

STATYSTYCZNE OPRACOWANIE WYNIKÓW KONTROLI JAKOŚCI ROBÓT ZIEMNYCH

Mieczysław Połoński¹

1. Metodyka statystycznego opracowania wyników

Przeprowadzona kontrola zagęszczenia budowli ziemnych dostarcza kilkadziesiąt lub kilkaset wyników w postaci nieuporządkowanych zbiorów liczbowych. Stanowią one statystyczną próbę poszczególnych parametrów gruntu traktowanych jako zmienne losowe. Dotyczy to gęstości objętościowej gruntu, gęstości objętościowej szkieletu gruntowego, wilgotności oraz wskaźnika lub stopnia zagęszczania. Na podstawie parametrów charakteryzujących zagęszczenie określa się, czy spełnione są odpowiednie wymogi techniczne. Dotyczą one odchylenia standardowego, które określa rozrzut wyników od wartości średnich, czyli jednorodność zagęszczenia na całym obiekcie, oraz ilość wyników, których wielkość jest mniejsza od wymaganego minimum dla danej klasy budowli. Zebrane wyniki pomiarów terenowych stanowią próbę do określenia cech populacji generalnej, tzn. zbioru wszystkich teoretycznie możliwych wyników pomiarów dla danego parametru gruntowego na danym obiekcie; im liczniejsza jest próba tym większe jest prawdopodobieństwo prawidłowego oszacowania parametrów populacji generalnej.

Statystyczne opracowanie wyników kontroli zagęszczenia obejmuje:

- określenie podstawowych parametrów próby losowej (zbioru danych wejściowych),
- zbudowanie histogramu danych empirycznych,
- postawienie hipotezy o zgodności danych empirycznych z jednym z rozkładów teoretycznych,
- odrzucenie lub stwierdzenie braku podstaw do odrzucenia postawionej hipotezy przy pomocy testów zgodności,
- wybór rozkładu teoretycznego najwierniej opisującego dane empiryczne na podstawie wyników testów zgodności,
- obliczenie na podstawie dobranego rozkładu teoretycznego procentowej ilości wyników niespełniających wymaganego warunku,
- sformułowanie końcowego wniosku.

1.1 Określenie podstawowych parametrów próby losowej

Na podstawie posiadanych wyników pomiarów oblicza się podstawowe parametry próby, które wstępnie charakteryzują badany zbiór danych. Są to:

- liczebność próby,
- wartość minimalna i maksymalna,
- średnia arytmetyczna,
- odchylenie standardowe,
- współczynnik asymetrii i zmienności.

Wyniki obliczeń należy dokładnie przeanalizować i określić czy wszystkie elementy próby nadają się do dalszej analizy. Eliminuje się dane wyraźnie odbiegające wartością od pozostałych, gdyż mogą być obciążone błędami, np. pomiarowymi, laboratoryjnymi itp. Liczebność zbioru, jaki zostanie przyjęty do dalszych obliczeń nie powinna być mniejsza niż 25 elementów.

1.2 Zbudowanie histogramu danych empirycznych

¹ dr hab., Szkoła Główna Gospodarstwa Wiejskiego

Na podstawie wstępnie zweryfikowanych danych ustala się empiryczny rozkład badanego parametru budując histogram. W tym celu na osi odciętych odkłada się przedziały, w których znalazły się wszystkie obserwacje, a na osi rzędnych liczbę elementów w każdym przedziale. Liczba przedziałów, na które dzielimy cały zakres obserwacji zależy od liczebności zbioru i powinna być rzędu $k = 1 + 3,3 \log(n)$, gdzie n jest liczebnością zbioru, jednak liczba przedziałów nie powinna być mniejsza od 5 (Benjamin, Cornell 1977). Na wykresie możliwa jest wizualna ocena układu wyników w próbie, tzn. których wyników jest najwięcej, które odbiegają od wartości średnich itp. Możliwe jest również wstępne oszacowanie, do którego ze znanych typów rozkładów zbliżony jest rozkład empiryczny badanej próby w celu postawienia hipotezy o zgodności danych empirycznych ze wskazanym rozkładem teoretycznym i zastąpienia otrzymanego histogramu empirycznego funkcją matematyczną opisującą rozkład badanego parametru.

1.3 Postawienie hipotezy o zgodności danych empirycznych z jednym z rozkładów teoretycznych

Należy poszukiwać rozkładu, który najwierniej oddaje charakterystykę analizowanego parametru na podstawie posiadanej próby losowej. Najczęściej używanym rozkładem prawdopodobieństwa jest rozkład normalny o funkcji gęstości określonej wzorem:

$$f_x(x) = \frac{1}{s_x \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x - m_x}{s_x} \right)^2 \right] \quad \text{dla } -\infty \leq x \leq +\infty \quad (1)$$

gdzie:

x – zmienna losowa (badany parametr gruntowy),

s_x – odchylenie standardowe z próby,

m_x – wartość średnia z próby.

Poza rozkładem normalnym wskazane jest także rozpatrzenie innych rozkładów teoretycznych, a szczególnie (Benjamin, Cornell 1977):

- log normalny przesunięty o postaci

$$f_x(x) = \frac{1}{\sqrt{2\pi} \sigma_{\ln y}} \exp \left[-\frac{1}{2} \left(\frac{\ln(x-a) - m_{\ln y}}{\sigma_{\ln y}} \right)^2 \right] \quad (2)$$

gdzie:

$\sigma_{\ln y}$ – odchylenie standardowe logarytmu naturalnego zmiennej $Y = x - a$,

$m_{\ln y}$ – średnia $\ln Y$,

a – przesunięcie.

- beta przesunięty o postaci

$$f_x(x) = \frac{1}{(b-a)^{r-1} \cdot \beta} (x-a)^{r-1} \cdot (b-x)^{r-1} \quad \text{dla } a \leq x \leq b \quad (3)$$

gdzie:

a, b – dolne i górne ograniczenie rozkładu,

r, t – parametry kształtu,

a – przesunięcie.

Rozkład beta jest rozkładem cztero-parametrycznym, umożliwia więc dokładne dobranie kształtu rozkładu do posiadanej próby.

- gamma przesunięty o postaci:

$$f_x(x) = \frac{\lambda^k \Gamma(k) x^{k-1} e^{-\lambda(x-a)}}{\Gamma(k)} \quad \text{dla } x \geq a \quad (4)$$

gdzie:

$\Gamma(k)$ – funkcja gamma,

k, λ – parametry rozkładu,

a – przesunięcie.

Przeanalizowanie wyżej wymienionych rozkładów powinno zapewnić dobranie rozkładu właściwego dla danej próby losowej. Można rozszerzyć obliczenia o inne niewymienione tu rozkłady.

1.4 Odrzucenie lub stwierdzenie braku podstaw do odrzucenia postawionej hipotezy przy pomocy testów zgodności

Do zastąpienia rozkładu empirycznego rozkładem teoretycznym nie wystarczy intuicyjne stwierdzenie podobieństwa ich przebiegu. Służą do tego specjalnie skonstruowane testy zgodności, przy czym określają one czy nie ma podstaw do odrzucenia hipotezy o zgodności rozkładu teoretycznego i empirycznego. Najczęściej stosuje się dwa: test Pearsona chi-kwadrat i test Kołmogorowa-Smirnowa. Pierwszy oparty jest na histogramie zwykłym, drugi na histogramie skumulowanym. Testowanie przeprowadza się na określonym poziomie istotności (5, 10, 20%), przy czym jako minimalny należy przyjąć poziom 5%.

1.5 Wybór rozkładu teoretycznego najwierniej opisującego dane empiryczne na podstawie wyników testów zgodności

Jeśli przeprowadzony test wykaże brak podstaw do odrzucenia postawionej hipotezy o zgodności rozkładów empirycznego i teoretycznego (na badanym poziomie istotności) można do dalszych obliczeń przyjąć określony typ rozkładu teoretycznego. Stosunkowo często może się zdarzyć, że tę samą próbę można opisać różnymi typami rozkładów wykazującymi podobne wyniki testowania zgodności. Do dalszych obliczeń należy przyjąć ten rozkład, który daje najmniejszy sumaryczny błąd kwadratowy w teście Pearsona. W przypadku, gdy wszystkie rozpatrywane rozkłady odrzucają postawione hipotezy należy dokładnie przeanalizować wartości parametru w posiadanej próbie czy nie są obciążone błędami. Jeśli nie dysponujemy inną próbą, orientacyjne wyniki dalszej analizy można uzyskać zakładając zgodność danych empirycznych z rozkładem normalnym. Przy interpretacji takich obliczeń należy pamiętać, że są to tylko dane przybliżone.

1.6 Obliczenie na podstawie wybranego rozkładu teoretycznego procentowej ilości wyników niespełniających wymaganego warunku

Obliczenie procentowej ilości wyników mniejszych od wymaganej wartości granicznej należy przeprowadzić na podstawie rozkładu teoretycznego, który został przyjęty do dalszych obliczeń w punkcie piątym. W przypadku rozkładu normalnego obliczenia można przeprowadzić na podstawie tablic statystycznych. Jednak zarówno w tym, jak i w pozostałych przypadkach najwygodniej jest

przeprowadzić całkowanie funkcji gęstości danego rozkładu teoretycznego w określonych granicach i na tej podstawie obliczyć procentowy udział poszczególnych wartości badanego parametru.

1.7 Sformułowanie końcowego wniosku.

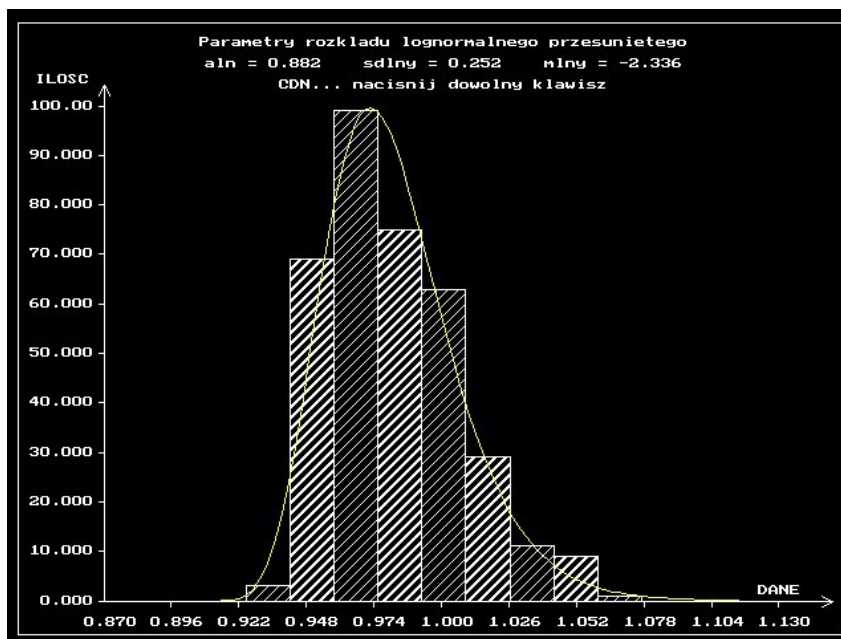
Na podstawie określonych parametrów podstawowych obiektu, a szczególnie wartości średniej arytmetycznej, odchylenia standardowego oraz procentowej ilości wyników, które nie spełniają stawianych wymagań możliwe jest sformułowanie wniosku o prawidłowej jakości wykonanych robót lub konieczności wykonania poprawek i przeprowadzenie ponownej kontroli zagęszczenia oraz porównania z wymogami technicznymi ustalonymi dla poszczególnych klas budowli.

2. Opis programu obliczeniowego SOROZ

W celu ułatwienia i przyspieszenia procesu analizy statystycznej autor referatu opracował specjalny program obliczeniowy pod nazwą SOROZ (System Oceny Robót Ziemnych).

Program dzieli się na dwie podstawowe części. Pierwsza z nich jest edytor, w którym umożliwiono wprowadzanie danych z klawiatury i z dysku, przeglądanie i modyfikowanie danych, zapis na dysk, wyprowadzenie danych na drukarkę, oraz uzupełniono o kilka podstawowych funkcji systemu operacyjnego. Ponadto edytor umożliwi eksport i import danych w postaci tekstowej do komunikacji z innymi programami statystycznymi czy bazami danych. Dane, które zostały wprowadzone do programu (z dysku lub klawiatury) mogą być wielokrotnie modyfikowane. Przewidziano następujące operacje aktualizacji danych: dopisywanie nowych wartości z klawiatury, dołączanie innego zbioru danych z dysku, zmiana wartości danej, szukanie określonej wartości w zbiorze, maskowanie (zasłanianie do obliczeń) pojedynczej danej lub całej grupy, odsłanianie zamaskowanych wartości, sortowanie, przeglądanie. Dzięki tak skonstruowanemu edytorowi możliwe jest dowolne operowanie danymi. Można np. wykonać oddzielnie obliczenia dla poszczególnych warstw a następnie automatycznie połączyć je w jeden zbiór i realizować obliczenia dla całej budowli. Dopuszczone jest również przesłanie wartości ekstremalnych i sprawdzanie jak wpływają one na wyniki obliczeń itp.

Druga część programu dotyczy wykonania analiz statystycznych. Wszystkie wyniki obliczeń prezentowane są w dwóch postaciach: analitycznej i graficznej, z możliwością kierowania wyników obliczeń na ekran i drukarkę. W danych podstawowych obliczane są następujące wartości: liczba elementów zbioru, wartość minimalna, maksymalna, średnia arytmetyczna, błąd średniej, odchylenie standardowe, współczynnik zmienności próby i współczynnik asymetrii. Oddzielna opcja umożliwia wyświetlenie danych w postaci histogramu.



Rys.1. Przykładowy widok ekranu programu SOROZ z histogramem i krzywą rozkładu lognormalnego przesuniętego

Obliczenia dotyczące rozkładów mogą być przeprowadzone w trybie ręcznym lub automatycznie. W pierwszym z nich operator programu może wskazywać jeden z czterech typów rozkładu, (normalny, logonormalny, beta i gama) oraz sprawdzać jego zgodność z danymi empirycznymi przy pomocy dwóch testów (Chi-kwadrat i Kołmogorowa-Smirnowa). Proces obliczeniowy dotyczący danego rozkładu lub testu przeprowadzony jest bez potrzeby ingerencji użytkownika. W przypadku rozkładu beta operator ma możliwość deklarowania dwóch spośród czterech parametrów rozkładu (dolnego i górnego ograniczenia), co przy pewnej wprawie prowadzi do bardzo dokładnego dopasowania kształtu krzywej rozkładu do histogramu empirycznego. W przypadku obliczeń automatycznych nie ma potrzeby jakiegokolwiek interwencji użytkownika. Analizowane są wszystkie typy rozkładów oraz testowana jest ich zgodność z danymi empirycznymi.

Przepisy wymagają określenia, jaki procent wyników nie spełnia stawianych wymagań (Wolski i in. 1994). Program pozwala obliczyć dla dziesięciu wartości badanego parametru gruntowego jego procentowy udział w populacji generalnej. Określone jest to na podstawie całkowania dystrybuanty danego rozkładu teoretycznego. Rozkładem, który będzie przyjęty do obliczeń udziału procentowego poszczególnych wartości cechy w trybie automatycznych obliczeń jest ten, który najlepiej spełnia wymogi testu zgodności. Te dane w połączeniu z obliczonym odchyleniem standardowym pozwalają jednoznacznie określić czy badana budowla ziemna spełnia wymagania stawiane jej w obowiązujących przepisach.

3. Podsumowanie

Wynikiem przeprowadzonych obliczeń w programie SOROZ jest wskazanie teoretycznego rozkładu, który najlepiej opisuje dane z próby, poziom istotności, na którym nie ma podstaw do odrzucenia hipotezy o zgodności wskazanego rozkładu i danymi, oraz podstawowa informacja dla użytkownika, tzn. procentowy udział badanego parametru mniejszy od zadanych wartości granicznych. Dodatkowo, podawane są uzupełniające wartości statystyczne, takie jak: liczba danych, wartość minimalna i maksymalna, średnia arytmetyczna, odchylenie standardowe, błąd średniej, oraz współczynniki zmienności i asymetrii. Całość obliczeń zamieszczona jest na postronicowanym wydruku w postaci tekstu, wyników numerycznych jak i rysunków, które nadają się do

bezpośredniego zamieszczenia w dokumentacji, dotyczącej kontroli i odbioru badanego nasypu ziemnego. Dzięki zastosowaniu tego programu możliwe jest prawidłowe opracowanie danych zgodnie z założeniami Warunków Technicznych Wykonania i Odbioru Robót Ziemnych (Wolski i in. 1994) w każdym przedsiębiorstwie wykonawczym lub terenowym laboratorium, bez konieczności zagłębiania się w teoretyczne szczegóły interpretacji statystycznej.

Literatura:

- Benjamin J.R., Cornell C.A. 1977: Rachunek prawdopodobieństwa, statystyka matematyczna i teoria decyzji dla inżynierów. Wydawnictwa Naukowo-Techniczne. Warszawa.
- Wolski W., Mirecki J., Mosiej K., Połński M. 1994: Warunki techniczne wykonania i odbioru robót, 59-71. Broszura ze środków Narodowego Funduszu Ochrony Środowiska i Gospodarki Wodnej.