ZESZYTY NAUKOWE POLITECHNIKI ŚLĄSKIEJ

Nr 115 Budownictwo z. 14 1964

Mgr FRANCISZEK PRZYBYLAK

Katedra Matematyki A

NOWA METODA OPRACOWYWANIA WYNIKÓW OBSERWACJI (wg J. Neymana)

Referat niniejszy nie zawiera własnych wyników rozważań autora na podany temat. Celem głównym autora jest pokazanie zastosowania nowej metody i zachęcenie do stosowania jej przy własnej pracy badawczej. Tutaj przedstawiono tylko fragment jej zastosowania. Bardziej wszechstronne informacje na ten temat można znaleźć w książce J.W. Linnika "Metoda najmniejszych kwadratów i teoria opracowywania obserwacji". (PWN) Warszawa 1962 r. .

Chcę tutaj nawiązać do pracy Stefana Hausbrandta opublikowanej w kwartalniku naukowym "Geodezja i kartografia" t. IX z. 1 pt. "Parę uwag w sprawie możności wykorzystania zdobyczy statystyki matematycznej do szacowania pomiarów inżynieryjnych". Autor tej pracy w uwagach wstępnych pisze: "... w inżynieryjnej praktyce pomiarowej spotyka się właśnie specjalnie dużo układów o małej ilości spostrzeżeń nadliczbowych - chociażby np. wyznaczanie wielkości kąta lub długości z podwójnych pomiarów obejmujące - chyba powierdzieć można bez przesady około 90% czynności mierniczych, zmierzających do udzielenia nam informacji o stosunkach liczbowych panujących w otaczającym nas świecie. W dziedzinie tych codziennych podwójnych pomiarów wnioski, do których doprowadziłoby nas stosowanie funkcji Studenta w sposób proponowany przez probabilistów, byłyby po prostu zaskakujące". Cytat ten zanalizujemy później.

Najpierw podam parę pojęć potrzebnych w dalszych rozważaniach.
Przez $x_1 \ldots x_n$ będziemy oznaczali wyniki pomiarów wielkości a.
O wartościach $x_1 \ldots x_n$ zakładamy, że $\Delta_i \in N(0,\sigma)$ gdzie $\Delta_i = x_i - a$ $(i = 1, \ldots n)$ tzn., że błędy losowe są niezależne i łącznie normalne

$$E(\Delta_i) = 0 \qquad (1)$$

$$E(\Delta_i^2) = \sigma^2 \qquad (2)$$

Równość (1) oznacza, że pomiary są nieobciążone (bez błędu systematycznego) natomiast równość (2) oznacza, że pomiary są jednakowo dokładne.
Oznaczmy

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Korzystając z metody najmniejszych kwadratów otrzymujemy oszacowanie dla a

$$a \approx x$$

Można obliczyć korzystając z (1) i (2), że

$$E(\bar{x}) = a \text{ i } D^2(\bar{x}) = \frac{\sigma^2}{n}$$

Średnia $\bar{x}$ jest zmienną losową o rozkładzie normalnym, tak więc $\bar{x} \in N(a, \frac{\sigma^2}{n})$.

Stąd mamy

$$P(|\bar{x} - a| > \frac{2{,}58\sigma}{\sqrt{n}}) = 0{,}01 \qquad (3)$$

Oszacowanie powyższe jest dosyć proste, jednakże największy kłopot sprawia fakt, że nie znamy wariancji obserwacji.
Otrzymane oszacowanie otrzymaliśmy metodą tzw. klasyczną.

Teraz zajmiemy się nową metodą podaną przez statystyka amerykańskiego J. Neymana (Uspiechy mat. nauk 10 1944 str. 207-229).
Metodę tę zastosujemy jedynie do bezpośrednich pomiarów grupowych jednakowo dokładnych.
Przy zastosowaniu tej metody skorzystamy z twierdzeń:

Twierdzenie 1

Wariancja z próby s^2 oraz $\bar{x}$ są niezależnymi zmiennymi losowymi oraz s^2 ma ten sam rozkład co zmienna $\frac{\sigma^2}{n}\chi^2_{n-1}$ gdzie $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

Twierdzenie 2

Iloraz

$$t_{n-1} = \frac{\bar{x} - a}{s}\sqrt{n - 1}$$

ma rozkład Studenta o n - 1 stopniach swobody.

Podane stwierdzenia pozwalają oszacować wielkość a za pomocą tzw. przedziałów ufności.

Z tablic rozkładu Studenta można znaleźć taką wartość dla n - 1 stopni swobody, że np.

$$P\left\{\left|\frac{x - a}{s}\sqrt{n - 1}\right| < \gamma\right\} = 0,99 \qquad (4)$$

*) Przez pomiary grupowe będziemy rozumieli takie pomiary kilku wielkości, które są dokonywane jednym i tym samym przyrządem, przy czym rozkład błędu dla każdej z nich jest normalny.

Wzór ten czytamy w następujący sposób: prawdopodobieństwo, że nieznana wielkość a będzie zawarta w przedziale

$$I = \left[\bar{x} - \frac{\gamma s}{\sqrt{n-1}},\ \bar{x} + \frac{\gamma s}{\sqrt{n-1}}\right]$$

jest równe 0,99.

Przedział I nazywamy "przedziałem ufności", zaś wartość 0,99 nazywa się "poziomem ufności", natomiast wielkość $E\left(\frac{2\gamma JS}{\sqrt{n-1}}\right)$ "dokładnością oszacowania".

Obecnie zajmiemy się cytatem. Niech wielkości $a_1 \dots a_k$ będą wielkościami mierzonymi, przy czym wielkości a_i odpowiadają dwa pomiary x_{i1}, x_{i2} $x(i = 1 \dots k)$. Zakładamy, że $x_{ij} \in N(a_i, \sigma)$ i wartości x_{ij} są niezależne.

Rozpatrzmy teraz wielkości

$$\bar{x}_i = \frac{1}{2}(x_{i1} + x_{i2})$$

$$s_i^2 = \frac{1}{2}\left[(x_{i1} - \bar{x}_i)^2 + (x_{i2} - \bar{x}_i)^2\right] = \frac{1}{2}\left[(x_{i1} - \frac{x_{i1} + x_{i2}}{2})^2 + \right.$$

$$\left. + (x_{i2} - \frac{x_{i1} + x_{i2}}{2})^2\right] = \frac{1}{4}(x_{i1} - x_{i2})^2 \quad (i = 1 \dots k)$$

Z twierdzeń 1 i 2 wynika, że zmienna $2\, s_i^2/\sigma^2$ ma rozkład χ_1^2 a więc dla sumy k niezależnych zmiennych $2\, s_i^2/\sigma^2$ otrzymujemy

$$\frac{2}{\sigma^2}\sum_{i=1}^{k} s_i^2 = \chi_k^2$$

Korzystając teraz z twierdzenia 2 otrzymujemy dla a_i przedziały ufności

$$I_i = \left[\bar{x}_i - \frac{\gamma_p S}{\sqrt{2k}}, \quad \bar{x}_i + \frac{\gamma_p S}{\sqrt{2k}}\right] \tag{5}$$

gdzie

$$S^2 = 2\sum_{i=1}^{k} s_i^2 = \frac{1}{2}\sum_{i=1}^{k}(x_{i1} - x_{i2})^2 \tag{6}$$

a γ_p znajdujemy z tablic rozkładu Studenta dla danego poziomu ufności p i dla k stopni swobody.

Wyniki powyższych rozważań zastosujemy w następującym przykładzie (przykład ten można znaleźć w cytowanej już książce Linnika).

W podanej niżej tablicy podany jest ciąg podwójnych pomiarów kąta pomiędzy przedmiotami nadbrzyżnymi, wykonanych za pomocą sekstansu.

α_i	x'_{i1}	x'_{i2}	$x'_{i1}-x'_{i2}$	$(x'_{i1}-x'_{i2})^2$	$\bar{x}_i$
60 18′	20″	30″	-10	100	25
112 16	10	10	0	0	10
47 2	20	15	5	25	17,5
83 37	30	40	-10	100	35
17 12	40	40	0	0	40
32 12	35	40	-5	25	37,5
8 10	20	30	-10	100	25
51 19	30	20	10	100	25
73 51	40	30	10	100	35
90 3	55	50	5	25	52,5
Suma				575	

Ze wzoru (6) obliczamy $S^2 = \frac{1}{2}\,575 = 287,5$ stąd $S=16,96$. Jeżeli za poziom ufności p przyjmiemy 0,99 to z tablic rozkładu Studenta znajdujemy dla $k = 10$ $\gamma_p = 3,17$.

Tak więc $\frac{\gamma_p S}{\sqrt{2r}} = \frac{3.17\ 16,96}{4,472} = 12,01$

Dla kąta z pierwszego wiersza otrzymaliśmy więc oszacowanie

$$60^\circ 18' 25'' - 12\ 01 \leqslant \alpha_1 \leqslant 60^\circ 18'\ 25'' + 12'',01$$

Oszacujmy teraz przedział, w którym znajduje się rzeczywista wielkość mierzonego kąta metodą klasyczną. Skorzystamy ze wzoru (3).

W tym przypadku oceną nieobciążoną dla $\frac{\sigma}{\sqrt{n}}$ jest wartość

$$\frac{s_1}{n} = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n(n-1)}} = \sqrt{\frac{25+25}{2.1}} = 5$$

Otrzymujemy więc

$$P(60^\circ 18' 25'' - 5.2,58 \leqslant \alpha_1 \leqslant 60^\circ 18' 25'' + 5.2,58) = 0,99$$

Widzimy, że przedział ten jest szerszy od przedziału otrzymanego przy korzystaniu z rozkładu Studenta.

Wyznaczymy teraz przedział ufności dla wartości kąta α_1 korzystając z wartości x_{11} i x_{12}.
Korzystamy ze wzoru (4) i ze wzoru na wartość x_i^2.
Mamy

$$s_1 = \frac{1}{2}\left|x_{11} - x_{12}\right| = 5$$

Teraz dla poziomu ufności $p = 0,99$ i dla 2 - 1 stopni swobody znajdujemy z tablic rozkładu Studenta $\gamma = 63,66$.

Otrzymujemy przedział ufności

$$60^{\circ}18'25'' - 5.63{,}66 \leqslant \alpha_1 \leqslant 60^{\circ}18'25'' + 5.63{,}66$$

Z obliczonych powyżej trzech przedziałów można wyciąg nąć następujące wnioski. W przypadku gdy jednym i tym samym przyrządem dokonujemy tylko dwóch pomiarów mierzonej wielkości, to stosowanie rozkładu Studenta dla tych dwóch pomiarów rzeczywiście nie daje rozsądnego wyniku (przy tak dużym poziomie ufności).
Natomiast gdy jednym przyrządem dokonujemy więcej pomiarów podwójnych (i nie tylko podwójnych), to zastosowanie rozkładu Studenta daje nam nawet węższy przedział, w któ rym znajduje się mierzona wielkość niż rozkład normalny.

Widzimy więc, że przy grupowych pomiarach (nawet podwójnych, lecz w dość licznej grupie) a chyba te właśnie stanowią 90% ogółu pomiarów, o których mówi Stefan Hausbrandt stosowanie rozkładu Studenta daje wyniki zadawalające.