМЕ

В предварительных полевых опытах с целью квалификации торфяных залежей и их ресурсов в разных категориях естественного исследования большое значение имеет метод рассеяния пунктов. В этих работах исследователь должен преодолеть основную трудность, а именно установить такое число бурений на исследованной территории, которое гарантирует четкость оценки исследованного параметра залежи.

Автор исходит из формулы на необходимую численность пробы 5, которая является модификацией известной формулы:

$$n_s = \frac{t^2 \cdot s^2}{\delta^2}$$

Так как исследователю недоступно, ибо варианция исследованной особенности залежи не известны, её плотность и среднюю ценность следует оценивать с помощью эмпирической варианции данных, полученных на основе приведенных в настоящей работе теоретических рассуждений и обильного статистического материала из четырёхсот исследований (5—26) на территории Польши. На основе исследованного материала принимается комбинация двухпунктового и прямоугольного расположения и принимается численное превосходство s^2 над σ^2

$$s^2 = {}^1/_3\,(x - k_1) \cdot (k_2 - \overline{x})$$

где $k_1 \leqslant x_{min}$ и $k_2 \leqslant x_{max}$ можно поэтому принять, что малые и большие ценности исследованной особенности создают два прямоугольные расположения, в которых k_1 — наименьшая ценность, а k_2 — наибольшая. Это является обоснованным упрощением и дает возможность определения ошибки оценки, что в других случаях проблематично. Автор указывает, что при употреблении формулы (5) и вспомагающей её (14) можно рассчитать численность зондированных пунктов на исследованной территории. На конкретном примере в таблице 1 отображена численность бурений на двух торфяниках при учете разных ценностей.

SUMMARY

In the introductory field works on the qualification of the peat layers and their store, in various categories of scientific investigation the method of disperse points appears to be very useful. In these field works the explorer has to overcome the fundamental difficulty depending in the establishment of the bore-holes number within a given area, such a number of bore-holes to make an exact estimation of the investigated parameter of the layer being examined.

The author comes from the equation (5) that is a modification of well known equation — $n_s = \dfrac{t^2 \cdot s^2}{\delta^2}$

As δ is not given because the variation of the facture being examined as well as its mean value is unknown, one could establish it by means of the empiric variation obtained on the basis of theoretical calculation and statistic materials from 400 peatbog investigation in Poland (5—26).

On the basis of the investigated material that was worked out I assume the combination of two points distribution and rectangular distribution. It has been magnifying σ^2 by value S^2

$$s^2 = \frac{1}{3}\,(\overline{x} - k_1) \cdot (k_2 - \overline{x}) \text{ where}$$

$k_1 = x_{min}$ and $k_2 = x_{max}$. It appeared to be possible to assume that small values of investigated feature and its great values form two rectangular distributions: k_1 being the least value, and k_2 the greatest one. Though the task is thus being made simplified, it is substantiated because in this way it is possible to determine the error in the estimation, what in other cases would be problematical.

The author shows that by using of the equation (5) and equation (14) supporting it, it is possible to calculate the number of the bore-holes on investigated area. On the table 1 the bore-holes number is given for two peat bogs by applying different δ values.